The contrastive topic requirement on specificational subjects

DANIEL MILWAY
University of Toronto
dan.milway@gmail.com

Abstract

This paper offers a discourse-pragmatic account of the constraint on indefinite DPs as subjects of specificational copular clauses (*a doctor is Mary*). Building on Mikkelsen's (2004) proposal that specificational subjects are topics, I argue that they must be contrastive topics which *properly* contain F-marked constituents. I show that this can account for the absolute ban on simple indefinite subjects, and allow for more complex indefinites to be subjects. Finally, I discuss the syntactic analysis that would be predicted given my pragmatic analysis, and the puzzles that arise from it.

Keywords: Specificational copular clauses, Topic, Indefinite subjects, Pragmatics

Résumé

Cet article offre une analyse de type discours-pragmatique de la contrainte sur les DP indéfinis qui sont sujets de clauses copulaires spécificationelles (*un médecin est Marie*). En partant de l'argument de Mikkelsen (2004), qui avance que les sujets des clauses spécificationnelles seraient topiques, je propose qu'on les analyse en tant que topiques contrastifs qui *incluent* un constituant F marqué. Je démontre que cette analyse explique l'interdiction absolue sur des sujets indéfinis simples, tout en permettant des sujets indéfinis plus complexes. Pour clore, je discute de l'analyse syntaxique prédite par mon analyse pragmatique, et des énigmes qu'elle soulève.

Mots clés: clauses copulaires spécificationelles, topique, sujets indéfinis, pragmatique

1. INTRODUCTION

Specificational clauses form one of the classes of copular clauses identified by Higgins (1973), characterized by a predicative DP (type $\langle e, t \rangle$) in subject position (DP1) and a

referential DP (type e) in post-copular position (DP2).[1] In contrast, predicational copular clauses are a class in which DP1 is referential and DP2 is predicative.[2]

(1)   a.   **Specificational**
           [My favourite book]$_{DP1}$ is [*War and Peace*]$_{DP2}$.

      b.   **Predicational**
           [*War and Peace*]$_{DP1}$ is [my favourite book]$_{DP2}$.

This description and these examples suggest an analysis according to which predicational clauses and specificational clauses (SCs) are inverses of each other.[3] One such analysis by Moro (1997) (hereafter referred to as *the inversion analysis*) refers to predicational clauses as *canonical copular clauses*, and to specificational copular clauses as *inverted copular clauses*. One of the hurdles for this analysis is that the set of DPs that can be specificational subjects is not coextensive with the set of DPs that can be DP2 in predicational clauses. As demonstrated in (2), for instance, while indefinite DPs can freely be DP2 in predicational clauses, their ability to be SC subjects is restricted.

(2)   a.   **Specificational**
           *A book is *War and Peace*.

      b.   **Predicational**
           War and Peace is a book.

This restriction on indefinite SC subjects, hereafter referred to as *the indefinite restriction*, which this paper addresses,[4] presents a puzzle for any syntactic or semantic analysis of SCs because it is not an absolute ban on indefinite DPs in subject position. Rather, the fact that some indefinite DPs can act as SC subjects, as demonstrated in (3), means that, before we can adduce the indefinite restriction as evidence for or against a particular analysis of SCs, we must first understand the provenance of the restriction.

(3)   a.   i.   * A doctor is Mary.

           ii.   A newly-minted doctor is Mary.

      b.   i.   * A linguist is Eric Lenneberg.

           ii.   An underrated linguist is Eric Lenneberg.

---

[1]See also Romero (2005) for a different semantic analysis.

[2]Higgins (1973) identifies two other classes of copular clauses: Identificational clauses like (1) in which DP1 is a demonstrative and DP2 is a referential nominal, and equational clauses like (2), in which both DP1 and DP2 are referential nominals.

(i)   That (book) is *War and Peace*.

(ii)   *War and Peace* is *Vojna i mir*.

[3]Abbreviations used: CT: contrastive topic; CT-F: contrastive topic with focus; d-tree: discourse-tree; Exh: exhaustive focus; F: focus; LF: logical form; QUD: question under discussion.

[4]It is also discussed by Halliday 1967, Higgins 1973, Heggie 1988, Heycock 1994, Williams 1997, and Mikkelsen 2005 among others.

    c.   i. * A building is Robarts.

         ii. A building no-one likes is Robarts.

