Origin of the English Double-*is* Construction

Elizabeth Coppock and Laura Staum

December 27, 2004

1 The Remarkable Double-*Is*

We are concerned with the pattern exemplified in (1a)-(1e). We refer to these as double-*is* constructions.[1]

(1) a. The thing is, is we've got to be strong. (Massam 1999)

 b. The point is is I'm not in business to be loved. (McConvell 1988)

 c. What'll happen is is that somebody'll get hurt. (Bolinger 1987)

 d. The problem is, is that we can't find the evidence. (Massam 1999)

 e. The hitch is, is that it seems to be occurring in the opposite direction. (Massam 1999)

For the sake of discussion, we will refer to the first *is* as be_1, and the second *is* as be_2.

(2) The thing is, is that ...
 be_1 be_2

2 Previous analyses

2.1 *What*-drop

According to Massam (1999), there's an empty *what* before the copula, represented schematically in (3).

(3) ~~what~~ the problem is, is that ...

If this were true, then we should be able to insert a *what* in all double-*is* constructions, maintaining acceptability. This falsely predicts (4b) and (5b) to be grammatical.

(4) a. What I'm saying is is that ...

 b. *What what I'm saying is is that ...

(5) a. That can't be a very welcome outcome, is that rates will now rise. (McConvell p. 302)

 b. *What that can't be a very welcome outcome, is that rates will now rise.

---

[1]This construction is not limited to American English; as McConvell shows, both Australian and British speakers produce such utterances.

2.2 Blend

McConvell's analysis is that double-*is* constructions represent a blend between the two analyses of pseudo-clefts given in (6a) and (6b):

(6) a. [What I'm saying] [is that ...]

  b. [What I'm saying is], [that ...]

Putting them together yields (7):

(7) [What I'm saying is] [is that ...]

One of the problems with this analysis is that *what I'm saying is* cannot easily be considered a syntactic constituent. McConvell's argument for its status as such relies on intonation, which does not constitute sufficient evidence for constituency; as is well known, intonation contours can fail to coincide with syntactic constituent boundaries. One piece of evidence against *what I'm saying is* as a syntactic constituent is that it cannot be replaced by a pro-form, as shown in (8).

(8) *It that ...

Another problem with this analysis is that it predicts that be_2 should agree in number with the sentence subject, but number agreement is not only optional but prohibited, as shown in (9).

(9) a. The cruel facts of life **are, is** that not every person who teaches Art is a good artist himself.

  b. *The cruel facts of life **are, are** that not every person who teaches Art is a good artist himself.

2.3 Focus Particle

Both Massam and McConvell mention the possibility that be_2 is in fact a focus particle, although they do not incorporate this into their analyses.[2] This idea forms the basis of the analysis given by Tuggy (1995). In particular, Tuggy analyzes *is that* as a "unit complementizer", and Koontz-Garboden (2001) gives the same analysis in LFG. This particular version of the focus-particle hypothesis is not tenable because it does not account for situations in which *is* acts as a focus particle and yet is not followed by *that*. In fact, the focus-particle *is* can be followed by complementizers other than *that*:

(10) The question is, is **whether** we should do that.

It can also be followed by a null complementizer:

(11) The thing is is I don't think he likes me.

It can even be followed by a noun phrase:

(12) The thing is is just my dad.

However, we accept the notion that be_2 is a focus particle. This avoids the problems with the *what*-drop and blend analyses; it does not predict examples such as (4b) or (5b) to be grammatical, and it accounts for the fact that we do not get number or person agreement between be_2 and the subject as shown in (9). In general, it accounts for the fact that be_1 is more flexible than be_2. Whereas be_1 can be non-finite, be_2 cannot, as shown in (13).

---

(13) a. The problem **being, is** that I'm probably going to test positive. (Massam, p. 349)

b. *The problem **is, being** that I'm probably going to test positive.

c. *The problem **being, being** that I'm probably going to test positive.

One fact that this analysis does not account for is the fact that be_2 can appear either not only as *is* but also as *was*, as shown in (14).

(14) My feeling **was, was** that she doesn't have a professional hold on the situation.

These two variants are not in complementary distribution, as shown by the grammaticality of examples like (15).

(15) The thing **was, is** that we had no control over the situation. (Massam, p. 349)

They are not in free variation either, as demonstrated by the ungrammaticality of example (16).

(16) *The thing **is, was** that we had no control over the situation.

It appears that be_2 can only appear as *was* when be_1 also takes the past tense form. be_2 can appear as *is* regardless of the tense of be_1, as shown by (15). Koontz-Garboden offers an explanation of this pattern in LFG with a lexical entry for *was* as a focus particle that constrains the tense of the main sentence to PAST. We take this to be essentially correct; *is* is a focus particle that does not impose any tense constraints, and *was* is a particle that imposes tense constraints.

