DE GRUYTER MOUTON
DOI 10.1515/ling-2013-0059 — Linguistics 2014; 52(1): 167–203

Jong-Bok Kim

English copy raising constructions: Argument realization and characterization condition

Abstract: Together with raising constructions, English employs the so-called copy raising (CR) constructions. The traditional wisdom for the treatment of these CR constructions has been that the subject of the highest embedded clause governed by the CR verb is raised to the matrix subject, leaving behind a coreferential pronoun (pronominal copy) in the subject position. This kind of movement-based analysis raises both empirical and analytical issues, when considering great variations in the position of the pronominal copy. This paper shows that copy raising predicates are basically classified into two main types, genuine and perception ones, while displaying similarities as well as distinctive properties. Together with the augmentation of empirical data through extensive corpus search, the paper further suggests that the similarities and differences of these two types as well as the variations of the CR construction can be followed from tight interactions among the lexical properties of these two types organized in a hierarchical structure, constraints on the argument realization specifying how the arguments are mapped on syntactic and semantic elements, and characterization constraints on the CR construction involved.

Keywords: copy raising, theta-role, interpretive constraint, perception, pronominal copy

Jong-Bok Kim: School of English, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul, 130-701, Korea. E-mail: jongbok@khu.ac.kr

1 Introduction

Along with the infinitival raising construction in (1a), English employs the so-called copy-raising construction as given in (1b):

(1) a. *The lifeguards seem to be dancing across the water.*
b. *The lifeguards$_i$ seem like they$_i$ are dancing across the water.*

The main characteristic of the CR (copy raising) in (1b) is that the referent of the matrix subject is identical to that of the embedded clause's subject. In terms of truth-conditional meaning, both the typical raising and CR examples in (1) are synonymous to their counterparts with the expletive subject:

(2) a. *It seems that the lifeguards are dancing across the water.*
b. *It seems like the lifeguards are dancing across the water.*

Compared to the similar raising construction, the CR construction has not received much attention in theoretical linguistics except a few (see Rogers 1971, 1972; Potsdam and Runner 2001; Asudeh 2002; Landau 2009, 2011; Asudeh and Toivonen 2012). In order to capture the synonymous relation between (1b) and (2b), Rogers (1971, 1972, 1973) and subsequent traditional movement analyses (e.g., Ura 1998; Moore 1998; Rezac 2004) have introduced a movement operation dubbed *copy raising* or *Richard* as sketched in the following:[1]

(3) a. [*e*] *seem like* [*the lifeguards*] *are dancing across the water.*
b. [*The lifeguards*]$_i$ *seem like they*$_i$ *are dancing across the water.*

As shown in (3a), the movement operation raises the highest embedded clause's subject to the matrix subject, leaving behind its coindexing pronoun in the subject position. This process of *pronominal copy* accompanied with the movement will then generate sentences like (3b). If no movement operation occurs, the subject can be filled with the expletive *it* as in (2b). This type of movement-based *copy-raising* analysis seems to capture the systematic relation between the CR and its putative source, but also raises several intriguing questions, as pointed out by Potsdam and Runner (2001). For example, the first question concerns how the subject of the matrix clause can be raised from the subject of the finite embedded clause which is a Case position. Movement from a Case position is quite unorthodox in traditional movement analyses, violating the Tensed S Condition that blocks A-movement from a tensed clause (Chomsky 1982). A related question also arises with respect to the theta role of the matrix subject. If the matrix subject were raised from the embedded clause's subject, would the matrix subject in the CR receive its theta role from the embedded clause's predicate? If it were not raised from the embedded clause, what would assign a theta

1 The name *Richard* is originated from Rogers (1971, 1972) for a transformation rule deriving examples like *Richard seems like he is in trouble* from *It seems like Richard is in trouble*. The name *copy raising* reflects resemblance to subject-to-subject raising.

role to the matrix subject? In addition, the question arises of how the movement leaves an overt pronoun (a copy of some sort).

Moreover, authentic data tell us that the copying process is much more complex than the one sketched in (3a). For example, consider the following corpus examples:[2]

(4) a. *The girl seemed as if **her** mom was dying.*
(COCA 2001 FIC)
b. *The bed appeared as if someone had recently been dragged from **it**.*
(COHA 1827 FIC)
c. . . . *the scene appeared as though the children were up in the clouds falling through with the snow.*
(COCA 1998 FIC)

In (4a), the putative copy of the matrix subject is not in the subject position of the embedded clause, but in its specifier position. In (4b), the assumed copy of the matrix subject occurs in the prepositional object in the VP. Examples like (4c) give us another challenge within the movement-type analysis since the embedded clause contains no pronoun co-referential with the matrix subject *the scene*.

The pronominal copy in the non-subject position and even its absence illustrated in (4) have also been noted by the literature too, motivating different approaches (see, among others, Heycock 1994, Potsdam and Runner 2001, Landau 2009, 2011, and Asudeh and Toivonen 2012). For example, recognizing the problems in movement-based approaches, Potsdam and Runner (2001) offer a non-movement, base-generation analysis. Their analysis builds on the assumption that the CR subject is nonthematic when the subject binds a pronominal copy in the embedded clause's subject position (Sub-CR type), but it is thematic when there is no pronominal copy in the highest embedded clause's subject position (Nonsub-CR type). This dichotomy is based on the observation that the CR subject can have a *there*-expletive or an idiom chunk when the pronominal copy occurs in the embedded clause's subject position (Potsdam and Runner 2001; Landau 2009):

(5) a. *%There looks like there's gonna be a riot.*
b. **There seems like John expects there to be an election.*

2 The corpus examples we use are extracted from the corpora COCA (Corpus of Contemporary American English) and COHA (Corpus of Historical American English), both of which are freely available online at http://corpus.byu.edu/coca/ and http://corpus.byu.edu/coha.

(6) a. *%Advantage appears like it was taken of the workers.*
b. **Tabs appear as if the government keeps them on.*

Given the general assumption that the expletive and idiom chunk are indicative of the verb's not assigning a thematic role to the position in question (see Postal and Pullum 1988), we may conclude that unlike (5b) and (6b), the matrix subject in (5a) and (6a) is nonthematic.

With the distinction between the thematic and nonthematic subject, Potsdam and Runner (2001) take the Sub-CR type to involve a non-movement A-chain between the two subjects as illustrated in the following (cf. Rezac 2004 and Fuji 2005):[3]

(7) [*Richard*$_i$ *seems* [$_{XP}$ *like* [$_{TP}$ [*he*$_i$ [*is in trouble*]]]]]

This analysis can avoid the issue of the Tensed S Condition, but at the same time opens analytical and empirical issues, as also acknowledged by Potsdam and Runner (2001). For example, questions remain of what kind of roles the A-chain in CR plays and why the same A-chain cannot be formed with the *that*-complementizer CP as in **John seems that he is ill*. Another ensuing question concerns the treatment of Nonsub-CR examples like (4), as Landau (2011) points out. Consider the following two examples which differ with respect to the pronominal copy:

(8) a. *Your house sounds like nobody enjoys cleaning *(it).*
b. *That noise sounds like somebody's cleaning.*

Within Potsdam and Runner's account, the subjects in both examples here are identically thematic since there is no pronominal copy in the highest embedded subject position. This, however, then fails to capture the difference between the two: unlike (8b), the pronominal copy in (8a) is obligatory and in addition the property of the subject here seems to be different in that unlike *that noise* functioning as the perception, the subject *your house* is not directly linked to the predicate *sounds*.

As an attempt to solve such an issue, Landau (2011) resorts to the condition of P-source (perceptual source): the subject of CR is interpreted as the source of perception. Landau's main point, given in the following, is that a pronominal copy is

3 Potsdam and Runner (2001) assume that the Nonsub-CR type involves simple coindexation between the two thematic positions, but offer no clear account for this.

necessary in the complement clause if the matrix subject is not a P-source (or nonthematic):

(9) The P-source-Copy Generalization
Given a sentence "DP_i V_{perc} (to DP_j) like CP", where V_{perc} includes *seem, appear, look, sound, feel, smell, taste*, a copy (= pronoun coindexed with DP_i) is necessary in CP iff DP_i is not a P-source.
(PCG, Landau 2011: Ex. 26)

Landau's PCG then explains the examples in (8): since the subject *your house* in (8a) is not a P-source, there must be its pronominal copy. The situation is different in (8b): the subject *that noise* functions as a P-source and this makes the pronominal copy optional. Landau (2011) also suggests that the CR's complement CP is a predicate when the matrix subject is non-thematic and in such a case the pronominal copy is necessary. Meanwhile, when the CP is propositional, the matrix subject is thematic. In this case, the CP forms an "aboutness" relation with the matrix subject, eventually licensing non-pronominal copy cases. Appealing though the analysis sounds, it is questionable if there are any syntactic constraints on the position of the pronominal copy: it says nothing about the possible position of the pronominal copy. It is also doubtable why the aboutness relation holds only when the CP is propositional (the matrix subject is thematic). This paper recognizes the role of the aboutness condition, and further develops it into the similar notion of *characterization condition* applying not only in limited cases, but for all the CR examples (see Section 3.2).