In the remainder of this paper, I develop and defend a hypothesis that the indefinite restriction is pragmatic in nature. Specifically, I propose that SC subjects are contrastive topics (CT) in the sense of Büring (2003) and Constant (2014). That is, they are topics that contain F(ocus)-marked constituents. I further propose that the F-marking in SC subjects cannot apply to the entire DP. It is important to note that this claim is not only about *indefinite* SC subjects, but SC subjects in general. As such, I will present evidence that definite DPs also meet this requirement. In section 3, I introduce some of the theoretical machinery required for my analysis and in section 4, I present my main claim and the arguments in its favour. Finally, in section 5, I discuss some residual issues and conclude.

1.1 A note on the use of acceptability judgement markers

The asterisk (*) and the hash mark (#) are commonly taken to indicate syntactic and semantic/pragmatic deviance, respectively. This paper strays from this convention slightly.

Since I will be proposing a discourse-pragmatic analysis in this paper, it will be useful to be able to differentiate between expressions which are *per se* deviant and those that are deviant in a given discourse context but otherwise acceptable. To this end, I use the asterisk (*) to mark expressions as *per se* deviant and the hash mark (#) to mark expressions that are deviant in a given discourse context. All other judgement markers are used according to convention.

2. PREVIOUS ANALYSES OF THE INDEFINITE RESTRICTION

The baseline for the current debate about the nature of the indefinite restriction seems to be the *equational analysis* of Heycock and Kroch (1999), which states, contra Moro (1997), that both DPs in an SC like (1a) are referring expressions, and that the specificational copula serves to identify the referent of DP1 with the referent of DP2. Since (non-specific) indefinites are non-referential, this analysis predicts that DP1 of an SC cannot be a (non-specific) indefinite.

Mikkelsen (2005) notes that the version of the indefinite restriction that Heycock and Kroch (1999) predict is a categorical ban on non-specific indefinite SC subjects.

The prediction of such a categorical ban, Mikkelsen argues, is falsified by SCs such as (4–7) which have indefinite subjects and yet are judged acceptable.

(4) **A philosopher who seems to share the Kiparskys' intuition on some factive predicates** is Unger (1972) who argues that … [5]

---

[5]Delacruz (1976: 195 fn8) cited by Mikkelsen (2005: 117)

(5) **Another speaker at the conference** was the *Times* columnist Nicholas Kristof, who got … [6]

(6) **One Iraqi émigré who has heard from the scientists' families** is Shakir al Kha Fagi, who left Iraq as a young man and runs a successful business in the Detroit area.[7]

(7) **A doctor who might be able to help you** is Harry Barcan.[8]

Since the restriction is not categorical, she argues, it is not due to a semantic type mismatch, as Heycock and Kroch (1999) propose.

Instead, Mikkelsen argues that the indefinite restriction is pragmatic in nature. She points out (following, among others, Halliday 1967) that, unlike predicational clauses, SCs have a fixed information structure. As demonstrated in (8), SCs are infelicitous in contexts that focus the initial DP, while predicational clauses are more flexible.

(8) a. Q: Who is the winner?
         A1: The winner is JOHN.   [Specificational]
         A2: JOHN is the winner.   [Predicational]

     b. Q: What is John?
         A1: #The WINNER is John.   [Specificational]
         A2: John is the WINNER.   [Predicational]
                                                              (Mikkelsen 2005: 195)

Here Mikkelsen uses Wh-question-and-answer pairs as diagnostics for focus (F). In the answer to a Wh-question, the constituent that "replaces" the Wh-word must be focused. So, the question in (8a) requires an answer in which *John* is focused. The fact that both predicational and specificational clauses are felicitous answers here suggests that focus is compatible with both DP2 of SCs and DP1 of predicational copular clauses. The question in (8b) on the other hand, requires an answer in which *the winner* is focused. The fact that only a predicational clause is a felicitous answer here suggests that focus is compatible with DP2 of predicational clauses and, more importantly, that focus is incompatible with DP1 of SCs.