One consequence of analyzing be_2 as a non-projecting particle is that it does not change the category of the phrase it combines with. This is a welcome consequence, as demonstrated by the following examples, which show that no matter what category be_2 combines with, removing it yields a standardly grammatical sentence:

(17) CP: I'd like to say **is** that ...(Bolinger 1987)

(18) IP: Our kids are great on vacation, but when they come back, **is** they need to play. (Massam 1999)

(19) PP: The other problem is **is** on the demand side [as opposed to the supply side]. (Bolinger 1987)

(20) NP: What I'm looking for is, **is** reliability ...(Switchboard corpus)

In the same vein, analyzing be_2 as a focus particle has the advantage of explaining the grammaticality of examples (21)-(23), all found in the Switchboard corpus.

(21) And I think that's what makes one feel invaded **is** the fact that there seems to be little control ...

(22) That's what I like to make **is** just real neat stuff like that.

(23) That's, uh, I, that's what I've heard people say, **is** they'll never go north of the Red River again.

At first glance, these sentences appear to have a full sentence as their subject, a surprising violation of the rules of standard English. Analyzing *is* as a focus particle (be_2) allows us to interpret these examples in terms of familiar grammatical structures: *is* is a particle adjoined to an appositive clause.

2.4 Distribution

Simply identifying be_2 as a focus marker explains the otherwise mysterious limited distribution. In particular, it explains why examples such as (24) are not possible.

(24)   *My dog is, is a pain in the ass.

We assume that be_2 marks the same kind of focus that pseudoclefts mark. Items that receive this kind of focus must be selected from of a set of alternatives. Therefore, for *is* to be capable of marking this kind of focus, a set of alternatives must be established. In a pseudocleft, the set of alternatives is established through the *wh*-NP. Double-*is* constructions do not have *wh*-NP topics, so their topic NPs must be of the type that can establish a set of alternatives without a *wh*-pronoun; appositional NPs such as *the thing* and *the problem* are capable of establishing a set of alternatives, but NPs like *my dog* cannot.

3   How it came to be

Tuggy (1995) calls the double-*is* construction "a marginal structure in the process of becoming grammatical". It is the details of this process of grammaticalization with which we are now concerned.

3.1   Grammaticalization

Our diachronic analysis bears an analogy to Elizabeth Traugott's analysis of the development of *going to*. According to Traugott, the development of *going to* had three stages, represented schematically in (25).

(25)   MOTION → MOTION+FUTURE → FUTURE

This process can be represented more abstractly as in (26).

(26)   A → A+B → B

In our analysis, A corresponds to the copula, and B corresponds to focus marking, shown in (27).

(27)   COPULA → COPULA+FOCUS → FOCUS

3.1.1   Step A: Association

Before *going to* became a future marker, whenever *going to* was followed by a verb, the action described would take place in the future, and would take place intentionally. According to Traugott, a process of "semanticization" takes place, which we interpret as follows: whereas when *going to*+V could only be used in motion contexts, the future meaning had to be inferred, later the future meaning was marked by the construction, or part of the lexical entry for it.

Before *is* became a focus particle, whenever it occurred in a basic pseudocleft as in (28), it acted as a copula, and it introduced a focussed phrase.

(28)   What I said **is** that I hate you.

The correlate to Traugott's "semanticization" step here is the association of *is* in this context with focus. The strength of this association rests on the fact that *is* occurs as the matrix verb in 86% of pseudoclefts, according to a preliminary mining of the Switchboard corpus; 141 randomly-selected pseudoclefts were analyzed, and 121 of them had *is* as the matrix verb (*was* constituted 12%).

3.1.2 Step A+B: Ambiguity

Crucial to Traugott's analysis of *going to* is that there were ambiguous contexts, ones in which the future meaning was definitely present, but the motion meaning was optional, as in example (29).

(29) Thys onhappy sowle... was goyng to be broughte into helle for the synne and onleful [unlawful] lustys of her body. (1482, Monk of Evesham [OED **go** 47b])

Assuming that no speaker would have used *going to* without the motion meaning until having received evidence that this is possible, it is necessary to assume that this type of ambiguity took place.

It is not so easy to identify correspondingly ambiguous contexts in our case. We would need to find a context where *is* can be interpreted either with or without its copula function. We can imagine the following possibilities.

1. ***What*-deafness** (pseudocleft misinterpretation)
   In order for pseudoclefts to provide the necessary context, we would have to imagine that hearers consistently fail to process the *wh*-pronoun *what* in pseudoclefts, leaving them essentially with double-*is* constructions. We find it unlikely that consistent processing failure takes place; moreover, this would have the same empirical difficulties as the *what*-drop analysis.

