Topic, Focus, and Syntactic Representations

Caroline Heycock and Anthony Kroch
University of Edinburgh and University of Pennsylvania

1. Introduction

Copular sentences like (1) may appear straightforward; they are certainly common.

(1)   The person I most wanted to meet was Tom Lehrer.

Particularly since the work of Higgins 1979, however, it has become clear that sentences like (1), together with pseudoclefts like (2), to which they are intimately related, pose challenging questions to our understanding of linguistic representations.

(2)   What I was hoping for was his autograph.

In this paper we will indicate briefly what is at issue, discuss in passing some recent contributions to understanding the problem, and then concentrate on one aspect that we believe is essential to understanding what is happening: the information structure of such sentences. This aspect of specificational sentences was argued to be important in Heycock and Kroch 1999, but only partially integrated into the analysis presented there; in this paper we attempt to go further in developing an account in which the role of information structure is central.

2. Setting out the problem: specificational sentences

It is commonly assumed that at least in some sentences the copula *be* is not an argument-taking verb, but rather an identity operator on predicates (but see Rothstein 2001 for a different view). On this view the "basic" predicate in the main clause in (3a) is the noun phrase *a nice woman*, just as it is in the small clause in (3b):

(3)   a.   Joan is a nice woman.
        b.   I consider [$_{SC}$ Joan a nice woman]

Slightly more controversial is the claim/assumption that a definite noun phrase can also constitute a predicate, so that again in (4a) the copula contributes nothing to the meaning beyond the tense information that it carries by virtue of its inflection, as suggested by the existence of Small Clauses as in (4b):

(4)   a.   Joan is the best singer in the country.
       b.   I consider [$_{SC}$ Joan the best singer in the country].

We will at least for the moment assume that the examples in (3a) and (4a) should indeed be grouped together, as examples of **canonical copular constructions**.

    More serious controversy arises when we consider cases like (5):

(5)   a.   The best singer in the country is Joan.
       b.   Her goal is total world domination.

These are what Higgins called **specificational** copular sentences, and which have more recently been termed **inverse** copular constructions (see in particular Moro 1997). Characteristically, the noun phrase after the copula appears to be used referentially rather than attributively or predicatively (it picks out a specific individual); also, such sentences do not have an acceptable small clause variant that can occur as a complement to *consider*:

(6)   a.   *I consider the best singer in the country Joan.
       b.   *I consider her goal total world domination.

    Higgins further claimed that sentences like those in (7) fall into the category of specificational constructions:

(7)   a.   What I want is that book over there/you.
       b.   What Joan is is a genius/extraordinarily intelligent.
       c.   What I *don't* have is any bread.

The examples in (7) are **specificational pseudoclefts**. It is crucial to note that these are to be distinguished from the superficially similar examples in (8):

(8)   a.   What I want is unobtainable.
       b.   What Joan$_i$ is is important to her$_i$.
       c.   What I don't know would fill many books.

The sentences in (8) are just straightforward sentences (canonical copular constructions, in the case of (a) and (b)) the subject of which happens to be a free relative. So they are interpeted very much as simpler cases like (9):

(9)   a.   Fame is unobtainable.

    b.   Joan$_i$'s status is important to her$_i$.
    c.   My faux pas would fill many books.

The interpretation of specificational pseudoclefts like those in (7), however, is quite different. For example, in (7b) *being intelligent* is not predicated of Joan's status, in the way that *being important to her* is predicated of Joan's status in (8b); rather, it is predicated of Joan. Thus the sentences in (7) are paraphrasable as in (10):

(10)   a.   I want that book over there/you.
           b.   Joan is a genius/extraordinarily intelligent.
           c.   I don't have any bread.

    The reason that specificational pseudoclefts have proved of such interest to syntactians is that they seem to exhibit the same patterns with respect to various (supposedly) syntactic constraints as their simple sentence paraphrases. Thus for example a reflexive is licensed in (11a), coreference is prohibited in (12a), and a Negative Polarity Item is licensed in (7c), repeated here as (13a). This appears to parallel exactly the pattern in their paraphrases in the (b) examples, but is in contrast to the non-specificational examples in the (c) examples:

(11)   a.   What she$_i$ is is proud of herself$_i$.
           b.   She$_i$ is proud of herself$_i$.
           c.   *What she is$_i$ shocked herself$_i$.

(12)   a.   *What he$_i$ claimed was that Ian$_i$ was innocent.
           b.   *He$_i$ claimed that Ian$_i$ was innocent.
           c.   What he$_i$ claimed was typical of Ian$_i$.

(13)   a.   What I don't have is any bread.
           b.   I don't have any bread.
           c.   *What I don't have surprised anyone.

The contrast between the (b) and the (c) examples is exactly as predicted by standard syntactic theories governing the licensing of anaphors, the restrictions on coreference, and the licensing of NPIs: in the (b) examples the antecedent or licensor c-commands the anaphor/R-expression/NPI, while in the (c) examples it doesn't, as it is buried inside the free relative. The startling fact is that in the (a) examples the lack of c-command doesn't seem to matter.[1] This property of pseudoclefts is called **connectivity** or **connectedness**; we have used the latter term in previous work, but to avoid terminological

---

1.   The examples given so far have been unambiguously specificational or predicational, this is however not always the case. For example, (i) is ambiguous between the two readings:

overlap with a completely unrelated concept developed by Richard Kayne, here we will refer to the unexpected pattern of licensing as **connectivity** (effects).

Connectivity is why specificational sentences are interesting. It seems to suggest either that the abstract syntactic representations these constraints are stated over are much more different from the surface string that one might otherwise suppose, or that the relevant constraints are operating on a different type of representation altogether (essentially, that they are semantic, rather than syntactic, constraints).

3. Some equatives are more equal than others

In previous work (Heycock and Kroch 1999) we have argued that specificational pseudoclefts are a type of equative. The equative analysis yields a number of nice results, and has been argued for by others also, including Sharvit 1997, 1999, and Schlenker 2001. But it also has some shortcomings. In particular, it predicts a symmetry which goes against our intuitions about these sentences, and which is called into question by some more subtle properties of copular sentences.