Mikkelsen argues that this fixed information structure of SCs follows from SCs being inversion structures. Following Birner (1994, 1996), she assumes that the discourse function of inversion is to mark the inverted material as linking a clause to previous discourse. The inverted material, then, must be more discourse-familiar than the post-verbal logical subject. Mikkelsen then argues that these discourse familiarity considerations can explain the acceptability of (4)–(7). Specifically, she argues that the additional material (i.e., the adejctives, PPs, and relative clauses) in the subjects of (4)–(7) serve to link those subjects to prior discourse, thus ensuring their topicality. Informally, then, Mikkelsen's argument is that "heavier" indefinites

---

[6]Seymore M. Hersh "The Stovepipe", *The New Yorker*, Oct 27, 2003: 86 cited by Mikkelsen (2005: 118)

[7]Seymore M. Hersh "The Stovepipe", *The New Yorker*, Oct 27, 2003: 86 cited by Mikkelsen (2005: 118)

[8]Mikkelsen (2005: 118)

make better topics and more topical DPs make better SC subjects, therefore "heavier" indefinties make better SC subjects.

Mikkelsen's proposal suggests that the acceptability of indefinite SC subjects should be gradient and context-dependent. However, the restriction on simple indefinite SC subjects seems to be categorical and context-independent.

That is, even if the material in a simple indefinite is familiar, as in (9), the indefinite cannot be the subject of an SC.

(9) Bill is a doctor. #A doctor is John (too).

Mikkelsen suggests that this is due to a requirement that indefinites introduce novel discourse referents. This requirement, called the Novelty Condition (Heim 1982), clashes with the proposed topicality condition on SC subjects. Mikkelsen points out, however, that this cannot be the entire story, since the Novelty Condition is the requirement that indefinites introduce new discourse referents, while the topicality requirement is a condition on the familiarity of the descriptive content of the DP. Since the two instances of *a doctor* in (9) do not share a discourse referent, the Novelty Condition does not rule out the indefinite subject.

Mikkelsen also suggests that those instances of familiar yet unacceptable simple indefinite SC subjects might be infelicitous because there is a general ban on repeating indefinites, as in the example below.

(10) Sally is a doctor. #A doctor lives next door.

This, however, does not seem to hold. Utterances, such as (10), that are barred because of repeated indefinites are made better if the first occurrence of the indefinite is modified. If the barred utterance has an SC with an indefinite subject, as in (9), then only changing the SC will improve it.

(11) I know many doctors.

     a. # A doctor is Patrick.

     b. A doctor lives next door.

To sum up, Mikkelsen observes that there seems to be a requirement that SC subjects be topical. She attempts to use this requirement to explain the restriction on indefinite subjects, arguing that topics must be given, while indefinites tend to be novel, so indefinites are not good topics and, as a corollary, indefinites tend to make poor SC subjects. She notes, however, that this account runs into a problem in that even when simple indefinites can be made topical, they cannot be SC subjects.

Addressing the indefinite restriction and responding in part to Mikkelsen's analysis, Heycock (2012) begins with the information structure pattern shown in (8), which she frames as a restriction on focusing SC subjects. She notes that this is parallel to a fact about scrambling in German observed by Lenerz (1977), reproduced in (12).

(12) a. Wem   hat Peter das   Futter gegeben?
        who.DAT has Peter the.acc food   given
        'To whom has Peter given the food?'

    i. Peter hat der   Katze das   Futter gegeben.
       Peter has the.DAT cat   the.ACC food given
       'Peter has given the cat the food.'   [Default order]

    ii. Peter hat das   Futter der   Katze gegeben.
       Peter has the.ACC food the.DAT cat   given
       'Peter has given the food to the cat'   [Scrambled order]

  b. Was   hat Peter der   Katze gegeben?
     what.ACC has Peter the.DAT cat   given
     'What has Peter given (to) the cat?'
      i. Peter hat der   Katze das   Futter gegeben.
         Peter has the.DAT cat   the.ACC food given
         'Peter has given the cat the food'   [Default order]

      ii. #Peter hat das   Futter der   Katze gegeben.
         Peter has the.ACC food the.DAT cat   given
         'Peter has given the food to the cat'   [Scrambled order]

What (12) demonstrates is that the canonical order for ditransitive objects in German is DAT ≺ ACC). The scrambled order, ACC ≺ DAT, is unavailable when the accusative argument is focused, as shown in (12b-ii). Just as SCs subjects cannot be focused in English, scrambled objects cannot be focused in German.