2. **Stuttering**
   Another potential ambiguous context stems from speaker-error: a speaker unintentionally utters a double-*is* construction through stuttering, and a hearer interprets this as intentional. Taking this to be the case would require us to explain what differentiates this case from other cases of stuttering, and why other cases of stuttering do not lead to grammaticalization.

3. **Post-digression copying**
   A third potential ambiguous context would involve performance considerations, though not necessarily error. Speakers may produce an extra copy of the verb after a digression in order to reorient the hearer to his position in the sentence. This phenomenon occurs, and is exemplified in (30).

   (30) What they're, what they're working on **is**, uh, the way I understand it **is** the computer system that, you know, in the future and who knows how far in the future they, they believe that computers will be able to understand human voice where instead of having to have a keyboard ... (Switchboard corpus)

   This notion is subject to the same criticism as the stuttering hypothesis; if this is a mechanism of change, it should take place in every context in which this widely-used technique occurs.

4. **Anacoluthon** [3]
   We find many examples of the type shown in (31) in the Switchboard corpus.

   (31) ... that's what I'm doing, **is** working part time because I can put my kids in a day care situation for a few hours in the week ... (Switchboard corpus)

   (31) can be regarded as an example of anacoluthon; the speaker finishes the sentence as if the *wh*-NP is the subject of the final clause, despite having begun the sentence as if the *wh*-NP is the object of the initial clause. In order for this sentence to be interpreted as grammatical, either the speaker must allow the subject of the sentence to be, itself, a sentence, or the *is* in bold must not be interpreted as a copula. If the latter is the case, *is* must be interpreted merely as a focus particle.

---

[3]*Anacoluthon*: "inadvertent or purposed deviation in the structure of a sentence by which a construction started at the beginning is not followed out consistently." (Smyth, H.W. Greek Grammar, Harvard University Press, 1920, p. 671).

5. **Garden Path effect**

Another possibility is that the speaker uttered a sentence which was genuinely ambiguous, such as (32).

(32)   You know what's really funny, um, is that I get calls, you know.

In this sentence, there are two possible interpretations, (33a) and (33b).

(33)   a.   You know, [what's really funny … is that I get calls …]
       b.   [You know what's really funny], … is that I get calls …

(33a) is standardly grammatical, and *is* has a copula function. In (33b), *is* cannot have a copula function; therefore the hearer is forced to interpret it as a focus marker if it is to be interpretable at all. The interpretation of *is* as a focus marker is licensed by the fact that *is* has become associated with focus when preceded by *wh*-NPs, and followed by *that*-clauses, in pseudoclefts. (33b), moreover, is the first interpretation that the hearer would make, because *know* is normally a transitive verb, so a noun phrase following it would be interpreted as its object.

Of these options, we favor the last one, because it does not rest on the assumption of systematic error, either on the part of the speaker or on that of the hearer. To return to our analogy to Traugott's analysis of *going to*, examples like (32) correspond to examples like (29).

3.1.3   Step B: Autonomy

In Traugott's example, we see autonomy when *going to* is used in contexts where an interpretation involving motion is unavailable, as in in examples like *going to be upset* or *going to faint*. The survival of this usage depends on the existence of a functional motivation for it. The meaning expressed by *going to* is arguably indistinguishable from that of *be to*, so its introduction does not at first glance appear to fill a functional gap; what it accomplishes could already be done. However, the *going to* construction has the advantage over the *be to* construction that it is not syntactically anomalous; while the *am to* construction involves a recursive IP node (an auxiliary with an IP complement), the *going to* construction involves a main verb with an IP complement. This regularization explains why *going to* survived as a future marker.

Autonomy takes place in our case when *is* is used in a context where it cannot be interpreted as a copula without sacrifice of grammaticality, such as (34).

(34)   That's what I said is that I hate you.

The survival of be_2 as a focus marker can also be explained functionally. Both its use in sentential subject constructions (see below) and its use in double-*is* constructions allow for more efficient communication in spoken language.

Examples like (21)-(23) provide one source of functional motivation for the survival of be_2 as a focus marker. These examples not only have the focus function of a pseudocleft, they also link the topic to the discourse via the demonstrative pronoun, which can only refer to a known discourse entity. These two functions were previously incompatible because the topic in a pseudocleft, which necessarily takes the form of a *wh*-NP, cannot be linked to the discourse. Having be_2 as a focus particle allows the pseudocleft-type focus to be expressed simultaneouly with discourse-linking.

At first glance, double-*is* constructions appear to perform a function that is already available; they express the same kind of focus as pseudocleft constructions do; (35a) is hardly distinguishable, functionally, from (35b).

(35) a.   What the thing is is that my mother hates him.

     b.   The thing is is that my mother hates him.