Also recognizing the pivot role of the P-source, Asudeh and Toivonen (2012) differentiate CR verbs with *seem* and *appear* from those with perception verbs (*look, sound, smell, feel* and *taste*).[4] This distinction is motivated from the following contrast (Asudeh and Toivonen 2012: (14)):

(10) a. **Thora seems/appears like Chris has been baking sticky buns.*
b. *Thora smells/looks/sounds/feels/tastes like Chris has been baking sticky buns.*

As illustrated from the contrast here, the verbs *seem/appear* require a copy pronoun in the embedded clause's subject position but the perceptual verbs do not.

4 One main difference between Landau (2011) and Asudeh and Toivonen (2012) has to do with the status of P-source. Landau (2011) takes the P-source as a theta-role while Asudeh and Toivonen (2012) take it as a "semantic" role, a broader notion that subsumes temporal and locative adverbs as well.

This dichotomy also accompanies the assumption that the true CR subject must be interpreted as a P-source. Their analysis, couched upon the LFG framework, also provides a base-generation approach in which the copy raising subject and the copy pronoun are taken to be in a standard anaphoric binding relationship. This means that the matrix subject of the CR binds a pronoun somewhere in the complement clause. There is thus no limit to where the copy pronoun occurs in the embedded clause (see Asudeh and Toivonen 2012). As we will discuss in this paper, the naturally occurring data indicate that it is hard to distinguish the CR verbs *seem/appear* from the perception verbs (see Sections 3.1 and 5.2). They behave alike with respect to the copy raising.

As we have discussed so far, the previous main approaches of the CR in English have centered on the differences between *seem/appear* CR verbs and perceptual CR verbs, thematic role of the matrix subject, the role of P-source, and so forth. Building on the previous work, this paper tries to investigate the authentic uses of the construction and offer a new perspective based on the lexicalist, constraint-based framework, HPSG (Head-driven Phrase Structure Grammar). In particular, we first investigate authentic uses of the CR construction, using the online available corpora COCA and COHA (see Davies 2009, 2012). Based on the corpus search as well as the previous literature, we discuss main grammatical properties of the construction. We show that the authentic data give us no clear distinction between *seem/appear* CR verbs and perception verbs. Based on these observations, we argue that the licensing of the CR is closely tied up with the lexical properties of the verb involved, constraints on the argument realization, and characterization conditions in the CR.

2 General properties

2.1 On the predicate types and subject properties

As we have seen in the previous section, the CR is often found with traditional raising verbs like *appear* and *seem* as well as with physical perception verbs like *smell, feel, sound, look, taste*, and so forth. This is further evidenced from the following corpus data:

(11) a. *She seems like she is laughing hysterically.*
(COCA 2010 NEWS)
b. *She tried not to appear as if she was rushing away from the screaming house.*
(COCA 2003 FIC)

(12) a. *O'Brien was looking as if he expected to get shot at any moment.*
(COCA 2005 FIC)
b. *I felt as though I was in heaven.*
(COCA 2008 NEWS)
c. *The ham tasted like it had been in the icebox too long.*
(COCA 1991 FIC)
d. *How is it that today you smell as if you have just been eating onions?*
(COCA 2003 FIC)

All these examples are synonymous with those with the expletive subject *it*, showing a systematic alternation between the non-expletive and expletive subject CR sentences.

One main constraint we can observe from the CR data is that the matrix subject and embedded clause's subject are in a coreferential relation and the latter serves as the former's pronominal copy. The literature has noted that the violation of this coreferential and pronominal copy condition seems to yield ungrammatical sentences, in particular with the verbs *appear* and *seem* (see Postal 1974; Potsdam and Runner 2001; Landau 2011; Asudeh and Toivonen 2012):

(13) a. **The lifeguards appear as if **he** was dancing across the river.*
b. **He seems as though **she** could either crack a smile.*
c. **There seems like **John** expects there to be an election.*

The pronominal copy condition between the matrix and embedded clause's subject also holds with the expletive subject *it* and *there*. This condition also seems to be true in the corpus examples:

(14) a. *It seems like **it**'s always cold outside and too hot inside.*
(COCA 1993 FIC)
b. *It seems as if **it**'s no fun being an actor anymore.*
(COCA 2009 MAG)

(15) a. *All I know is that there seemed like **there** was always plenty of food.*
(COCA 1995 FIC)
b. *There sounds like **there** was a very cold side to her.*
(COCA 2008 SPOK)

When there is no agreement relation between the two subjects, we will have ungrammatical examples:

(16) a. *There seemed like **it** was raining.
b. *John seems like **there** is no tomorrow.

One thing worth noting here is, as an anonymous reviewer points out, acceptable examples like the following:

(17) a. *It seems like there is more at stake.*
(COCA 2012 NEWS)
b. *It seemed like there was a lot more energy last year.*
(COCA 2011 NEWS)
c. *Does it seem like there ought to be a premium adjustment on the price?*
(COCA 2005 SPOK)

In these examples, the matrix subject *it* and the embedded clause's subject *there* bear different agreement features. For example, *it* carries third-person and singular-number agreement features, while *there* has only "uninterpretable" third-person agreement features (see Radford 2010: 246). These sentences are thus acceptable simply because they are not CR examples, but base-generated ones with the expletive *it* as the matrix subject. When the matrix subject is the expletive *it*, there is thus no requirement for the two subjects to be coreferential (see Section 5.2).

With the index value of the subject including person, number, and gender, we can expect that the coreferential relation between the two subjects also affects the subject-verb agreement in the matrix and embedded clause (Kaplan-Myrth 2000; Rezac 2004):

(18) a. *There **looks** as if there **is** a problem.*
b. *There **look** as if there **are** problems.*
c. **There **looks** as if there **are** problems.*
d. *It looks like there are problems.*

Given that the expletive *there* in the embedded clause gets its agreement features from the postcopular NP in the embedded clause, the matrix subject *there* in (18a), coreferential with the embedded clause's subject *there*, must be singular too. Unlike this, the matrix subject *there* in (18b) and (18c) must be plural because of the postcopular plural NP *problems*. This is why (18c) is unacceptable. Nothing is wrong with (18d) where the matrix subject is the expletive, not related to the embedded clause's subject.

The observations we have made so far indicate that the CR construction, licensed by intransitive raising and perception verbs, is sensitive to several gram-

matical constraints including the agreement condition between the matrix subject and the highest embedded clause's subject. However, as we have hinted, there are cases where such an agreement condition does not hold. For example, consider the following examples:

(19) a. *The fact she went alone seems like* ***she*** *wasn't afraid.*
(COCA 2009 MAG)
b. *Richard seemed like the judges had decided to support Mary's complaint that* ***he*** *cheated.*
(Asudeh and Toivonen 2012: (79))

The two coreferential NPs in these examples are in the remote, deeper syntactic positions, challenging an account that links the matrix subject to the highest embedded clause's subject. Such variations in the position of the putative pronominal copy imply that syntax alone cannot answer the behavior of the CR constructions, which we will discuss in this paper.

2.2 On the property of the embedded clause

As we have seen from the examples in Section 2.1, the embedded clause is introduced only by *like*, *as if*, or *as though*, and the clause must be finite:[5]

(20) a. *His parents seem like/*that they are more active with their children.*
(COCA 1997 FIC)
b. *Prince appears as if/*that he bears the weight of the world on his narrow shoulders.*
(COHA 2007 MAG)
c. *Max looked as though/*that he'd been plunged into deep sea.*
(COCA 1994 NEWS)

In all these examples, the expressions *like*, *as if*, and *as though* cannot be replaced by the canonical complementizer *that*.

As noted by Bender and Flickinger (1999) and others, the embedded clause acts like a complement clause selected by the matrix predicate. The first argument for its complementhood comes from the obligatoriness of the *as-if* clause:[6]

5 The CR thus does not license examples like **Their parents seem like to be more active with their children*.

6 We use the *as-if* clause as the cover term for the clauses headed by *like*, *as if* and *as though*.

(21) a. *The lifeguards appear *(as if they were dancing across the water).*
b. *She seems *(like she is laughing hysterically).*
c. *I felt *(as though I was in heaven).*

In addition, these verbs select an AP as its predicative complement and can be replaced by the *as-if* clause (Kaplan-Myrth 2000, Asudeh 2002):

(22) a. *His imagery appears* { *xeroxed.* / *as if it is xeroxed.* }
b. *The wines taste* { *good.* / *as if they are good.* }

The examples with the expletive subject also support the complementhood of the clause:

(23) a. *It seemed as though all the doors formerly open to her were now shut in her face.*
(COCA 1993 MAG)
b. *To the casual observer it may appear as if the whole lake is affected.*
(COCA 1992 MAG)

Given the general assumption that verbs in English have at least one argument, the *as-if* clause here is the only possible candidate. In addition, note that the examples here cannot be thought of as extraposition – intraposing the clause back into subject position induces ungrammatical examples.

The extraction possibility can give us another indicator for the clause's complementhood. Consider the following contrast (data from Bender and Flickinger 1999):

(24) a. *The president that he looked* [*as if he was imitating* ____] *was Ford.*
b. **The president that he fell* [*as if he was imitating* ____] *was Ford.*

The traditional assumption is that constituents can be extracted from complement phrases but not from adjunct phrases because in the latter case the trace would not be properly governed (see Huang 1982). The extraction of the embedded clause's object in (24a) is possible, but the same process is not possible in (24b) in which the verb *fell* is an intransitive with no object, implying the *as-if* clause here is an adjunct clause. This also explains why the extraction here is not possible.