    With this information structure parallel established, Heycock (2012) attempts to argue that the parallel between English SC subjects and German scrambled objects is due to their parallel semantic properties.

    Following de Hoop (1992) and Diesing (1992), Heycock assumes that scrambled DPs in German are necessarily interpreted as strong DPs. She claims that SC subjects are also restricted to strong interpretations. As evidence for this claim, she presents another parallel. A property of weak indefinites, according to Milsark (1974), is that they cannot serve as subjects of Individual-Level predicates, as shown in (13).

(13) I had been struggling with a complicated set of data …

    a. ?* A problem was particularly hard.

    b. One problem was particularly hard.

    c. {?A/one} problem that I came across was particularly hard.

    d. One of the problems was particularly hard.   (Heycock 2012)

Heycock argues that the same pattern holds for indefinite SC subjects as shown in (14).

(14) a. ?* A problem was that we didn't understand all the parameters.

    b. One problem was that we didn't understand all the parameters.

    c. {A/one} problem that I came across was that we didn't understand all the parameters.

    d. One of the problems was that we didn't understand all the parameters.

                                           (Heycock 2012)

Given these similarities, Heycock proposes that the indefinite restriction is actually a restriction on *weak* indefinites as SC subjects.

Heycock's use of the terms *strong* and *weak* comes from Milsark (1974), via de Hoop (1992) and Diesing (1992), who distinguishes them on the basis of presupposition triggering. A DP with a strong determiner always carries a presupposition of the existence of the DP's referent, while a DP with a weak determiner is ambiguous between a reading that presupposes existence and a reading that does not presuppose existence. Heycock's proposal, then, amounts to a proposal that SCs must presuppose the existence of their subjects' referent. According to this proposal, the SC in (15) must presuppose the existence of at least one side-effect.

(15) A side-effect of the treatment is baldness.

While (15) does entail the existence of a side-effect, that entailment disappears in the antecedent of a conditional, as in (16).

(16) If a side-effect of the treatment is baldness, then Jo won't recommend it.

Compare this pair to (17) and (18), which both entail the existence of a side-effect, demonstrating the presuppositional reading of the definite SC subjects.

(17) The side-effect of the treatment is baldness.
(18) If the side-effect of the treatment is baldness, then Jo won't recommend it.

Since non-presuppositional indefinites can be SC subjects and only weak DPs have non-presuppositional readings, we can conclude that weak indefinites can be SC subjects, thus falsifying Heycock's proposal.

Heycock (2012), further argues that what Mikkelsen (2005) cites as SCs with indefinite subjects may, in fact, be predicate inversion constructions formed by A' movement. According to Heycock, SCs and inversion structures can be distinguished in English by the agreement of the copula: In SCs, the copula agrees with the first DP, as in (19), while in inversion constructions, the copula agrees with the DP to its right, as in (20).

(19) My favourite team {is/*are} the Maple Leafs.
(20) A threat to the Habs' dominance {*is/are} the Maple Leafs.

However, using Heycock's diagnostic for SCs, we can construct SCs with indefinite DP1s. Consider the examples (21) and (22), each presented with their corresponding predicational clause, for contrast.

(21) A committee that I'd hate to present to is Sue, Jerry, and Alex.
        (Sue, Jerry, and Alex are a committee that I'd hate to present to.)
(22) A team I've cheered for all my life is the Maple Leafs.
        (The Maple Leafs are a team I've cheered for all my life.)

So it seems, contra Heycock, that weak indefinite SC subjects are possible.

2.1   Summary

Each of the two approaches to explaining the indefinite restriction reviewed in this section has its own issues. The pragmatic approach of Mikkelsen (2005) covers a greater portion of the data but lacks a precise and cohesive account of it. The semantic approach of Heycock and Kroch (1999) and Heycock (2012) is more precise, but at the expense of its empirical coverage.

In the following sections I will argue that the solution to the indefinite restriction must be that indefinite SC subjects contain contrastive topics which properly contain F-marked constituents. This will be shown to have greater empirical coverage than the accounts addressed in this section.

3.   THEORETICAL BACKGROUND

What I refer to as contrastive topic (CT) here is closely related to what Jackendoff (1972) refers to as the B-accent in his now classic examples, reproduced here in (23).