3.1. Asymmetries in connectivity

First, as hinted already in Higgins 1979, and discussed in some detail in den Dikken, Meinunger, and Wilder 2000, the crucial **connectivity** properties of specificational pseudoclefts are affected by the order of the two arguments in the sentence, which is not expected if specificational sentences are symmetrical equatives. In particular, Negative Polarity Items can appear in **canonical specificational pseudoclefts** (those in which the first element is the free relative, like (14a)) and be licensed by a negative within the free relative; but this is not possible in **reverse pseudoclefts** like (14b):

(14)   They bought a lot of textbooks, but ...
     a.   what they didn't buy was any novels.
     b.   *any novels was/were what they didn't buy.

Another difference is that there are no "Antireconstruction" effects in canonical pseudoclefts like (15a); but there are in reverse pseudoclefts like (15b):

(15)   a.   *What he$_i$ has always claimed is that Cain$_i$ is innocent.
     b.   That Cain$_i$ is innocent is what he$_i$ has always claimed.

(i)   What Joy is is unimportant.

This could either mean that Joy's status is unimportant, or that Joy is unimportant. The latter is the specificational reading.

These properties of reverse pseudoclefts will be discussed in greater detail in Section 7.2—the facts are actually more complex than appears here, and are also not quite as presented in den Dikken, Meinunger, and Wilder 2000.

3.2. Pronominalisation in specificational sentences

There are sentences which, it appears, must be given an equative analysis, under anyone's assumptions. One obvious class consists of tautologies like those in (16):

(16)   a.   War is war.
       b.   Happy is happy.
       c.   Joan is Joan.

And presumably copular sentences of the following type are also equatives:

(17)   a.   To live is to suffer.
       b.   England is not Scotland.

There are good reasons, which we do not discuss here for lack of space, to think that the equative reading is independent of the copula. A possible analysis of equative semantics, therefore, is to assume the availability of Partee's *ident* typeshifting operator, which both lifts the type of an $< e >$ type argument and introduces the equative meaning. Thus, the result of letting *ident* apply to the "basic" *e*-type meaning of a proper noun like *Joan* is as in (18b).

(18)   a.   Joan is Joan
       b. $(\lambda \, x)(x = [\![\text{Joan}_{<e>}]\!])$

For reasons discussed in Heycock and Kroch 1999, we continue to assume that English at least does not allow a predicate in subject position to combine with a following argument; thus *ident* must only be able to apply to the post-copular DP. An alternative possibility is to adopt the proposal of Graff 2001, according to which definites always have a "predicate-type semantic value."

In either case, we now have a semantics for canonical copular sentences with definite noun phrases in the predicate nominal position, as in (19):

(19)   a.   Mary was the winner.
       b.   If I were you . . .
       c.   My cousin is his uncle.

But at this point we discover a difference between these sentences and clearly specificational sentences like (20a,b):

(20)   a.   The culprit is you.
        b.   ?Who she met was John.

The difference is that a specificational sentence following on from (20a,b) cannot refer back to the subject noun phrase with a gendered pronoun, but must use the singular neuter *it*:

(21)   a.   The culprit isn't you, *he/*she/it is Fiona.
        b.   Who she met was John? I don't think so: *he/*she/it was Bill.

This behaviour is quite unexpected under an equative analysis, as pointed out in Büring 1998, Mikkelsen 2002.

One solution that suggests itself is that the first noun phrase in a specificational sentence necessarily denotes an individual concept, whose denotation can depend on at least some temporal index, and that this is what forces the use of the neuter pronoun. But this predicts a much wider use of the neuter pronoun than we actually find; consider for example (22):

(22)   In the early days of the Church the pope was poor, but in the eighteenth and nineteenth century he/*it was typically very rich indeed.

Such examples involve personal pronouns which do not denote individuals, but rather something which can depend on at least some temporal index. But of course the neuter pronoun is impossible in (22), where the postcopular phrase is not nominal and the relevant clause cannot be specificational. Note the minimal contrast with (23); since the indefinite can denote an individual or a property both the neuter and the masculine pronoun are possible (with a subtle change in meaning)

(23)   …… it/he was typically a very rich man

Essentially the same facts hold in German, as shown in Büring 1998. So it seems that non-neuter pronouns can have some kind of "individual concept" reading—but they are still excluded as the subjects of specificational sentences.

Büring's own solution is that the *it* is the expletive found in cleft sentences, so that the second sentence in (21) is really an elliptical cleft:

(24)   The culprit isn't you, it's Fiona ~~who is the culprit.~~

This solution looks promising, but at least in English it cannot be quite correct. As Mikkelsen (2002) points out, the neuter pronoun can occur as the **tag** to a specificational question:

(25)   a.   Our real problem is you, isn't it?

b.   The culprit turned out to be you, didn't it?

On standard assumptions the pronoun in the tag is anaphoric to the subject of the clause (here *our real problem* and *the culprit*). If we somehow allow ellipsis of an entire cleft here, as in (26a), we will have to explain why it is excluded in other examples like (26b):

(26)   a.   Our real problem here is YOU, isn't it ~~you that is the real problem~~?
        b.   *I bought THAT BOOK, isn't it ~~that book that I bought~~

Further, in English *it*-clefts are perfectly possible with plural foci; the agreement on the copula is invariably singular:

(27)   a.   It was/*were John and Bill who won.
        b.   It was/*were John and Bill who were the best candidates.

The analysis in terms of an elliptical *it*-cleft then predicts that (28a) should be grammatical. But this is clearly a wrong prediction; the contrast with a tag question actually based on a cleft is sharp:

(28)   a.   *The best candidates were John and Bill, wasn't it ~~John and Bill who were the best candidates~~?
        b.   It was John and Bill who were the best candidates, wasn't it?

Thus, while it is quite likely that some English sentences of the form *it–be–DP* are reduced clefts, some are not, and the neuter pronoun does seem to be anaphoric in some way—but to what?