There are clear cases where the embedded clause introduced by *like*, *as if*, and *as though* occurs in the adjunct position:

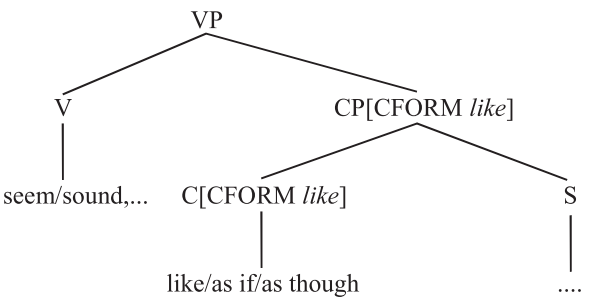
(25) a. *People switch jobs* [*as if they were double-parked*].
(COCA 1994 MAG)
b. . . . *but he spoke* [*as if he did not feel very sure of what he said*].
(COHA 1828 FIC)
c. *The little gals cried* [*as if their hearts would break*].
(COHA 1845 FIC)
d. *He ran* [*as if the track were the top of a hot stove*].
(COHA 1995 MAG)

The underlined main verbs here do not select the *as-if* clause. The *as-if* clauses here are all optional and function as modifiers. One main difference from the CR's complement clause is that such examples do not license the counterparts with the expletive *it* as the subject:

(26) a. **It spoke* [*as if he did not feel very sure of what he said*].
b. **It ran* [*as if the track were the top of a hot stove*].
c. **It squinted and shrugged* [*as if he had no idea what had brought on her attitude*].

Reflecting these syntactic properties, we can conclude that the CR predicate selects a sentential complement headed by the complementizer *like*, *as if*, or *as though*, projecting a structure like the following:

(27)



The structure illustrates that the CR verb selects a CP whose CFORM (complementizer form) value is *like*, subsuming all the three CFORM values *like*, *as if* and *as though*.[7] This value, originated from the head C, is projected to the mother CP as a head feature: this then will enable the CR verb to refer to the complement CP's

7 The main reason to assume *like* as the representative CFORM value is that the value has the highest frequency in the COCA and COHA.

form value, observing the locality condition.[8] Following Huddleston and Pullum (2002), we take *as if* and *as though* as a single compound word. The compound treatment receives support from the fact that the *if*-clause or *though*-clause cannot be coordinated or repeated:[9]

(28) a. **Prince seems as* [*if he has got everything*] *and* [*if he knows everybody*].
b. **It seems as* [*though he's got everything*] *and* [*though he knows everybody*].

These examples support the view that both *as if* and *as though* are nonseparable and single compound expressions.

3 Genuine copy raising vs. physical perception verbs

3.1 Similarities and differences

As hinted so far, the predicates in the CR can be classified into two groups: GCR (genuine copy raising) and PCR (perception copy raising). The former includes verbs like *seem* and *appear* while the latter has verbs like *smell*, *feel*, *look*, *sound*, and *taste*. As noted by Rogers (1972, 1973), Lappin (1984), Asudeh and Toivonen (2012) and others, this dichotomy appears to get support from the fact that the subject copy raising (or pronominal copy in the embedded clause's subject) is necessary in the GCR but is optional in the PCR type (data adopted from Asudeh and Toivonen 2012):

(29) a. **Jane seems like everything has gone wrong.*
b. *Jane looks like everything has gone wrong.*

As such, at first glance, there seems to be a clear contrast between the GCR and the PCR in allowing the pronominal copy. However, as noted by the literature including Rogers (1971), Heycock (1994), Potsdam and Runner (2001), and Lan-

8 For the motivations to introduce form values for English complementizers and the discussion of locality issues, see Sag et al. (2003: 487), Kim and Sells (2008: 90), and Sag (2012).
9 As suggested by Asudeh (2002) and Asudeh and Toivonen (2012), we might treat *like* or *as* as a preposition selecting a finite S or a clause headed by *if*, but may need an independent constraint to block examples like (28).

dau (2009, 2011), the GCR type also allows cases with no pronominal copy of the matrix subject in the embedded clause's subject position:

(30) a. *When I talked to her, she seemed like there would be no issues.*
b. *The situation appeared as if they were trying to hide who they really are.*

In the embedded clause of these examples, there is no expression coreferential with the matrix subject. Our corpus search also supports this position. Both the GCR and the PCR allow the violation of the coreferential relation between the two subjects, leaving the presumed pronominal copy in various positions or being inferred from the context. For example, our corpus search yields many cases where the matrix subject is coreferential with the specifier of the embedded clause's subject in both types:

(31) a. *He appeared as if **his heart** were broken by her speech.*
(COHA 1828 FIC)
b. *The girl seemed as if **her mom** was dying.*
(COCA 2001 FIC)
c. *The judge looked as if **his candy** had been stolen.*
(COHA 1951 FIC)
d. *LeRoi felt as if **his heart** had been ripped out of his chest.*
(COHA 2001 FIC)
e. *She sounded as though **her thoughts** were a million miles away.*
(COCA 2010 FIC)

The corpus also give us a variety of cases where the genitive specifier of the matrix subject in the GCR and PCR type is coreferential with the highest embedded clause's subject:

(32) a. *Her skin appeared as if **she** didn't take a bath for years.*
(COCA 2009 NEWS)
b. *His voice seemed as if **he** shouted all night.*
(COCA 2001 FIC)
c. *Her breathing sounds as though **she** is sleeping – she's faking it.*
(COCA 2007 FIC)
d. *His face looks as though **he** has emptied himself of every thought.*
(COCA 1991 NEWS)

There is also no difference between the GCR and the PCR in allowing the coreferential NP to occur as the object of the embedded clause:

(33) a. *The Peugeot appeared as if dust had created **it**.*
(COCA 2002 ACAD)
b. *. . . but so many of your story lines sound as if you take **them** right out of the headlines.*
(COCA 1998 SPOK)
c. *The lawn looked as if someone had brushed **it**.*
(COCA 1993 MAG)
d. *. . . , where everything else smelled like you'd squished **it** out between your toes.*
(COCA 2007 FIC)

In both types, the matrix subject can be also linked to the pronoun in the prepositional object position:

(34) a. *. . . , till the forest appears as if a tornado had passed over **it**, . . .*
(COHA 1850 MAG)
b. *The cabinets looked as though someone had thrown the pots into **them** from across the room.*
(COCA 1998 MAG)
c. *The others tasted as if all of the moisture and character had been wrung out of **them**.*
(COCA 1997 NEWS)

In addition to these cases, both the GCR and the PCR behave similarly in that the embedded clause may include no expression coreferential with the matrix subject at all, as evidenced from the following:

(35) a. *For me, studying Yiddish seemed as though I were traveling, instead, through the streets of a long-forgotten hometown.*
(COCA 2000 ACAD)
b. *In spite of that, or just for that reason, she appeared as if everything were finally in its place.*
(COCA 2002 FIC)
c. *Domestic in scale but collapsed in volume, they look as if air had been blown in and then sucked out.*
(COCA 2003 MAG)
d. *You sound as if the man has no choice in the matter.*
(COCA 1992 FIC)

The naturally occurring data we have seen so far tell us that both the GCR and the PCR behave alike in licensing non-pronominal copy cases. There is no telling difference between the two types with respect to the licensing positions of the pronominal copy in the embedded clause. Both types can license the pronominal copy not only in the subject but also in other positions such as the specifier of the subject, verbal object, and prepositional object position. There are also many cases where the highest embedded clause does not include a pronominal copy at all. This observation argues against the position to distinguish between the GCR with *seem* and *appear* and the PCR with the perception verbs (see Asudeh and Toivonen 2012).

3.2 Interpretive constraints

A variety of authentic data indicates that the pronominal copy constraint depends on context. It seems that, as argued by Rogers (1971) and Landau (2011), as long as the CR construction in question observes certain interpretive conditions, the pronominal copying constraint can be overridden. Consider the following (see Heycock 1994 and Landau 2011):

(36) a. *This noise seems/appears/sounds/*feels/*looks/*tastes like Eric is responsible for the production.*
b. *The moon seems/appears/looks/*feels/*sounds/*tastes like the orbit is nearly circular.*
c. *This book ?appears/?seems/sounds/looks/??feels/??smells like everyone should own a copy.*

In (36a), from perceiving the sound of the noise, the perceiver can infer Eric's responsibility for the production, but we cannot look or taste or even feel the noise. In (36b), the visual perception of the moon helps us infer about the orbit, but there is no way for us to feel or hear it. For instance, the auditory stimulus (*sound*) cannot be the source of an inference about the orbit, either. As for the book in (36c), the most natural verb is *sound* or *look* in the context where people are talking about the book. However, note that even the verbs *feel* or *smell* may be possible since we can have a context where one has direct experience with the book (e.g., hold it, smell it). As such, the CR construction requires that the embedded event (or state) be inferable from the matrix perceptual event. Adopting this observation from Rogers (1972), Landau (2011), Asudeh and Toivonen (2012), we also assume the following condition for the the matrix subject of CR to be interpreted as the source of perception:

(37) P-source Condition:
The matrix subject of the CR needs to serve as the (psychological) source of perception (P-source) in the eventuality involved.