(23)   a. (What about FRED? What did HE eat?)
           FRED$_B$ ate the BEANS$_A$.

       b. (What about the BEANS? Who ate them?)
           FRED$_A$ ate the BEANS$_B$.   (Jackendoff 1972: 261)

Jackendoff identifies the A and B pitch accents with a falling contour and a rise-fall-rise contour, respectively, and addresses their discourse pragmatics. Since Jackendoff's work, there has been research on the pragmatics, semantics, syntax, and prosody of these phenomena, some of which I outline in this article. Büring (2003), building upon Rooth (1992) and Roberts (2012), develops formal models of the semantics and pragmatics of contrastive topics in terms of alternative semantics and d(iscourse)-trees, respectively. I describe these models in section 3.1. Constant (2014) develops a syntactic account of Büring's models which I discuss in section 3.2.

3.1   The discourse pragmatics of contrastive topics (Büring 2003, 2016)

The analysis of contrastive topics developed by Büring (2003) has two components: Alternative semantics and d-trees. The first component is essentially an extension of Rooth's (1992) theory of focus. According to Rooth, in addition to ordinary interpretations ($[\![\cdot]\!]^{\mathcal{O}}$), sentences with focused elements receive a focus interpretation ($[\![\cdot]\!]^{f}$)—a set of alternatives which is derived from the ordinary interpretation and the focused constituent. For instance, (26) has the ordinary interpretation (27), and the focus-value (28), where the formeer is the proposition expressed by the sentence, and the latter is the set of propositions generated by replacing the focused material with a variable.

(24)   [Mary]$_F$ answered Sue.

(25) $[\![24]\!]^{\mathcal{O}} = [answered\,(\mathbf{m}, \mathbf{s})]$

(26) $[\![24]\!]^{f} = \{answered\,(x, \mathbf{s})|x \in D_e\}$

Note that the focus-value of (24) is equivalent to the ordinary interpretation of the question *Who answered Sue?* following Hamblin (1973).

Büring, following Roberts (2012), extends this analysis to contrastive topics, proposing that an expression with a CT-F structure (i.e., an expression with both focus and contrastive topic) not only have an ordinary interpretation and a focus-value, but also a CT-value, which is generated from the expression's focus-value. The CT-value of a sentence is formed following the procedure in (27).

(27) CT-value formation:

   step 1: Replace the focus with a *wh*-word and front the latter; if focus marks the finite verb or negation, front the finite verb instead.

   step 2: Form a set of questions from the result of step 1 by replacing the contrastive topic with some alternative to it.

(Büring 2003)

Note, as demonstrated below, this generates a set of questions, which is a set of sets of propositions. In the case of (28a), the CT-value is equivalent to the interpretation of a multiple-*wh* question (*Who ate what?*).

(28) a. [Hilary]$_{CT}$ ate [bagels]$_F$.

   b. CT-value formation:
   step 1: What did Hilary eat?

       step 2: $\left\{ \begin{array}{l} \text{What did Hilary eat?} \\ \text{What did Robin eat?} \end{array} \right\}$

   c. $[\![$Hilary$]\!]_{CT}$ ate [bagels]$_F.]\!]^{ct} = \{\{x \text{ ate } y | y \in D_e\} | x \in D_e\}$

Büring proposes that the CT- and F-values of a sentence represent the question and a subquestion, respectively, such that the sentence answers the subquestion and partially answers the question. That is, the interpretation of a sentence with a CT-F structure projects a discourse structure that Büring represents in the form of a d (iscourse)-tree. For instance, d-tree projected by (28a) is given in (29).

(29)
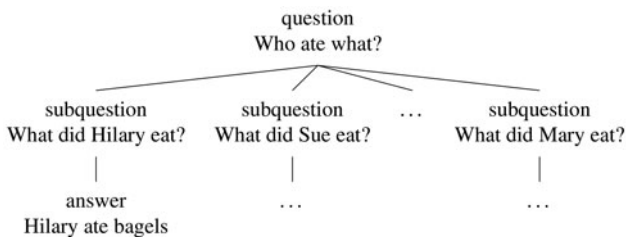
Note that the CT- and F-marking in (28a) are swapped, as they are in (30a), a different d-tree, namely (31) is projected.
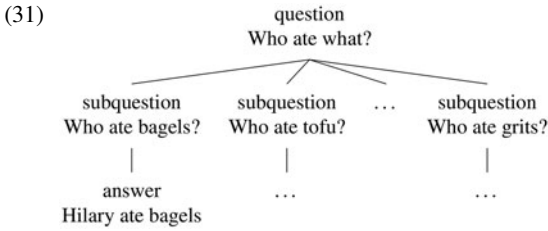
(30) a. [Hilary]$_F$ ate [bagels]$_{CT}$.