4. Pseudoclefts as "self-answering questions"

In recent work by den Dikken, Meinunger, and Wilder 2000, Schlenker 2001, Romero 2001 it has been proposed that specificational sentences are "self-answering questions." Though we are glossing over some serious differences in the individual proposals, the essential idea is that the initial phrase in a specificational sentence (in the case of a pseudocleft, this is the free relative) is a concealed question (or rather, a "concealed answer") similar although not identical to the type found in the complement to verbs like *know*.

(29)   I don't know the time.

There are various interesting issues that arise that we do not have space to go into here. One is evidently the *necessity* for concealment. DPs that have a question denotation typically appear as the complement to verbs that can also take regular interrogative CP complements, but this is not the case with

English pseudoclefts: unambiguous interrogative CPs are *ungrammatical* in the initial position (contra den Dikken, Meinunger, and Wilder 2000). We limit ourselves here to pointing out that if (30a) is a specificational sentence, with a concealed question denotation for the intial phrase (by hypothesis), then presumably (30b) is an equation of two concealed questions:

(30) a. My goal is total world domination.
   b. Your goal is my goal.

The relevance of this will become clearer when we consider a facet of specificational sentences that has often been neglected: their information structure.

5. Information structural asymmetry in specificational sentences

It has been pointed out by many people that specificational sentences have a fixed information structure. As an illustration, consider the examples in (31)–(32), from Heycock 1994. The same canonical order copular sentence can be used felicitously in both (31) and (32):

(31) A: Who was the culprit? (John or Bill?)
   B: JOHN was the culprit.

(32) A: What was John? (Was John the culprit or the victim?)
   B: John was THE CULPRIT.

In contrast, the specificational sentence only allows the postcopular DP to be the focus:[2]

(33) A: Who was the culprit? (John or Bill?)

---

2. As pointed out in Heycock 1994, the B sentence in (34) is grammatical, but only in a different context, one in which the sentence is not taken specificationally:

(i) A: So I met the culprit and the victim both. But I still don't know which one was John.
  B: THE CULPRIT was John.

As evidence that B's answer is not a specificational sentence, note that corresponding to "the culprit" we can get a gendered, not a neuter pronoun:

(ii) B: THE CULPRIT, HE/*IT's John.

Further, in specificational sentences it is possible to get an accentual peak in the precopular phrase *if there is also an accentual peak in the postcopular phrase* (in fact this was the most common pattern reported for pseudoclefts in Delin 1989); this can be analysed as a contrastive topic followed by a contrastive focus:

(iii) The CULPRIT was JOHN, the VICTIM was JOAN.

      B: The culprit was JOHN.

(34)   A: What was John? (Was John the culprit or the victim?)
          B: *THE CULPRIT was John.

Further confirmation of the fixed information structure of specificational pseudoclefts can be found in the careful corpus work of Delin 1989. So it is a characteristic of specificational sentences that the precopular free relative is the part that holds a connection to the previous discourse ("ground," in the terminology of Vallduví 1992, "theme" in that of Steedman 2000a, 2000b), while the postcopular phrase is the part that provides information about the theme (Vallduví's "focus"; Steedman's "rheme"). Now, observe that that this does *not* follow from the claim that the free relative is a concealed answer, since it is perfectly possible for the concealed answer reading to coincide with focus in true equatives and in other types of sentence:

(35)   A: So what do you know about John's goal?
          B: John's goal is [$_F$ MY GOAL].

(36)   A: What kind of information do you have about the next meeting?
          B: I know [$_F$ THE DATE].

6. Integrating information structure

What we propose, therefore, is that the correct characterisation of specificational sentences is very slightly different: the initial phrase is not (the answer to) a question; rather it is the **ground** of the assertion made by the sentence. The special fact about specificational sentences is that they allow the ground to be constructed from the initial noun phrase, rather than from the entire sentence. Given this possibility, the connectivity properties then arise from the normal mechanisms for putting ground and focus together.

Suppose, contrary to many common assumptions, but following Steedman 2000a, 2000b, that the focus/ground (rheme/theme) division of a sentence is not established at some late point in its interpretation but rather in the initial syntactic parse. The focus is a parsed syntactic constituent that contains the word bearing the nuclear accent of the sentence. The ground is the rest of the sentence with the focus constituent removed, leaving behind perhaps a copy of the root node of that constituent. To illustrate, consider the sentence in (37):

(37)   Mary read THE NEWSPAPER.

The capitalized constituent should be taken as the focus of the sentence, which would be the case if the sentence were used to answer the question

What did Mary read? or in a similar context. Under our proposal the parse of (37), would yield the pair of constituents in (38):

(38)   [$_{\text{GROUND}}$ Mary read NP], [$_{\text{FOCUS}}$ the newspaper]

The obvious syntactic representation is as in (39):

(39)
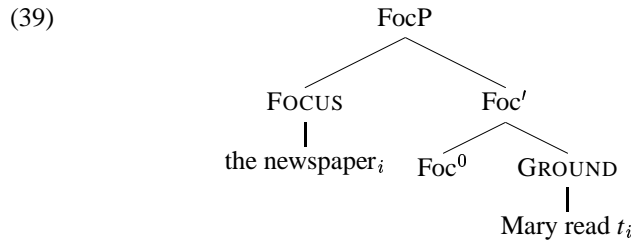
The ground of a sentence is presupposed material; that is to say, it corresponds to shared beliefs of the speaker and hearer of the sentence before the sentence is uttered. Often the presupposition associated with the ground is represented as an existential sentence or formula. For example, as (40) for the sentence in (37):

(40)   Mary read something.

A common alternative is to think of the ground as corresponding to a function created by abstraction over the position of the "gap" left by the focus:

(41) λx[Mary read x]

This is the proposal that we will assume here.

As we have seen, there is a problem for interpreting sentences like *The culprit is John* on its specificational reading. One possible derivation for a copular sentence with a definite in final position is to allow the postcopular phrase to have (Graff) or to acquire (Partee) the type of a predicate. This yields a grammatical sentence: but not a specificational sentence, as we have seen. Further, we continue to maintain the conclusion of Heycock and Kroch 1999 that it is not possible for a predicate in initial position to take the postcopular phrase as its argument.