Rogers (1972) points out that the subject of the CR construction must be understood as the perceptual source of the report, different from the expletive variant:

(38) a. *Harry looked to me like he was drunk.*
b. *It looked to me like Harry was drunk.*

This "cognitive presupposition" of the CR tells us the entailment difference here: (38a) entails the speaker's seeing Harry but (38b) does not. In a similar manner, Asudeh and Toivonen (2012) suggest that the matrix subject is an entailed participant in the eventuality in question. For example, if there is no contextual clue providing the P-source for the CR subject, examples like (39a) are infelicitous unlike those like (39b):

(39) a. *#Tom seems like he's cooking.*
b. *It seems like Tom is cooking.*

If we see Tom doing something at the stove in the kitchen, but we cannot infer if he is cooking or not, (39a) is not a proper utterance, as agreed upon by native speakers.

This P-source condition, however, seems not to be the whole story, as pointed out by Heycock (1994) and Landau (2011). Consider the following examples:

(40) a. *Your car sounds like it needs tuning very badly.*
b. *John looks like he has failed the exam.*

When the speaker listens to the wincing sound of the car's engine, the subject in (40a) can serve as the P-source. However, the speaker can make this statement without experiencing the car's sound directly but hearing the friend's describing the bizarre noises that the car is making. The subject *your car* is not a direct P-source: there is no sensory stimulus to which the CR verb *sounds* is responsive. (40b) can be also uttered with no P-source reading for the subject in a situation where the speaker utters this, looking at the post of the official exam results. In both examples, the speaker has no direct perceptual contact with the referent of the subject.

In addition, one important property we observe in the CR is that the subject in general represents given information. The CR sentence cannot be given in the beginning of any context, or the matrix subject cannot be indefinite:

(41) a. **A lifeguard seems like he or she is dancing across the river.*
b. **A girl seems like she is laughing hysterically.*

All the corpus examples we have found indicate that the matrix subject is definite or generic: no true indefinite subject is found in the CR construction. What this indicates is that the matrix subject functions as the topic or given information, while the remaining predicate serves as comment or new information (see Gundel 1988). We suggest that the matrix subject of the P-source in the CR construction needs to be characterized by the remaining predicate:[10]

(42) CR's Perceptual Characterization Condition (PCC):
The matrix subject of the CR construction, serving as the topic, is "perceptually characterized" by the rest of the utterance.

This characterization condition tells us that the utterance as a whole serves as a characterization of the matrix subject in the CR construction. This characterization condition can easily explain the unacceptability of examples like the following:

(43) a. **Bill appears as if Mary is intelligent.*
(Lappin 1984)
b. **Tina seems like Chris has been baking sticky buns.*
(Asudeh and Toivonen 2012)

The fact that Mary is intelligent does not say any characteristic about Bill. Neither does Chris's baking sticky buns describe any characteristic about Tina.

---

10 Takami (1992) introduces the characterization condition for English pseudo-passives:

(i) a. **I was waited for by Mary.*
b. *I don't like to be waited for.*

The simple fact that Mary is waiting for me does not characterize the property of *me*, but my tendency for not preferring to make someone wait for me can describe a characteristic about me. The same condition can tell the following *tough* construction apart:

(ii) a. **Friends are dangerous to meet in New York.*
b. *New York is dangerous to meet friends in.*

Meeting friends in New York does not tell any characterization property of the friends, but the statement (iib) describes the property or characterization of New York. See Kuno (1987) and Takami (1992) for further discussion of the characterization condition.

The characterization condition PCC also explains why examples like the following extracted from the corpora are natural even though there is no pronominal copy at all:

(44) a. *"The house smells like you've been cooking all day", says Duggan.* (COCA 2006 NEWS)
b. *The entire scene appears as if the Creator himself had wished it to be so.* (COCA 1992 FIC)
c. *Things appear as if you were standing at a window or in front of a view.* (COCA 2004 MAG)

The matrix subject in each case is an overt or inferrable P-source participant which is "characterized" by the event denoted by the embedded clause. For example, in (44a), the house's smell became its characteristic by the event of cooking all day. In (44b), the matrix verb *appear* contributes to the "perceived" nature of the subject's characteristic. That is, the complement clause characterizes the entire scene. Note that the subject *things* in (44c) also denotes a familiar situation around the speaker and hearer. The remaining parts of the utterance characterize the current state of affairs. This explains why we can generate these examples with no pronominal copy in the embedded clause at all.

The position we take is thus that the license of the CR construction does not depend on the pronominal copy condition alone, but rather depends upon the CR's characterization condition PCC too. As long as this interpretive, pragmatic constraint is observed, there is no need to resort to the co-reference constraint (or pronominal copying).

There are still questions that remain to be answered: (a) why the most natural CR examples are those with the coreferential relations with two subjects (b) why there are lexical variations in licensing the absence of the pronominal copy. What we observe is that the matrix subject of the CR is linked to an individual participating in the event described by the embedded clause in the following ranking suggested by Ariel (1990):

(45) subject > specifier of the subject > (prepositional) object > context-provided inferred individual

As argued in Ariel (1990), this hierarchy reflects a "referential accessibility hierarchy" that provides speakers with means to code the accessibility of the referent to the addressee. This explains why the matrix subject preferred to be coindexed with the most accessible individual in the embedded clause in the ranking provided in (45). The corpus data we have collected also support this hierarchy: the

higher frequency cases are those where the matrix subject is coindexed with the subject or its specifier of the highest embedded clause.

4 Lexical properties and theta-role assignments

Given the traditional assumption that the matrix subject is originated from the highest embedded clause's subject, we would expect that the matrix subject receives no theta role from the matrix CR predicate. This position seems to be supported by several phenomena at first glance. However, we suggest that the matrix subject of the CR is ambiguous with respect to the theta-role bearing, as also pointed out by Potsdam and Runner (2001) and others.

Arguments supporting that the matrix subject carries no thematic role can come from the fact that CR predicates, placing no selectional restriction on their subject, allow the expletive *it*, and idiom pieces as the subject. As noted in Potsdam and Runner (2001), the matrix subject can be part of an idiom as illustrated in the following (see Postal 1974; Rothstein 1991 also):

(46) a. *%There seem like there are problems.*
b. *It seems like it's raining harder than it is.*
c. *%The shoe looks like it's on the other foot.*
d. *%Exception seems like it was taken to the recounting of votes.*

The possibility of having the expletives *there*, the weather *it*, and part of the idiom supports the idea that the matrix verbs do not assign a thematic role to the matrix subject.

A further support for the nonthematic role of the matrix subject can be found from naturally occurring examples with no subject at all:

(47) a. *Seems as though I would have to take the first train for England . . .*
(COCA 2001 MAG)
b. *Seems like fishing brings out the best in a man.*
(COHA 1972 MAG)

(48) a. *Feels like I won the lottery.*
(COCA 2006 NEWS)
b. *Sounds like he is preaching exclusion.*
(COCA 2005 NEWS)
c. *Looks like they'll be building another wall.*
(COCA 2007 NEWS)

Our corpus search yields many spoken and written examples where the matrix subject is not realized at all in the CR construction.

The fact that no thematic role is assigned to the matrix subject means that the CR predicate selects the *as-if* clause as its unique semantic argument. That is, the CR verbs (both GCR and PCR) are monadic verbs selecting only one internal argument (see Section 5.2). This monadic treatment of the CR verbs will license examples with the expletive *it* as subject:

(49) a. *It seems like you are ready.*
b. *It sounds like you are misinformed.*

The monadic treatment also provides us with a way to explain the pronominal copy examples:

(50) a. *Prince appears as if **he** bears the weight of the world on his narrow shoulders.*
(COHA 2007 MAG)
b. *. . . it makes the tree appear as if **it** were covered with deep pink blossoms.*
(COHA 1947 NF)

In such examples, the matrix subject and the embedded clause's subject refer to the same individual, motivating the traditional pronominal copy analysis. In the monadic treatment, as long as we have a way of linking the matrix subject to the highest embedded clause's subject, we can keep the supposition that the matrix subject is not assigned a thematic role. In Section 5, we will see how this effect can be achieved without assuming movement operations.

In addition to this monadic case, following Potsdam and Runner (2001) and others, we accept the view that the matrix subject of the CR predicate can be thematic too. In particular, we assume that the subject is thematic in examples with the pronominal copy in a non-subject position, whose data we repeat here:

(51) a. *He appeared as if his heart were broken by her speech.*
b. *Her skin seemed as if she didn't take a bath for years.*
c. *The lawn looked as if someone had brushed it.*
d. *Her apartment sounds like there must be a wonderful view.*

In these examples, the assumed pronominal copy of the matrix subject is not in the embedded clause's subject position. The coreferential NP is in the specifier of the subject or in the object position. In (51d), there is even no coreferential pronoun in the embedded clause at all.

Evidence indicates that in such examples, the matrix subject gets a thematic role from the matrix CR predicate. The first evidence can be observed from coordination data like the following:

(52) a. *His hair* [$_{VP}$ *was blonde*] *and* [$_{VP}$ *looked as though he'd spent a lot of time fixing it*].
(COCA 2005 FIC)
b. *He* [$_{VP}$ *lay down*] *and* [$_{VP}$ *once again appeared like he was never going to get up*].
(COCA 1999 NEWS)

The subject *his hair* in (52a) serves as the subject of the two coordinated VPs too. Since the first VP *was blonde* surely assigns a thematic role (e.g., theme) to its subject, we may assume that the second VP's subject also has a thematic role. The same situation holds in (52b). If the subject of the second VP were non-thematic, it would mean that the subject of the VP *lay down* is also non-thematic. In this example, we have the pronominal copy in the subject, unlike (52a). This in turn means that the subject of the CR is ambiguous: it can be either thematic or non-thematic, supporting Potsdam and Runner's (2001) position.