   b. CT-value formation:
       step 1: Who ate bagels?

       step 2: $\left\{ \begin{array}{l} \text{Who ate bagels?} \\ \text{Who ate tofu?} \end{array} \right\}$

   c. $[\![$[Hilary]$_F$ ate [bagels]$_{CT}.]\!] = \{\{x \text{ ate } y | x \in D_e\} \ y \in D_e\}$ ($\neq [\![(30a)]\!]^{ct}$)

(31)
                              question
                            Who ate what?

       subquestion   subquestion   . . .   subquestion
       Who ate bagels?   Who ate tofu?   Who ate grits?
            | | |
         answer   . . .   . . .
   Hilary ate bagels

D-trees provide a perspicuous way of representing various aspects of discourse structure in a way that leverages a vocabulary already used by generative linguists. They allow us to define pragmatic notions such as assertions, questions, alternatives, etc. in terms of nodes, sisterhood, dominance, etc. For instance, assertions and questions are distinguished by the fact that the former are terminal nodes while the latter are non-terminal.

It should be noted that CT-F structures are used in a variety of discourse contexts to achieve subtly different conversational goals. Consider the following examples.

(32) A: When are you going to China?
      B: I'm going to [China]$_{CT}$ in [April]$_F$.   (Roberts 2012)

(33) A: What did the pop stars wear?
      B: The [female]$_{CT}$ pop stars wore [caftans]$_F$.   (Büring 2003)

(34) A: Who's a good psychiatrist?
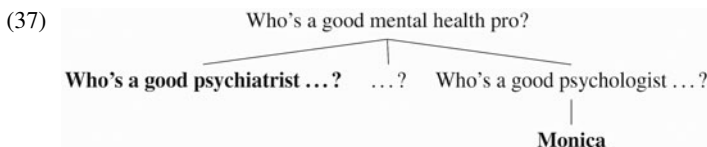   B: [My sister Monica]$_F$ is a [psychologist]$_{CT}$.

All of these instances of CT-F structures signal what Büring calls *implicit moves*. Each instance has a different sort of implicit move that can be easily represented by its d-tree. In (32) the assertion B directly answers the question A, but implies the existence of a relevant superquestion (*When are you going to which place?*). The d-tree in (35) shows this by marking the explicit moves in bold.

(35)   When are you going which place?

      When are you going to . . . ?   **When are you going to China?**
                                                    |
                                                 April
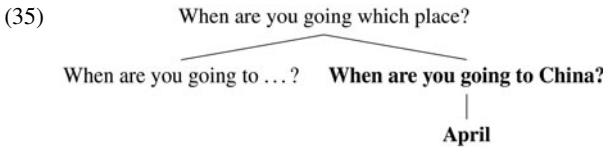
The assertion B in (33), on the other hand, does not answer the explicit question A, but instead answers an implied subquestion (*What did the female pop-stars wear*). Again this can be represented clearly in the d-tree in (36).

(36)   **What did the pop stars wear?**

      What did the male pop stars wear?   What did the female pop stars wear?
                                                            |
                                           The female pop stars wore caftans.
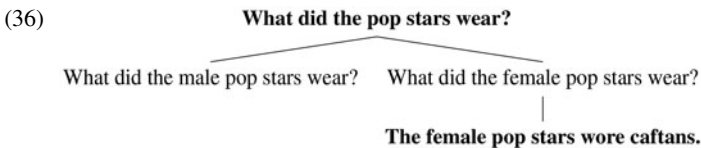
Finally, the assertion B in (34) answers neither the explicit question A, nor an implied subquestion. Instead, it answers an implicit subquestion of a superquestion of the explicit question, as we can see in its d-tree in (37).
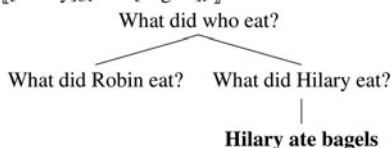
(37)

Who's a good mental health pro?