However, examples like (35a) are not attested in the Switchboard corpus. A corpus search revealed that normal pseudoclefts of the structure [*what* [NP] *is*], where [NP] is not a pronoun, are rare; only one example was found:

(36) I think what one of their problems is, is they, they get ...

In all other examples of normal pseudoclefts of the form [*what* [NP] *is*], [NP] takes the form *that*, *there*, or *it*. All such examples are given in (37)-(40)

(37) ...and what **it** is, is, um, normal pizza but then with spinach and broccoli.

(38) I think what **it** is, is everybody [yawning], ...

(39) Okay, well what **this** is, is, it's, it's really sponsored by the Y ...

(40) No, what **there** is, is very high property taxes because ...

    This indicates the existence of a processing constraint against pseudoclefts of this form. Despite its grammaticality, this form appears not to be available to speakers in spoken language. Hence, speakers have a functional gap: when the topic is a full NP instead of a pronoun, pseudoclefting is not available to focus the comment. It is this gap that the double-*is* construction fills. This fact explains why sentences like (35b) are the most common examples of double-*is* constructions. In particular, this explains why pronouns are never the subject of double-*is* constructions; the pseudocleft is available to express the same meaning.

    Just as *going to* survives as a grammaticalized future marker on the basis of its functionality, so be_2 survives as a grammaticalized focus marker.

3.2   Pseudoclefts with zero copula

One interesting prediction of this analysis of *is* is the existence of the following phenomenon. If *is* can be analyzed as a focus marker with no copula function, then an interpretation of pseudoclefts wherein that analysis is given to *is* becomes available. We posit that the *wh*-NP constituting the subject of a pseudocleft is interpreted as a topic phrase, left-adjoined to the sentence, and the *is* thereby becomes optional.

(41) So, what they're doing, they're counting for cash is the way they're ...

(42) And you'd go to court and you'd see these kids, and then, and **what they did, they usually hired kids to do this** because you would get a, you know, a lighter sentence or no sentence or probation or, you know, something trivial.

(43) You know, um, so, I think there, I, **what I learned from them there was a lot of resentment towards the Americans**, so, and it was, like they were Puerto Rican, and were Americans.

(44) Yeah, because, uh, I think **what they say that they'll take your name out of the hopper**, so that whatever is left in the hopper, you know, it's, it's, that's when you'd be tested and then when you were tested, you would be taken out of there.

4 Problems

4.1 Tense agreement

In §2.3, we suggested that *was* is a focus particle, with a tense constraint. While this analysis accounts for the data, it is problematic in that tensed particles are typologically rare. If an analysis could be found that does not require positing a tensed particle, this would be preferable.

On the other hand, if the use of *was* eventually drops out in favor of *is*, the particle analysis will be supported; a system with a tensed particle is unstable, and will change to eliminate the instability.

4.2 Co-occurrence with pseudoclefts

As David Beaver pointed out, the use of double-*is* constructions with pseudoclefts in such examples as (45) is mysterious under the assumption that be_2 performs the same discourse function that pseudoclefts do. In other words, how could (45) ever be produced, when the corresponding pseudocleft (46) is available?

(45) What I'm saying is is that I hate you.

(46) What I'm saying is that I hate you.

We have three potential ways of dealing with this objection:

1. be_2 does not perform the same function as pseudoclefts do.

2. be_2 strengthens the focus already given by the *wh*-NP.

3. be_2 is redundant; it expresses the same discourse function.

We see all of these as possibilities, and do not know how to choose from among them.

References

[1] Bolinger, Dwight. 1987. The Remarkable Double IS. *English Today* 9, 39-40.

[2] Byrne, Francis and Donald Winford, eds. 1993. *Focus and Grammatical Relations in Creole Languages*. John Benjamins Publishing Company.

[3] Givón, Talmy. 1990. *Syntax: A Functional-Typological Introduction*, Volume II. John Benjamins Publishing Company.

[4] Hopper, Paul and Elizabeth Traugott. 1993. *Grammaticalization*. Cambridge University Press.

[5] Koontz-Garboden, Andrew. 2001. LFG problem set: FOCUS and TENSE in the *thing is* construction. Unpublished manuscript, Stanford University.

[6] Massam, Diane. 1999. *Thing is* constructions: the thing is, is what's the right analysis? *English Language and Linguistics* 3.2: 335-352. Cambridge University Press.

[7] McConvell, Patrick. 1988. *To be* or double *be*? Current changes in the English copula. *Australian Journal of Linguistics* 8:287-305.

[8] Tuggy, David. 1995. The Thing Is Is That People Talk That Way. The Question Is Is Why? In: Eugene H. Casad (ed.), *Cognitive Linguistics in the Redwoods: The Expansion of a New Paradigm in Linguistics*, 713-752. Berlin: Mouton de Gruyter.