The second argument is noted by Potsdam and Runner (2001): when the pronominal copy is non-subject, we do not observe the canonical raising properties. For example, no idiom, no funny NP or no PP can serve as the main subject of the CR predicate when there is no pronominal copy in the embedded clause's subject position:

(53) a. **The other foot appears like the shoe is on it*.
b. **Much headway seems like we made it on that problem last night*.
c. **Under the bed seems like an unoriginal place to hide will be it*.

In addition, given the thematic subject position, we then expect it to be linked to the controller of control predicates. This prediction is supported by corpus examples:

(54) a. *He* ***attempted*** [*to sound like he is speaking Chinese*].
(COCA 2010 FIC)
b. *Frank* ***tried*** [*to look like he was wrestling with his conscience*].
(COCA 2007 FIC)

The verbs *attempted* and *tried* are control verbs whose infinitival VP complement is controlled by the matrix subject. Since the matrix subject receives a thematic

role from the matrix verb, we can conclude that the subject of the infinitival VP projected from the CR verb is also thematic.

5 A lexicalist, constraint-based analysis

5.1 Type hierarchy and argument realization

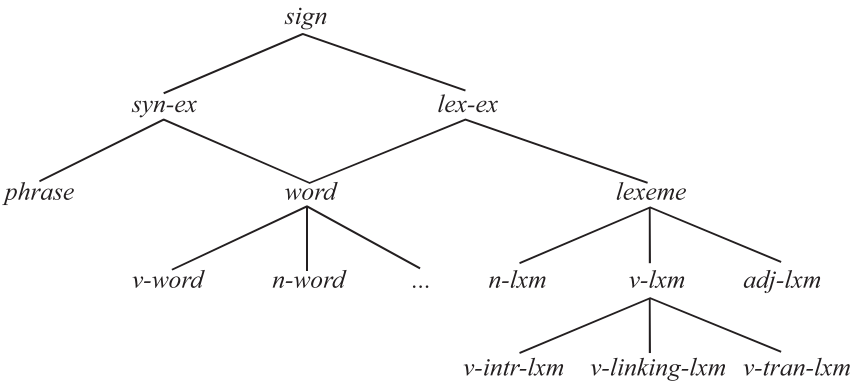
In accounting for the properties of the CR construction we have discussed so far, we adopt the lexicalist theory of HPSG (Head-driven Phrase Structure Grammar). HPSG, a model theoretical grammar framework, conceptualizes language as a system of signs or pairings of form with meanings. In HPSG, the grammar is merely a static description of what signs are in the language along with a basic mechanism for combining the signs which are modeled as typed feature structures. More formally, HPSG consists of a set of types arranged into a multiple inheritance hierarchy, a set of constraints on those types, and an initial symbol. Each type is a partial description of a kind of linguistic object such as a word or a phrase. The types can be used in feature descriptions to describe larger linguistic objects (see, among others, Pollard and Sag 1994, Ginzburg and Sag 2000, Sag et al. 2003, Sag 2012).

As exemplified in (55), the linguistic signs (*sign*) can be classified into syntactic (*syn-ex*) and lexical (*lex-ex*) expressions, and the latter of which in turn has two subtypes, *word* and *lexeme*.[11] Lexemic expressions are abstract proto-word or root-like expressions. The lexemes can be projected into stems and then into words. For example, the lexeme *appear* will give rise to genuine words at syntax such as *appears*, *appeared*, *appearing*, and even *appear*. Only word-level expressions can appear at syntax: lexeme expressions cannot (see Sag et al. 2003: 228, Kim and Sells 2008: 81, Sag 2012: 98). Like other linguistic types, lexemic expressions are organized as a type hierarchy in order to capture generalizations about similar classes. As given in a simplified version of the hierarchy (55), the type *lexeme* has subtypes to which the properties of the supertypes are inherited:[12]

11 The HPSG literature defines the subtypes of *sign* in slightly different ways. The type hierarchy given here is a simplified version of Sag (2012). See Ginzburg and Sag (2000), Sag et al. (2003), Kim et al. (2011), Sag (2012).

12 See Flickinger et al. (1985), Flickinger (1987), Pollard and Sag (1987, 1994), and Sag (2012) for the more detailed inheritance hierarchical structure of the lexicon. In addition, refer to Snider (2005) for a similar classification of the CR predicates and to Goldberg (2006) for the key roles the notion of multiple inheritance hierarchy plays in the Constructional Grammar perspective.

(55)



The inheritance mechanism allows properties of the type *v-lxm* to be inherited to its subtypes, *v-intr-lxm* (*intransitive-lxm*), *v-linking-lxm*, and *v-tran-lxm* (*transitive-lxm*). For example, all the instances of the type *v-lxm* will select at least one argument, as represented in the following feature description:[13]

(56) $\begin{bmatrix} \textit{v-lxm} \\ \text{ARG-ST} \left\langle \text{XP}, \ldots \right\rangle \end{bmatrix}$

The type declaration here ensures that the linguistic expressions belonging to the type *v-lxm* (verb lexeme) have at least one argument XP in the list value of its ARG-ST (argument-structure). This information will be inherited to its subtypes in the hierarchy (55). The inheritance of such a constraint is "default" in the sense that constraints on supertypes affect all instances of subtypes, unless contradicted by some other constraints on a given type (see Lascarides and Copestake 1999, Sag et al. 2003: 229). The default and defeasible constraints allow the system to capture linguistic generalizations as well as idiosyncrasies.

The lexemic expressions will undergo inflectional or derivational processes and be promoted to a *word* level expression (Sag et al. 2003, Kim and Sells 2008, Sag 2012). In these processes, the argument-structure information on the lexemic expressions will be linked to the syntactic valence information or grammatical relations such as subject (SUBJ) and complements (COMPS). The mapping rule,

13 The type of each feature structure is given as italics on the top and the attributes we use here include ARG-ST (argument-structure), COMPS (complements), IND (index), NFORM (noun form), RELN (relation), SUBJ (subject), SEM (semantics), VAL (valence), XARG (external argument). See Sag et al. (2003), Kim and Sells (2008), and Sag (2012) for details.

represented by the Argument Realization Principle, is straightforward in a language with a relatively rigid word order. The basic pattern is that the first element on the argument-structure list is realized as SUBJ and the rest as COMPS (see Manning and Sag 1998; Ginzburg and Sag 2000; and Van Eynde 2005).

(57) Argument Realization Principle (ARP):
The first element on the ARG-ST list is realized as SUBJ, the rest as COMPS in syntax.

This realization is obligatory in English; for example, the three arguments of *put* are realized as subject and complements, with the putter (agent) as subject. The violation of this principle will lead to ungrammatical sentences:

(58) a. *John put the book in the box.*
b. **John put in the box.*
c. **In the box put John the book.*
d. *#The book put John in the box.*

We see that the arguments selected by a lexical head should be all realized as SUBJ and COMPS, which are combined in the notion of VAL (valence) features.[14]

(59)
$$\begin{bmatrix} \textit{v-word} \\ \text{VAL} \begin{bmatrix} \text{SUBJ} \langle \boxed{1}\text{NP} \rangle \\ \text{COMPS} \langle \boxed{2}\text{NP}, \boxed{3}\text{PP} \rangle \end{bmatrix} \\ \text{ARG-ST} \langle \boxed{1}\text{NP}, \boxed{2}\text{NP}, \boxed{3}\text{PP} \rangle \end{bmatrix}$$

This information indicates that the ARG-ST has three elements represented by the numbers. The first element (1) is identified with the element on the SUBJ list while the remaining two (2 and 3) are linked to the COMPS list, observing the ARP (57). The ARP constraint blocks examples like (58c) in which the locative argument is realized as the subject, as shown in (60):

14 The term *valence* (VAL) refers to the number of arguments that a lexical item can combine with, to make a syntactically well-formed sentence. See Sag et al. (2003) and Kim and Sells (2008).

(60) *
$$\begin{bmatrix} \textit{v-word} \\ \text{VAL} \begin{bmatrix} \text{SUBJ} \langle \boxed{3}\text{PP} \rangle \\ \text{COMPS} \langle \boxed{1}\text{NP}, \boxed{2}\text{NP} \rangle \end{bmatrix} \\ \text{ARG-ST} \langle \boxed{1}\text{NP}, \boxed{2}\text{NP}, \boxed{3}\text{PP} \rangle \end{bmatrix}$$

The expression projected from this lexical entry violates the ARP, which requires the first element of ARG-ST be realized as the SUBJ.[15] In the following section, we will see how the hierarchical inheritance mechanism and argument realization constraints are interacting together to predict the complex behavior of the CR constructions.