The solution to this problem lies in the information structure of specificational sentences. Suppose that just in case the two phrases cannot combine because they are both referring expressions of the same type, instead of abstracting on the position of the focus to produce a function, the parsing algorithm makes use of a function that is already available: the predicate that is part of the initial referring expression.

Exactly how this is implemented depends, of course, on what one takes

to be the semantics of definite/specific noun phrases. For example, in the proposal of Heim 1982, or the related proposal in Graff 2001, definites are like indefinites in introducing a description and, if they occur in argument rather than predicate position, a variable which has to be bound by a quantifier introduced into the structure from some other source. Under this kind of analysis, all that is required is that the lambda presumably associated with the focus head bind the variable, rather than this being done by the existential quantifier that would be introduced by existential closure or its equivalent.

(42)   The culprit is $[_F$ John $]$
          Ground $= \lambda x[culprit(x) \& \forall y(culprit(y) \rightarrow y = x)]$; Focus $= John$

What about the rest of the sentence, though? In the usual case the natural assumption is that the ground is created by the (covert) movement of the focused element to the specificer of the focus head, as in (39) above, where the variable corresponds to the trace of movement. In a specificational sentence like (42), why are the copula and the trace of the focused element not part of the ground?

As far as the trace is concerned, we simply assume (as is fairly common) that traces do not occur, or can be deleted, when they are not required for interpretation. And recall that we are quite explicit *not* proposing that the copula in specificational sentences is a meaningful element; hence it too can be deleted at LF. Thus no special mechanism is required.

One attractive property of our approach to specificational sentences is that it applies without modification to specificational pseudoclefts. Thus, leaving aside issues of tense and temporal interpretation, the representation for both the simplex specificational sentence in (43a) and the pseudocleft in (43b) is the same, namely (43c)

(43)   a.   His claim was that John had fallen down.
          b.   What he claimed was that John had fallen down.
          c.
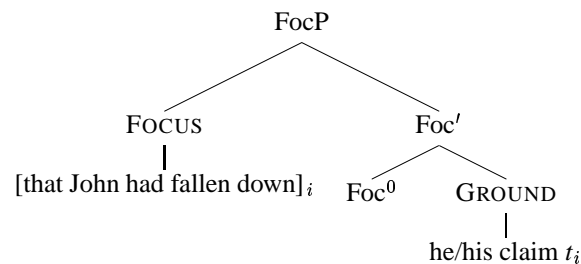
This follows from the analysis given above, together with the assumption that a pseudocleft free relative is a kind of definite referring expression. This result is promising since it provides a representational basis for the attested

parallel in connectedness between pseudoclefts and simplex specificational sentences. Here, for example, principle C applies to both examples, ruling out coreference between *he* and *John.*

Further, it is important to note that our Logical Form representation of a specificational pseudocleft or other specificational sentence and that of a non-cleft sentence with the same focus and ground assignment are indistinguishable. Thus, the representation in (43c) is valid not only for the two specificational cases but also for the canonical form of the sentence:

(44)   He claimed [$_F$ that John had fallen down].

This is, of course, just what we want if we hope to use Logical Form to explain connectedness.

Note that it seems to be the case that the ground can consist of the initial phrase only if the initial phrase is specific, in the sense of Enc 1991. That is, as we have seen, not only are adjectival predicates ruled out from this position, but indefinites are good only to the extent that they have a "specific" reading. This seems to be the same restriction as is found for the subject of Individual-level predicates:

(45)   a.   The man you need to talk to is John.
         b.   ONE man you need to talk to is John.
         c.   ?A man you need to talk to is John.
         d. ?*A tall man is John.
         e.   *A man is John.

(46)   a.   The man you need to talk to is intelligent.
         b.   ONE man you need to talk to is intelligent.
         c. ??A man you need to talk to is intelligent.
         d.   *A tall man is intelligent.   *Generic reading only.*
         e.   *A man is intelligent.   *Generic reading only.*

If this is the correct characterisation, we have to understand exactly what it is that these "presuppositional" or specific indefinites have in common with definites. For example, we would not want to pursue the kind of proposal in Diesing 1992, who treats this kind of indefinite as quantificational; but the correct analysis for this restriction is not yet clear to us.

We have claimed that specificational sentences have the same logical form as corresponding canonical sentences. in the case of pseudoclefts, this opens us to an objection because the uniqueness presupposition associated with the ground in these sentences appears to be stronger than in their single sentence paraphrases. For example (47a) seems less felicitous than (47b):

(47)   a.   What he read was this book. And he read that one too.

b.　He read THIS BOOK. And he read that one too.

However, given the well-known fact that there is often no clear marking of the boundary of the focus constituent (Steedman 2000a; Steedman 2000b) (in earlier terms, focus can "project" from the word or phrase that actually bears the pitch accent) it is not clear that the information structure of these two examples is the same. We believe in fact that there is no difference with respect to uniqueness presuppositions if these other interfering factors are controlled for. Note for example that in contexts where the presupposition/implicature of uniqueness is suspended in the case of prosodically marked focus, the same effect is seen in the corresponding pseudocleft, as illustrated in (48):

(48)　　A:　I need to buy some salami. Where can I buy salami around here?
　　　　B:　You can buy salami [$_F$ at Pathmark].
　　　　B2:　Where you can buy salami is [$_F$ at Pathmark].

7. Connectivity

Given the possibility of constructing the ground from the initial phrase in a specificational sentence, syntactic conditions will hold in these sentences just as in non-copular sentences with the same information structure. That is, all of (49a–c) have the same representation at LF (note crucially the relevant reading of (49c) is as focus-movement, with no second accentual peak within the clause). If we take this representation to be as in (50), this amounts to assuming total obligatory "reconstruction" of the Focus constitutent:

(49)　　a.　What she saw was two flamingos.
　　　　b.　She saw [$_F$ two flamingos].
　　　　c.　[$_F$ Two flamingos] she saw.