5.2 CR predicates and argument realizations

As we have discussed earlier, the pronominal copy analysis (with a movement rule) covers only a limited set of data, in particular, data with the pronominal copy in the embedded clause's subject. We have seen that the naturally occurring data extracted from corpora yield numerous examples where the GCR and the PCR construction are used with the pronominal copy in the non-subject position or with no pronominal copy at all in the highest embedded clause. Furthermore, as argued by Potsdam and Runner (2001), movement operations meet challenges in capturing the semantic difference between the assumed source sentence and the output one. Let us consider the scope possibilities in the raising and CR examples:

(61) a. *Two people seem to have won the lottery.*
b. *Two people seem like they have won the lottery.*

The raising sentence (61a) has the following two readings in which the quantifier *two people* has either a narrow or wider scope reading with respect to the predicate *seem*:

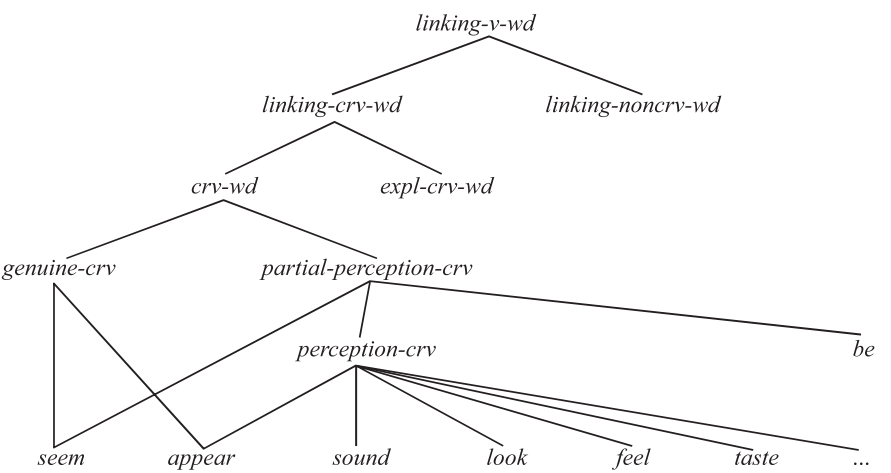
(62) a. *It seems like two people have won the lottery.*
b. *Two people are such that they seem like they have won the lottery.*

15 There can be some unusual argument realizations showing mismatches between the value of the ARG-ST and VAL. See, among others, Manning and Sag (1998), Van Eynde (2005), Kim (2003), Ball (2008), and Müller and Ørsnes (2013).

The two possible readings can be attributed to the fact that the quantifier is generated in the embedded clause's subject position and moved to the surface position (May 1985). However, the CR sentence (61b) does not induce the two readings: it has only the reading (62b) where the quantifier scopes over *seem* as noted by Lappin (1984). If the quantifier *two people* in (61b), leaving a pronominal copy in the embedded clause, were originated from the embedded clause, there would be no reason to disallow the verb *seem* to take scope over the quantifier, as in the typical subject-to-subject raising case in (61a). One plausible analysis to explain this difference, as argued by Potsdam and Runner (2001), seems to be the one assuming that in CR examples like (61b), the quantifier subject is generated in situ, hence allowing only one scope possibility with respect to the predicate *seem* and the quantifier.

Our analysis, supporting such a base-generation for the matrix subject of the CR, starts from the classification of CR predicates (*crv-wd*) as given in the following hierarchical structure:

(63)



The hierarchy reflects cross-cutting generalizations that hold across the types. Let us consider justifications for each type in the hierarchy. Linking verbs (*linking-v-wd*) include those like *seem*, *appear*, *sound*, *feel*, *remain*, *become*, *grow*, *prove*, *turn*, *become*, *be*, and so forth. This type is classified into two groups: those participating in the CR (*linking-crv-wd*) and those not (*linking-noncrv-wd*).[16] That is,

16 Verbs like *remain*, *grow* and *prove* thus will belong to the latter.

verbs participating in the CR construction belong to *linking-crv-wd* which in turn has two subtypes *crv-wd* and *expl-crv-wd* (*expletive-crv-word*). The following demonstrates an exemplar realization of these two types:

(64) a. *Richard seems/appears/sounds/looks/feels/tastes like he is in trouble*. (*crv-wd* examples)
b. *It seems/appears/sounds/looks/feels/tastes like he is in trouble*. (*expl-crv-wd* examples)

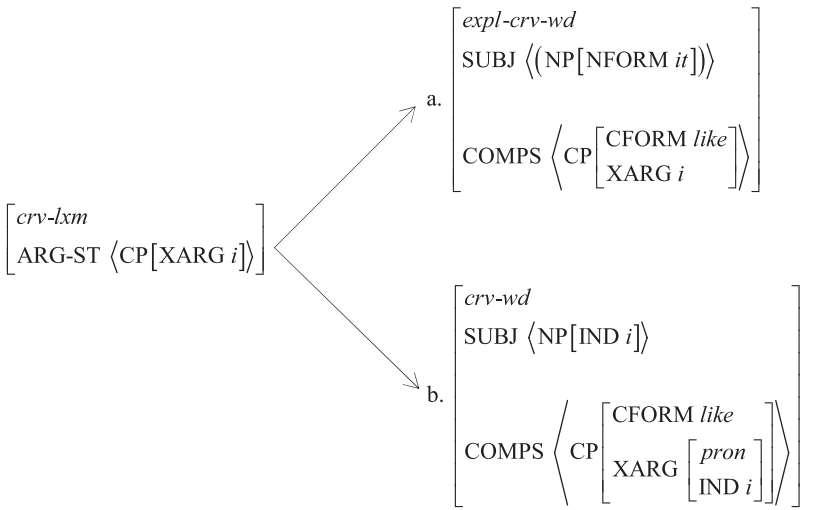
These two types of word are projected from the CR lexeme. One main lexical property that all the CR lexemes have is that they can select one sentential argument, as represented in the following lexemic information:

(65)
$$\begin{bmatrix} \textit{crv-lxm} \\ \text{ARG-ST} \left\langle \text{CP}\left[\text{IND } \textit{s1}\right] \right\rangle \\ \\ \text{SEM} \begin{bmatrix} \text{IND } \textit{s0} \\ \text{FRAMES} \left\langle \begin{bmatrix} \textit{crv-frame} \\ \text{ARG1 } \textit{s1} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

The lexical specification on the lexeme (*crv-lxm*) means that the lexeme requires as its argument one sentential expression (CP) whose index (IND) value denotes a situation (*s1*). The semantic (SEM) information indicates that the lexeme provides a semantic "relation" whose argument (ARG1) is linked with the selected argument CP.[17] The key idiosyncratic, lexical property of the lexeme *crv-lxm* is its argument realization. That is, it can serve as an input to the two different word expressions, as specified in the following:

17 For the detailed semantic representations, see Sag (2012).

(66)



The two realizations here can be interpreted as two lexical rules: the input *crv-lxm* yields two outputs: *expl-crv-wd* and *crv-wd*. As specified here, the lexeme *crv-lxm* can be realized either as the word *expl-crv-wd* selecting the expletive *it* as its syntactic subject or as the word *crv-wd* whose subject is coindexed with the CP's external argument (XARG).[18] As an illustration, consider instances of these two word types:

(67) a.
$$\begin{bmatrix} \textit{expl-crv-wd} \\ \text{FORM } \langle \textit{seem/sound} \rangle \\ \text{SUBJ } \langle (\text{NP}[\textit{it}]) \rangle \\ \text{COMPS } \langle \text{CP}[\text{IND } \textit{s1}] \rangle \end{bmatrix}$$

b.
$$\begin{bmatrix} \textit{crv-wd} \\ \text{FORM } \langle \textit{seem/sound} \rangle \\ \text{SUBJ } \langle \text{NP}[\text{IND } i] \rangle \\ \text{COMPS } \left\langle \text{CP}\begin{bmatrix} \text{XARG } i \\ \text{IND } \textit{s1} \end{bmatrix} \right\rangle \end{bmatrix}$$

As noted in the previous section, these two cases display atypical argument realizations. The lexical entry (67a) introduces the expletive subject that has no link-

18 The external argument (XARG) of a clause (CP or S) is its subject. The feature XARG is a kind of head feature originated from the verb in question and percolating up to the top of the clause. As noted by Sag (2012), both tag questions and CR constructions motivate the introduction of the feature XARG. For example, in tag questions like *They left, didn't they/*she?*, the XARG values of the main clause and the tag must be compatible. In CR constructions the XARG value makes information about the clause's subject accessible to the verb selecting the clause.

ing element in the ARG-ST, while the complement is mapped from the sole argument. This word level realization will project examples like the following:

(68) a. *It seems/appears/sounds like John is smart.*
b. *It seems/appears/sounds like there is no hope.*
c. *It seems/appears/sounds like it would rain tomorrow.*

Since the sole argument of the predicate is realized as the CP sentential complement (COMPS), there is no element to be mapped onto the subject. To satisfy the Extended Projection Principle (Chomsky 1982) in English in a sense, the verb looks for its subject with no thematic role, the expletive *it*. Note that there is no restriction on the type of the subject of the embedded clause.

The lexical entry (66a) also indicates that the expletive subject is optional here. No overt expression needs to be realized at syntax, as already witnessed from examples with no subject, whose data we repeat here:[19]

(69) a. *Seems like fishing brings out the best in a man.*
b. *Sounds like he is preaching exclusion.*

The present system licenses such examples too: since there is no external argument expression to be realized as the subject, the subject can either be the expletive *it* or be empty.[20]

Now consider the type *crv-wd* in (67b), the alternative argument realization of the type *crv-lxm*. What we can see here is that the sentential argument is realized as an internal CP complement (COMPS). Meanwhile, the subject has no linking element in the ARG-ST, but its index value is the same as that of the CP's external argument (XARG). This coindexation relation will make it unnecessary to assign a thematic role to the matrix subject, introducing an additional element in the ARG-ST. In fact, this additional way of realizing the subject makes the copy construction *sui generis*, eventually making it possible to generate the typical pronominal copying examples:

(70) a. *John$_i$ seems/appears like **he**$_i$ has to be right all the time.*
b. *John$_i$ looks/sounds like **he**$_i$ checks the mirror every morning.*

19 Non-CR verbs do not license the absence of the subject as in **(It) remains possible that bad weather could tear more holes in the tanker's hull.*

20 The copula verb *be* has additional idiosyncrasies: its subject cannot be omitted at all, which can be specified as its own lexical idiosyncrasy.

The matrix subject is coreferential with the embedded clause's subject, inducing the effect of having a pronominal copy in the latter position. This in turn means that the subject's IND value needs to be identical with the index value of the CP's external argument, ruling out examples like the following:[21]

(71) a. ***There** seemed like **it** was raining.*
b. ***There** seemed like **he** had a very cold side.*

The argument realization of the *crv-wd* in (67b) thus explains the agreement relation between the matrix subject and the embedded clause's subject. Note that the CP's external argument needs to be a pronoun (*pron*), ruling out examples like the following:

(72) a. **The lifeguards$_i$ seem like the lifeguards$_i$ are dancing across the water.*
b. **They$_i$ seem like the lifeguards$_i$ are dancing across the water.*

This condition reflects the property that the two subjects are in a type of anaphoric relation (see Asudeh and Toivonen 2012 for a detailed LFG approach).

Now let's consider the subtypes of the *crv-wd* in the hierarchy (63) together with the membership of its three main subtypes:

(73) a. *genuine-crv: seem, appear*
b. *perception-crv: appear, sound, look, feel, taste*
c. *partial-perception-crv: seem, appear, sound, look, feel, taste, be*

Genuine CR verbs *seem* and *appear* semantically select one sentential argument only, and are preferred to have the pronominal copy in the highest embedded clause's subject.[22] Perception verbs (*perception-crv*) are related to typical linking verbs selecting two arguments each of which is realized as the subject (SUBJ) and complement (COMPS) and linked to its own semantic argument. This is a canonical pattern of argument realization for linking verbs in English (see Kim and Sells 2008):

21 Note that the analysis allows examples like *It seemed like there is a very cold side to him*, as we discussed in Section 2. This example is not a CR sentence, but a base-generated one with the expletive *it* as the matrix subject. See examples in (68) also.

22 Slight modifications of the hierarchy may provide ways to reflect (dialectal or lexical) variations. For example, if a dialect disallows only *seem* to have the non-subject pronominal copy, we can remove the type *partial-perception-crv*.

(74) a. *He felt so guilty and angry about the whole issue.*
b. *He remained a formidable opponent.*

These linking verbs *felt* and *remained* semantically require two arguments, each of which is mapped onto the subject and predicative complement. The perception CR verbs also belong to these linking verbs (or projected from these linking verbs) with more specific constraints, as represented in the following:

(75)
$$\left[\begin{array}{l} \textit{perception-crv} \\ \text{VAL}\left[\begin{array}{l} \text{SUBJ}\left\langle \boxed{1}\text{NP}\left[\text{IND } i\right]\right\rangle \\ \text{COMPS}\left\langle \boxed{2}\text{CP}\left[\text{IND } s1\right]\right\rangle \end{array}\right] \\ \text{ARG-ST}\ \left\langle \boxed{1},\ \boxed{2}\right\rangle \\ \text{SEM}\left[\begin{array}{l} \text{IND } s0 \\ \text{FRAMES}\left\langle \left[\begin{array}{l} \textit{crv-relation} \\ \text{ARG1 } i \\ \text{ARG2 } s1 \end{array}\right]\right\rangle \end{array}\right] \end{array}\right]$$

The lexical information specifies that the perception CR verb selects a subject NP and a sentential complement, each linked to the corresponding semantic argument (ARG1 and ARG2). There is thus an isomorphic mapping relation between the arguments (ARG-ST) and the valence features (VAL) and also between the arguments and the semantic arguments (SEM).

Note that this lexical information requires no pronominal copy in the embedded clause's subject. That is, we can expect variations in the co-referred NP whose data we repeat here:

(76) a. *He appeared as if his heart were broken by her speech.*
(COHA 1828 FIC)
b. *The lawn looked as if someone had brushed it.*
(COCA 1993 MAG)
c. *. . . , till the forest appears as if a tornado had passed over it, . . .*
(COHA 1850 MAG)
d. *You sound as if the man has no choice in the matter.*
(COCA 1992 FIC)

In all these examples, there is no pronominal copy in the embedded clause's subject position. Given that the subject of the CR predicate can be thematic too or

gets a thematic role from the matrix predicate, nothing is wrong not to have a coreferential expression in the embedded clause, as long as the characterization condition in (42) is met.[23]

We now see that the lexeme *crv-lxm* has thus two options to be realized as a word level expression: one with the expletive subject (*expl-crv-wd*) and the other with the pronominal copy in the embedded clause's subject (*crv-wd*). In both realizations of the CR lexeme, we thus have an incongruous mapping between syntactic and semantic argument. There is one semantic argument with two syntactic arguments.

Now consider the motivation to introduce the type *partial-perception-crv*. The membership difference from the type *perception-crv* is the verb *seem* and *be*. We can observe that the verb *appear* can assign a thematic role to the subject when it is interpreted such as 'put on the appearance of', in particular when the verb is used as a non-raising verb. Consider the following set of data:

(77) a. *She appeared self-assured.*
b. *He is anxious to appear a gentleman.*
c. *It appears to have rained more at night then during the day.*

The verb *appear* in (77a) and (77b) is interpreted similar to *look*, selecting not one but two arguments. This is different from its use as a raising verb in (77c) where it assigns no thematic role to the expletive subject. Note also the difference between *seem* and *appear*: the verb *seem* is semantically most transparent verb with no clear perceptive meaning and it cannot be used as a pure intransitive:

(78) a. *John appeared.*
b. **John seemed.*

What these observations imply is that the verb *seem* in the CR is the least likely verb that assigns a thematic role to the subject, as seen from the following (cf. Asudeh and Toivonen 2012):

(79) a. **Jane seems like there would be no issues.*
b. *?Jane appears like there would be no issues.*
c. *Jane sounds like there would be no issues.*

23 Within the present system, the characterization condition can be formalized as a contextual or pragmatic condition on the construction projected from the expression *crv-wd*. We leave open its formalization here.

We can attribute this difference between *seem* and *sound* (or even *appear*) to the weaker semantic properties of the verb *seem*, assigning no thematic role to the matrix subject in the CR.

However, note that given a proper context supporting the meaning of *seem*, (79a) can be turned into an acceptable one as in (80):

(80) *When I talked to her, Jane seemed like there would be no issues.*

The verb *seem* is the most bleached perceptual source verb, imposing no substantive restriction on the sensory or mental mode through which the stimulus is perceived. Contextual cues can enrich its semantics.

Also consider the following with more context:

(81) a. *Studying a foreign language is much like travel in a foreign land: interesting, exciting, but at the same time often intimidating, difficult, discomfiting. For me, studying Yiddish seemed as though I were traveling, instead, through the streets of a long-forgotten hometown.*
(COCA 2000 ACAD)

b. *Tom Guenther, a banking regulator who flew US Airways on Monday morning to Washington from Charlotte, said flight attendants were polite and never mentioned the bankruptcy filing. "They just seemed like it was any other day," he said.*
(COCA 2002 NEWS)

As given here, the context can coerce the verb into a semantically substantive verb so that it can assign a thematic role to its subject. This being noted, we cannot segregate *seem* from the other perceptive verbs, leading us to posit the supertype *partial-perception-crv*.

One remaining case we need to consider is the CR with the copula verb *be*, which can be used as the subtype of *partial-perception-crv*. The copula verb *be* behaves similarly to the perception verbs, but in different way. Observe that the copula verb also can appear in the CR construction as instances of the *expl-crv-wd* or as instances of *crv-wd*:

(82) a. *It is as if people are having a difficult time.*
(COCA 2009 SPOK)

b. *It was as though some hidden battle were going on constantly.*
(COCA 1996 FIC)

c. *It was like he was desperately telling me his story.*
(COCA 2011 NEWS)

(83) a. *The actual shock was as if someone hit me with their whole strength with a club.*
(COCA 1999 MAG)
b. *The interior is as if you are in between World War I and World War II.*
(COCA 2001 MAG)
c. *Her tone is as if they were picking up a conversation suspended some time ago.*
(COCA 2000 FIC)

The semantic content of the copula verb in (83) is inferred from the context, like the CR verbs *seem*, *look*, or *sound*. For (83c), the subject *tone* can be easily linked to our auditory perception verb *sound* and the complement clause characterizes this perception source. The context thus allows the matrix subject to serve as the P-source in the eventualities, observing the characterization condition, PCC. For example in (83a), the shock represents a characteristic of someone's hitting me while in (83b) the interior characterizes the situation *you* went through. This implies that a variant of the copular verb can be used as a CR verb, with proper contextual support that can fulfill the characterization condition.

In sum, the CR lexeme selects one sentential argument, but this lexeme can be realized into two different types of word: one requiring the expletive *it* as its subject and the other placing an agreement relation between the matrix subject and the external argument of the CP complement. The former projects CR sentences with the expletive subject while the latter generates the so-called subject pronominal copy examples. These two realizations apply to all the typical CR verbs including *seem*, *appear* and perception verbs like *look*, *feel*, *taste*. The perception verbs (or partially perception verbs including *seem* and *be*) have an alternative lexical constraint, realized from the typical linking verbs selecting two arguments. For these, there is no need for the coindexation relation between the subject and an expression within the embedded clause selected by the CR verb. They serve as a subtype of the CR verbs (*crv-wd* in observing the characterization condition).