(50)

FocP
├── FOCUS
│   └── two flamingos$_i$
└── Foc′
      ├── Foc^0
      └── GROUND
            └── she saw t_i
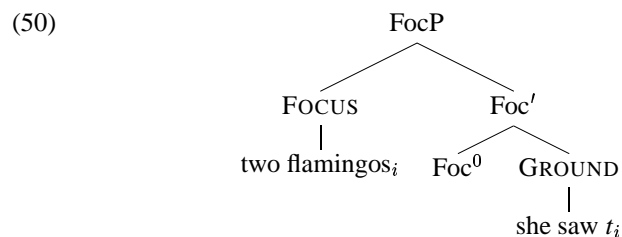
In the case of focus signalled by intonation alone, the need for total reconstruction or its equivalent is obvious. And as we will see below, there is evidence that focus movement contrasts with topicalisation in also requiring the same kind of total reconstruction. Topicalisation, on the other hand, is known to behave somewhat differently with respect to reconstruction; and as

we will see the pattern of reconstruction in topicalisation is matched exactly in pseudoclefts of a certain type.

7.1. Connectivity and information structure: the licensing of Negative Polarity Items

We have designed our syntax to capture the parallels between specificational sentences and canonical sentences with the same information structure, and it seems to do so at a reasonable cost. But we might ask whether we have gone too far in linking the two sentence types. After all, there is now no room for differentiating them. This might appear to result in problems given the differences in licensing of Negative Polarity Items, an issue pointed out in den Dikken, Meinunger, and Wilder 2000. In this section we will show that NPIs in fact behave as expected, modulo a constraint on their distribution that is not particular to pseudoclefts.

Consider, for example, the contrast between the two sentences of (51):

(51)   a.   *Any firemen weren't available.
        b.   What wasn't/weren't available was/were any firemen.

It seems at first quite surprising that (51b) should be grammatical, given that, on the one hand, *any* is not c-commanded by negation and, on the other, that the apparently equivalent simple sentence does not license the NPI. In fact, however, the difference we find here is actually expected, when we look more closely at the contexts that license NPIs. In her thesis, Linebarger (1980) pointed out that contexts in which a noun phrase is semantically inside the scope of negation license NPIs within that noun phrase even though, syntactically, the noun phrase seems to be outside the scope of negation. A classic example of this type is (52):

(52)   A doctor with any real knowledge of acupuncture wasn't available.

The pseudocleft version of (52), given in (53), is just as acceptable as its simple sentence paraphrase:

(53)   What wasn't available was a doctor with any real knowledge of acupuncture.

This fact shows that the connectivity effect for NPI licensing found for object NPIs in pseudoclefts extends correctly to subject position in those cases where the NPI is ordinarily licensed there, suggesting that our reconstruction account is not flawed. In fact, the grammaticality of (52) raises the question of whether the grammaticality of (51b) isn't the expected case and the ungrammaticality of (51a) the unexpected one.

As is well known, *available* is a stage level predicate whose associated negation can easily take scope semantically over its subject, although the subject, at least as far as the surface syntax is concerned, seems to be in the canonical position. This is illustrated by the preferred reading of a standard example like (54), in which negation over the implicit existential quantifier scopes over the subject:

(54)   Firemen weren't available.

If the subject is within the scope of negation, it would be expected that an NPI would be licensed in that position. That is what all of our examples except (51a) indicate. Apparently, an NPI cannot c-command its licenser in surface syntax. This requirement is awkward for an expression whose licensing is essentially semantic, but the grammaticality of (51b) argues for it, showing that even an environment as strongly subject to syntactic reconstruction as a pseudocleft does not reconstruct an NPI into an illicit position in syntax if it properly licensed semantically.[3] In support of this approach we note that a similar anti-c-command requirement seems to hold of reflexives and their binders. Consider the examples in (55)

(55)   a.   Each other's antics amuse them.
         b.   *Each other amuse them.

Verbs like *amuse*, which have a non-subject experiencer, show reverse binding into the subject, as shown in (55a), but this binding is impossible when the subject itself would be bound by the experiencer, as in (55b). It might seem that the problem here is not due to binding but to the fact that anaphors cannot appear in the nominative case, but the ungrammaticality of (56), where the case problem does not arise due to the 'for' complementizer, shows that there is still a binding problem:

(56)   *For each other to amuse them would be unexpected.

Note what happens in the pseudocleft versions of these sentences, given in (57):

(57)   a.   What amuses them is each other's antics.
         b.   What amuses them is each other.

---

3.   The "surface" nature of some of the constraints on NPI licensing is discussed in more length in McCloskey 1996:86–97, who mentions among other effects the contrast between (ia) and (ib):

(i)   a.   Which one of them doesn't anybody like?
         b.   *Which one of them does anybody not like?

The pattern is just the same as in the NPI case. These cases of lack of parallelism between pseudoclefts and simple sentences, far from undermining our strong syntactic reconstruction account, rather support it. Pseudocleft connectivity appears to be every bit as thoroughgoing in its effects as our account requires.[4]

7.2. Reverse pseudoclefts and connectivity

As was mentioned earlier, pseudoclefts appear to be possible with two orders: the "canonical" order, in which the free relative is initial (den Dikken et al's Type A), and the "reverse" order (Type B); see (58):

(58)   a.   What she is is rather too pleased with herself. *Type A: canonical*
        b.   Rather too pleased with herself is what she is. *Type B: reverse*

For ease of exposition, we will refer to the two types as "canonical" and "reverse" pseudoclefts.[5]

Reverse pseudoclefts do not show exactly the same connectivity effects as canonical pseudoclefts. This is a potential problem for all analyses which derive such effects from the semantics of equation (*e.g.* Sharvit 1997; Sharvit 1999, Heycock and Kroch 1999, Schlenker 2001). If the information structure of reverse pseudoclefts were identical to that of canonical pseudoclefts, the asymmetries in connectivity would also be problematic for the analysis we are presenting here. However, it turns out to be the case that reverse pseudoclefts do not have the same information structure as their canonical counterparts: the "counterweight" (the phrase that is not a free relative) is typically not the focus, but rather the topic.

Evidence that the initial phrase in a reverse pseudocleft can be part of the ground rather than being the focus comes from contrasts like that between (59) and (60). The B1 answer in (59) shows that the precopular noun phrase in a reverse pseudocleft can be a focus:

---

4.   Example (55b) might also be ruled out as a principle B violation, since the pronoun is c-commanded by a co-indexed noun phrase in the relevant local domain. Given that this is a case of reverse binding, however, it is not clear whether principle B should apply here. Furthermore, in cases where it does apply, as in (i) below,

(i) $John_i$ loves $him_{j/*i}$.

the binding prohibition survives in the pseudocleft, as we see in (ii):

(ii)   Who $John_i$ loves is $him_{j/*i}$.

5.   Crucially, note that the "canonical" order pseudocleft is a specificational sentence; this is the type of sentence that Moro, Williams, etc, have argued to be "inverted".

(59)   A: What is John's favorite food?
        B1: [$_F$ CHEESESTEAK] is what John loves.   *Reverse*
        B2: What John loves is [$_F$ CHEESESTEAK].   *Canonical*

The acceptablity of B1 in (60) shows that the precopular noun phrase in a reverse pseudocleft can also be a topic, hence part of the ground; this is impossible for the corresponding postcopular phrase in a canonical pseudocleft, as shown by the unacceptability of B2:

(60)   A: Whose favorite food is scrapple?
        B1: [$_{Topic}$ Scrapple] is [$_F$ what BILL loves].   *Reverse*
        B2:*[$_F$ What BILL loves] is [$_{Topic}$ scrapple].   *Canonical*

This difference is enough to account for the differences in connectivity between the two types. When the precopular expression in a reverse pseudocleft is not a focus (which it may be—as in B1's response in (59)—although this is not typical), the sentence seems to have the information structure of a **topicalization**, a conclusion which is again supported by the results in Delin 1989.

If the particular connectivity properties of canonical order pseudoclefts/specificational sentences arise from an operation made possible by their information structure, we then do not expect to find the same connectivity properties in reverse pseudoclefts. As den Dikken, Meinunger, and Wilder 2000 have shown, this is true, although, as will be pointed out below, we disagree with some aspects of their data presentation and analysis. What we will conclude is that reverse pseudoclefts show exactly the connectivity properties of leftward A′-movement in English. The most usual instance of this is topicalisation; however, depending on context and stress assignment the initial A′-moved phrase can be a focus: and in this case the connectivity facts are the same as in canonical order pseudoclefts.

Here we will itemize the connectivity properties that den Dikken, Meinunger, and Wilder 2000 discuss for pseudoclefts and show that in the general case they are identical to those found in topicalization.

- **De dicto/de re ambiguities**:

  De dicto readings of the counterweight phrase are possible in reverse pseudoclefts, just as in canonical order pseudoclefts.

  (61)   a.   What he is looking for is a cheap house. *Ambiguous*
          b.   A cheap house is what he is looking for. *Ambiguous*

  Note that apparent de dicto readings can also be found in non-cleft copular sentences like (62a–b):

(62)   a.   Her goal is a new house.
        b.   A new house is her goal.

They are also possible in topicalizations:

(63)   A cheap house, I have been looking for for some time now.

- **Bound variable readings of pronouns**

Bound variable readings are possible in reverse pseudoclefts, just as in canonical order pseudoclefts:

(64)   a.   What [no Scottish woman]$_i$ can live without is her$_i$ oatcakes.
        b.   Her$_i$ oatcakes is/are what [no Scottish woman]$_i$ can live without.

Again, these are possible also in topicalizations:

(65)   Her$_i$ oatcakes, [no Scottish woman]$_i$ can live without.

- **Binding**

Reverse pseudoclefts show connectivity effects with respect to Condition A:[6]

(66)   a.   Himself$_i$ is who John$_i$ saw.
        b.   Proud of himself$_i$ is what John$_i$ has always been.

and Condition C:[7]

(67)   *John$_i$ is who he$_i$ thinks they are about to fire.

---

6.   It is hard to test Condition B because with the kind of focal stress that is obligatory in pseudoclefts, apparent violations of Condition B can be amnestied even in non-cleft sentences; in addition, with this kind of stress it is in general more natural to use an emphatic reflexive.
7.   Note the contrast with (i), which has a non-pseudocleft reading that freely allows the given coindexation:

(i)   John$_i$ is who he$_i$ wants to be.

This coreference is possible even if *John* is focused:

(ii)   A: Who around here is happy with himself? Who is who he wants to be?
       [B:] [$_F$ JOHN$_i$] is who he$_i$ wants to be.

And again such effects are found in topicalizations:

(68) a. Himself$_i$, John$_i$ saw.
  b. Proud of himself$_i$, John$_i$ has always been.
  c. *John$_i$, he$_i$ thinks they are about to fire.

There is however a difference between canonical and reverse pseudoclefts with respect to Condition C violations. In reverse pseudoclefts only we find "anti-reconstruction" effects: R-expressions buried inside other referential expressions or inside CPs can be coreferential with a pronoun c-commanding the "gap." Thus, note the difference between the canonical and reverse pseudoclefts in (69c,d) and (70c,d) respectively:

(69) a. What he$_{*i/j}$ was was proud of John$_i$.
  b. What he$_{*i/j}$ will never do is scold John$_i$'s children.
  c. What he$_{*i/j}$ really missed was John$_i$'s dog.
  d. What he$_{*i/j}$ had always claimed was that John$_i$ was innocent.

(70) a. Proud of John$_i$ was what he$_{*i/j}$ was.
  b. Scold John$_i$'s children is what he$_{*i/j}$ will never do.
  c. John$_i$'s dog was what he$_{i/j}$ really missed.
  d. That John$_i$ was innocent was what he$_{i/j}$ had always claimed.