6 Conclusion

We have seen that the CR construction, which can be classified into GCR (genuine CR) and PCR (perception CR) type, raises several challenging issues to generative grammar, in particular, to movement analyses. The traditional analysis has followed the assumption that the embedded clause's subject is raised to the matrix subject, leaving behind a pronominal copy in its place. This raises both empirical

and theoretical issues. Theoretically, this runs against the traditional movement assumption that only non-cased expressions can be moved to a case-assigned position. Empirically, we have seen that the pronominal copy in the embedded clause's subject position covers only part of the data. There is a great variation in the pronominal copy.

In this paper, we have suggested that the complexity and variations of the CR construction have to do with the tight interactions among hierarchically organized lexicon, argument realizations, and pragmatic constraints. Both the GCR and the PCR type can select only one internal argument which is at syntax mapped onto the complement. This lexemic information can be realized into two different ways with respect to the subject value: one with the expletive or no subject and the other with the pronominal copy. The perception CR verbs are different in that they select two arguments with congruous mapping from argument structure to syntax as well as semantics. The CR construction, both with genuine CR verbs and perception verbs, needs to observe the "characterization" condition, allowing us to override the pronominal copy constraint. Once we have tight interactions among lexical properties, syntax, and semantics, the idiosyncratic properties of the construction in question fall out in a clear, simple way.

Acknowledgments: An earlier version of this paper was presented at the 1st World Congress of Scholars of English Linguistics at Seoul, Korea, in June 29–31, 2012. I thank the audiences for questions and suggestions. My thanks also go to Sung-Ho Ahn, Mark Davies, Jungsoo Kim, Okgi Kim, Sun-Woong Kim, Stefan Müller, Chongwon Park, and Peter Sells for helpful discussion and feedback. The questions and criticism from the anonymous reviewers of this journal also helped a lot to improve the quality of this paper. Any remaining errors and misinterpretations are of course mine.

References

Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London & New York: Routledge.

Asudeh, Ash. 2002. Richard III. In Mary Andronis, Erin Debenport, Anne Pycha, & Keiko Yoshimura (eds.), *Papers from the 38th regional meeting of the Chicago Linguistic Society (CLS 38)*, 31–46. Chicago, IL: Chicago Linguistic Society.

Asudeh, Ash, & Ida Toivonen. 2012. Copy raising and perception. *Natural Language and Linguistic Theory* 30(2). 321–380.

Ball, Douglas. 2008. *Clause structure and argument realization in Tongan*. Stanford, CA: Stanford University dissertation.

Bender, Emily, & Dan Flickinger. 1999. Diachronic evidence for extended argument structure. In Gosse Bouma, Erhard Hinrichs, Geert-Jan Kruij & Richard Oehrle (eds.), *Constraints*

and resources in natural language syntax and semantics, 3–19. Stanford, CA: CLSI Publications.
Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.
Davies, Mark. 2009. The 385+ million word corpus of contemporary American English (1990–2008+): Design, Architecture, & Linguistic Insights. *International Journal of Corpus Linguistics* 14. 159–190.
Davies, Mark. 2012. The 400 million word corpus of historical American English (1810–2009). In *English Historical Linguistics* 2010, 217–250. Amsterdam & Philadelphia: John Benjamins.
Flickinger, Dan. 1987. *Lexical rules in the hierarchical lexicon*. Stanford, CA: Stanford University dissertation.
Flickinger, Dan, Carl Pollard, & Tom Wasow. 1985. Structure-sharing in lexical representation. In William C. Mann (ed.), *Proceedings of the twenty-third annual meeting of the association for computational linguistics*, 262–267. Chicago, IL: Association for Computational Linguistics.
Fuji, Tomohiro. 2005. Cycle, linearization of chains and multiple case checking. In Sylvia Blaho, Luis Vicente & Erik Schoorlemmer (eds.), *Proceedings of ConSole 18*, 39–65. Leiden: Leiden University.
Ginzburg, Jonathan & Ivan Sag. 2000. *Interrogative investigations: The form, meaning, and use of English interrogatives*. Stanford, CA: CSLI Publications.
Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
Gundel, Jeanette. 1988. Universals of topic-comment structure. In Michael Hammond, Edith Moravczik & Jessica Wirth (eds.), *Studies in syntactic typology*, 209–239. Amsterdam & Philadelphia: John Benjamins.
Heycock, Caroline. 1994. *Layers of predication: The non-lexical syntax of clauses*. New York: Garland Publishing Co.
Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
Huang, C.-T. James. 1982. *Logical relations in Chinese and the theory of grammar*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
Kaplan-Myrth, Andrew. 2000. *The movement rule formerly known as Richard*. Ms. Yale University.
Kim, Jong-Bok. 2003. English locative inversion: A constraint-based perspective. *Ene* [Korean Journal of Linguistics] 28(2). 207–235.
Kim, Jong-Bok, & Peter Sells. 2008. *English syntax: An introduction*. Stanford, CA: CSLI Publications.
Kim, Jong-Bok, Jaehyung Yang, Sanghoun Song, & Francis Bond. 2011. Deep processing of Korean and the development of the Korean resource grammar. *Linguistic Research* 28(3). 635–672.
Kuno, Susumu. 1987. *Functional syntax: Anaphora, discourse, and empathy*. Chicago, IL: University of Chicago Press.
Landau, Idan. 2009. This construction looks like a copy is optional. *Linguistic Inquiry* 40(2). 343–346.
Landau, Idan. 2011. Predication vs. aboutness in copy raising. *Natural Language and Linguistic Theory* 29. 779–813.
Lappin, Shalom. 1984. Predication and raising. In Charles Jones & Peter Sells (eds.), *Proceedings of NELS* 14, 236–252. Amherst, Ma: GLSA.

Lascarides, Alex & Ann Copestake. 1999. Default representation in constraint-based frameworks. *Computational Linguistics* 25(1). 55–105.
Manning, Chris & Ivan A. Sag. 1998. Argument structure, valence, and binding. *Nordic Journal of Linguistics* 21. 107–144.
May, Robert. 1985. *Logical form: Its structure and derivation*. MIT Press.
Moore, John. 1998. Turkish copy-raising and A-chain locality. *Natural Language and Linguistic Theory* 16. 149–189.
Müller, Stefan & Bjarne Ørsnes. 2013. Passive in Danish, English, and German. To appear in *Proceedings of HPSG Conference*. CSLI Publications.
Pollard, Carl & Ivan A. Sag. 1987. *Information-based syntax and semantics*. Stanford, CA: CSLI Publications.
Pollard, Carl & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago, IL & London: The University of Chicago Press.
Postal, Paul. 1974. *On raising*. Cambridge, MA: MIT Press.
Postal, Paul M. & Geoffrey K. Pullum. 1988. Expletive noun phrases in subcategorized positions. *Linguistic Inquiry* 19, 635–670.
Potsdam, Eric, & Jeffrey Runner. 2001. Richard returns: Copy raising and its implications. In Mary Andronis, Chris Ball, Heidi Elston, & Sylvain Neuvel (eds.), *Papers from the 37th regional meeting of the Chicago Linguistic Society (CLS 37)*, 453–468. Chicago, IL: Chicago Linguistic Society.
Radford, Andrew. 2010. *An introduction to English sentence structure*. Cambridge University Press.
Rezac, Milan. 2004. *Elements of cyclic syntax: Agree and merge*. University of Toronto dissertation.
Rogers, Andy. 1971. Three kinds of physical perception verbs. In *Papers from the 7th regional meeting of the Chicago Linguistic Society* (*CLS 7)*, 206–222. Chicago, IL: Chicago Linguistic Society.
Rogers, Andy. 1972. Another look at flip perception verbs. In *Papers from the 8th regional meeting of the Chicago Linguistic Society* (*CLS 8)*, 303–315. Chicago, IL: Chicago Linguistic Society.
Rogers, Andy. 1973. *Physical perception verbs in English: A study in lexical relatedness*. Los Angeles, CA: University of California dissertation.
Rothstein, Susan. 1991. Syntactic licensing and subcategorization. In Susan Rothstein (ed.), *Perspectives on phrase structure: Heads and licensing, syntax and semantics* 25, 139–157. New York: Academic Press.
Sag, Ivan A. 2012. Sign-based construction grammar: An informal synopsis. In Hans Boas & Ivan Sag (eds.), *Sign-Based Construction Grammar*. Stanford: CSLI Publications.
Sag, Ivan A, Thomas Wasow, & Emily Bender. 2003. *Syntactic theory: A formal introduction*. CSLI Publications.
Snider, Neal. 2005. *English copy raising*. Manuscript. Stanford University.
Takami, Kenichi. 1992. *Preposition stranding: From syntactic to functional analyses*. Berlin & New York: Mouton de Gruyter.
Ura, Hiroyuki. 1998. Checking, economy and copy-raising in Igbo. *Linguistic Analysis* 28. 67–88.
Van Eynde, Frank. 2005. Argument realization in Dutch. In Ton van der Wouden, Michaela Poß, Hilke Reckman & Crit Cremers (eds.), *Computational linguistics in the Netherlands 2004: Selected papers from the fifteenth CLIN meeting, December 17, Leiden Centre for Linguistics*, 89–103. Utrecht: LOT.