The contrast between reverse pseudoclefts with AP/VP counterweights (70a,b) and those with DP/CP counterweights (70c,d) is mentioned as a puzzle in a footnote in den Dikken, Meinunger, and Wilder 2000: 84. However, it mirrors exactly the pattern that exists in topicalization (Huang 1993, Heycock 1995):

(71) a. Proud of John$_i$, he$_{*i/j}$ was.
  b. Scold John$_i$'s children he$_{*i/j}$ never will.
  c. John$_i$'s dog, he$_{i/j}$ really missed.
  d. That John$_i$ was innocent, he$_{i/j}$ had always claimed.

- **Negative Polarity Items (NPIs)**

Negative polarity items can occur as the focus of a canonical pseudocleft, but not as the first element of a reverse pseudocleft (den Dikken, Meinunger, and Wilder 2000: 43*ff*)—see the contrast between (72) and (73):

(72) a. He bought plenty of text books, but what he didn't buy was any novels.
  b. There was plenty of cider, but what nobody bought was any wine.

(73)   a. *...any novels was/were what he didn't buy.
        b. *...any wine was what nobody bought.

NPIs that are contained within a larger XP, however, may occur also in reverse pseudoclefts, as illustrated in (74b):[8]

(74)   a.   What we couldn't find was a doctor who knew anything about acupuncture.
        b.   A doctor who knew anything about acupuncture was what we couldn't find.

In cases where the wh-clause corresponds to the type of context where negation can license an NPI it does not c-command (Uribe-Echevarria 1994), the NPI is also possible in the initial position of a reverse pseudocleft—see (75):

(75)   They desperately needed someone with at least a little expertise in acupuncture, but
        a doctor who knew anything about acupuncture was exactly what *wasn't* available.

Again we find a perfect parallel with topicalization. Negative Polarity Items can generally not be topicalized:[9]

---

8.   In this judgement we disagree sharply with den Dikken, Meinunger, and Wilder 2000, who claim for example that (ia) (their (49d)) is ungrammatical, contrasting with (ib):

(i)   a.   that anybody was injured was what she didn't say
        b.   what she didn't say was that anybody was injured

We agree that (ia) is awkward, but not to the point of ungrammaticality (Schlenker 2001 shares this view (p. 52)) and more natural examples can be constructed, for example:

(ii)   Bluebeard had reservations about the conduct of the raid; that he had spared anyone was what he regretted the most.

The example in (75) is also cited by den Dikken, Meinunger, and Wilder 2000 as ungrammatical; again we disagree strongly with this judgement.
9.   It is not at all clear why NPIs cannot be topicalized. Note in particular that this is not simply because the NPI has to take scope under the negative, since narrow scope for *many* is clearly possible in the same context:

(i)   He has many acquaintances, but
        a.   *...any real friends he doesn't have.
        b.   ...many real friends he doesn't have.

It is possible that the ungrammaticality of examples like (ia) may not be the result of

(76) a. *I bought lots of text books, but any novels, I didn't buy.
      b. *There was plenty of cider, buy any wine, nobody bought.

Just as in reverse pseudoclefts, however, NPIs that are contained within a larger XP, as in (77), may occur in topicalizations (*cf* (74)):

(77) We found various doctors, but a doctor who knew anything about acupuncture, we couldn't find.

Since we have argued that the initial phrase in a reverse pseudocleft is typically a topic—in contrast to the postcopular phrase in a canonical pseudocleft, which must be a focus—the parallel that we have just seen is as expected. In particular, we expect to find antireconstruction effects. However, we have also argued that the initial phrase in a reverse pseudocleft can, alternatively, be a focus. It follows from our analysis, then, that if we can find reverse pseudoclefts that unambiguously have this information structure, they should not display antireconstruction effects. We believe that this prediction is borne out.

It is not straightforward to eliminate a topicalization reading for any given example, but we may appeal to the fact that topicalisation is normally assumed not to be possible with the "non-referential" (narrow scope) reading of amount QPs. Thus, while (78a) is ambiguous between topicalisation and focus-movement, (78b) can only be an instance of focus movement (given that the lexical semantics of a verb of creation exclude the possibility of a wide-scope existential reading for the QP *four stories*), and has the characteristic intonation pattern with the only accentual peak on the dislocated phrase:

(78) a. Four stories about the war, she would like to rewrite. *Topicalisation or Focus Movement*
      b. Four stories about the war, she would like to write. *Focus Movement only*

If we now turn to the corresponding reverse pseudoclefts, we find that indeed, where the initial phrase cannot be construed as a topic, the antireconstruction

---

a failure of "reconstruction"—*i.e.* the failure of the NPI to be licensed by the negation that it asymmetrically c-commands from its S-Structure position. Rather, it appears that for some reason, NPIs simply may not be topicalized, as shown by the following contrast:

(ii) a. ?Nobody thought that those books, they would burn.
      b. *Nobody thought that any books, they would burn.

In (iib), the embedded topicalization seems sharply worse than it does in (iia). But the NPI is still asymmetrically c-commanded by the negation that licenses it, so it appears that the topicalization alone must be responsible for the unacceptability.

effect disappears:

(79)   Twenty stories about John$_i$ is/are what he$_{*i/j}$ has absolutely refused to write.

Further, as predicted, the same is true of non-cleft sentences with the same information structure. In (80) this is demonstrated for topicalization versus focus movement:

(80)   a.   Twenty stories about John$_i$, he$_{i/j}$ has absolutely refused to rewrite.
   b.   Twenty stories about John$_i$, he$_{*i/j}$ has absolutely refused to write.

Thus the fine detail of the antireconstruction effects bears out the hypothesis that the initial phrase in a reverse pseudocleft can be either topic or focus, and that this affects its connectivity properties.

   Note further that since we have independent evidence that NPI licensing is affected by Surface Structure—the NPI must not c-command the licensor—it is also expected that only an embedded NPI can appear in a reverse pseudocleft, just as with a simplex sentence subject. Thus we predict the difference, already noted, between (14a) and (14b), repeated here as (81a–b):

(81)   They bought a lot of textbooks, but ...
       a.   what they didn't buy was any novels. *Canonical*
       b.   *any novels was/were what they didn't buy. *Reverse*
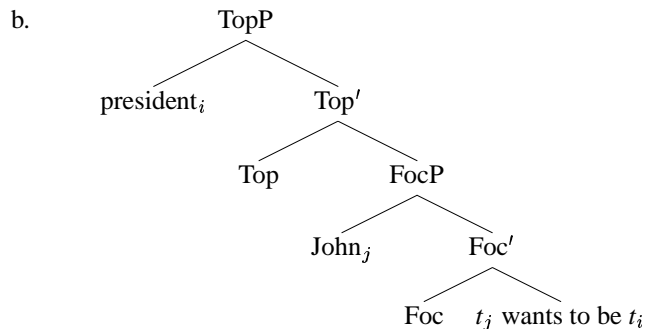
8. Representing topicalized sentences and reverse pseudoclefts in Logical Form

   We have yet to address the question of the structure to assign to topicalizations, whether in cleft or simplex sentences. Just as we have assumed that there is a head for Focus, so we adopt the relatively common assumption that there is also a head whose specifier hosts the Topic, higher in the structure:
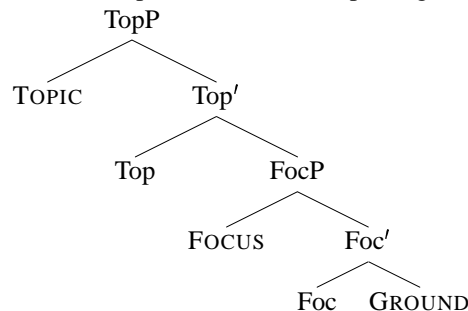
Note that in this structure the COMMENT constituent is here identified with FocP, and the TOPIC appears outside the constituent labeled GROUND. While in the most straightforward cases the variable that is bound by the element in Spec,TopP is the result of movement to that position, we assume that it can also be the variable introduced by a free relative or equivalent expression: this is what happens in a reverse pseudocleft:

(82) a. Mary is not interested in being president, [$_{Topic}$ president] is what John wants to be.

   b. TopP

```
            TopP
           /    \
   president_i   Top'
                /    \
              Top    FocP
                    /    \
                John_j   Foc'
                        /   \
                      Foc   t_j wants to be t_i
```

One question that our proposal does not yet answer and that we leave for the future is why the "reconstruction" properties of the elements in Spec,TopP and Spec,FocP should differ. Given that there is independent evidence that they do, however, the different connectivity properties of canonical and reverse pseudoclefts follows directly.

9. Summary and conclusions

In this paper we have discussed some problems inherent in a treatment of specificational sentences as equatives; we have proposed that in order to resolve these it is necessary to allow the information structure of such sentences to play a crucial role. Specifically, we have proposed that a marked option for building the ground is the basis for their striking connectivity properties. We have further discussed the contrast in connectivity between canonical and reverse pseudoclefts, and argued that this follows directly from the different information structural properties of the two orders.

References

Büring, Daniel. (1998). Two types of identity statements. Draft of a talk given at SALT 8, MIT. Ms, Cologne University.

Delin, Judy. (1989). *Cleft Constructions in Discourse*. Ph. D. thesis, University of Edinburgh.

den Dikken, Marcel, André Meinunger, and Chris Wilder. (2000). Pseudoclefts and ellipsis. *Studia Linguistica 54.1*, 41–89.

Diesing, Molly. (1992). *Indefinites*. Cambridge, Mass.: MIT Press.

Enc, Mürvet. (1991). The semantics of specificity. *Linguistic Inquiry 22.1*, 1–25.

Graff, Delia. (2001). Descriptions as predicates. *Philosophical Studies 102*, 1–42.

Heim, Irene. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Ph. D. thesis, University of Massachusetts, Amherst.

Heycock, Caroline. (1994). The internal structure of small clauses. In J. Beckman (Ed.), *Proceedings of NELS 25*, Volume One, Amherst, Mass., pp. 223–238. GLSA.

Heycock, Caroline. (1995). Asymmetries in reconstruction. *Linguistic Inquiry 26.4*, 547–570.

Heycock, Caroline and Anthony Kroch. (1999). Pseudocleft connectedness: Implications for the LF interface level. *Linguistic Inquiry 30.3*, 365–397.

Higgins, F. Roger. (1979). *The Pseudo-Cleft Construction in English*. New York: Garland.

Huang, C.-T. James. (1993). Reconstruction and the structure of VP: some theoretical consequences. *Linguistic Inquiry 24.1*, 103–138.

Linebarger, Marcia. (1980). *The Grammar of Negative Polarity*. Ph. D. thesis, MIT.

McCloskey, James. (1996). On the scope of verb movement in Irish. *Natural Language and Linguistic Theory 14*, 47–104.

Mikkelsen, Line. (2002). Two types of definite description subjects. In *Proceedings from the Student Session at ESSLLI 14*, Trento. ESSLI 14.

Moro, Andrea. (1997). *The Raising of Predicates: Predicative Noun Phrases and the Theory of Clause Structure*. Cambridge: Cambridge University Press.

Romero, Maribel. (2001). Specificational subjects and concealed questions. Talk given at the Copular Sentence Workshop, University of Pennsylvania.

Rothstein, Susan. (2001). *Predicates and their Subjects*, Volume 74 of *Studies in Linguistics and Philosophy*. Dordrecht, Boston: Kluwer.

Schlenker, Philippe. (2001, July). A note on the connectivity problem. Ms., University of Southern California.

Sharvit, Yael. (1997). *The Syntax and Semantics of Functional Relative Clauses*. Ph. D. thesis, Rutgers University.

Sharvit, Yael. (1999). Connectivity in specificational sentences. *Natural Language Semantics 7*, 299–339.

Steedman, Mark. (2000a). Information Structure and the syntax-phonology interface. *Linguistic Inquiry 31.4*, 649–690.

Steedman, Mark. (2000b). *The Syntactic Process*. Cambridge, Mass.: MIT Press.

Uribe-Echevarria, Maria. (1994). *Interface Licensing Conditions on Negative Polarity Items: A Theory of Polarity and Tense Interactions*. Ph. D. thesis, MIT.

Vallduví, Enric. (1992). *The Informational Component*. New York: Garland.