Who's a good psychiatrist …?   …?   Who's a good psychologist …?

Monica

So, although a given CT-F structure can be mapped onto a single d-tree in a predictable way, the context in which it is uttered determines its place in and effect on the discourse. Implicit in Büring (2003) is an informal condition on CT felicity which I give in (38).
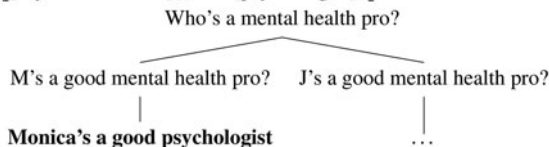
(38) S sentence that uses a CT-F structure.
      Q is a question.
      S is felicitous in the context of the QUD Q iff the S projects a d-tree DT such that Q is represented in DT.

Though informal, this condition can effectively rule out several examples of infelicitous CT-F structures. The infelicity of the CT-F structures in (39) and (40) is predicted by the fact that the explicit question that they answer is not found in the d-trees they project.

(39) a. A: Who ate bagels?
          B: #[Hilary]$_{CT}$ ate [bagels]$_F$.

      b. $[\![$[Hilary]$_{CT}$ ate [bagels]$_F$$]\!]^{ct}$
          What did who eat?

          What did Robin eat?   What did Hilary eat?

          Hilary ate bagels

(40) a. A: Who's a good psychiatrist?
          B:# [My sister Monica]$_{CT}$ is a [psychologist]$_F$

      b. $[\![$[My sister Monica]$_{CT}$ is a [psychologist]$_F$$]\!]$
          Who's a mental health pro?

          M's a good mental health pro?   J's a good mental health pro?

          Monica's a good psychologist   …

3.2   The Topic Abstraction analysis of CT (Constant **2014**)

In his thesis, Constant (2014) proposes and argues for a comprehensive revision of Büring's theory of contrastive topic. This revision includes a more nuanced analysis

of the syntax of CTs and a more precise description of the prosody associated with CTs in English,[9] which I will outline in turn in this section.

Constant analyzes the pitch contour in terms of Pierrehumbert and Hirschberg's (1990) ToBI formalism[10] and shows that the the characteristic rise-fall-rise contour of CTs is analyzable as a pitch accent (L+H*) followed by a low phrase tone (L-) and a high boundary tone (H%), as shown in Jackendoff's (1972) classic examples reproduced here as (41) and (42).

(41) A: What about FRED? What did HE eat?
      B: FRED …   ate the beans.
          L+H* L-H%

(42) A: What about the BEANS? Who ate THEM?
      B: Fred ate the BEANS …
                          L+H*   L-H%

Constant further notes that the pitch accent and boundary tones are associated to different things. The pitch accent, he argues, is associated with an F-marked constituent, while the boundary tone is associated with the right edge of the phrase that contains the F-marked constituent. This can be seen in (43) and (44), where the placement of the L + H* accent depends on the discourse context, while the L-H% boundary tone is associated with the edge of the DP.

(43) A: What did the singers wear?
      B: The   FEMALE singers … wore caftans.
              L+H*   L-H%

(44) A: What did the female performers wear?
      B: The female   SINGERS   …   wore caftans.
                        L+H*   L-H%

Based on this pattern (and other reasons), Constant proposes that what Büring calls CT-marking is identical to F-marking, and the distinction between CT and exhaustive focus (hereafter Exh, following Constant) is due to the structural configuration of those phrases that contain F-marked constituents.

In order to capture the distinction between CT and Exh, Constant proposes an operator in the left periphery, CT-λ whose specifier is interpreted as a CT. So, CT phrases are raised, sometimes covertly, to the left periphery (in the sense of Rizzi 1997) and the CT-λ operator cliticizes to the intonational phrase, yeilding the L-H% boundary tone. The proposed LF structure of (42) then, is given in (45), where a dashed arrow indicates covert movement.

---

[9]The thesis also includes a proposed semantics CTs and an analysis of the Mandarin discourse particle -*ne* as a CT marker. These topics, however, are beyond the scope of this paper, so I will not address them.

[10]Constant (2014: 14–16) provides a succinct description of the ToBI system, so rather than reproduce that description, I encourage interested readers to seek out this portion of the thesis and works cited therein.

(45)
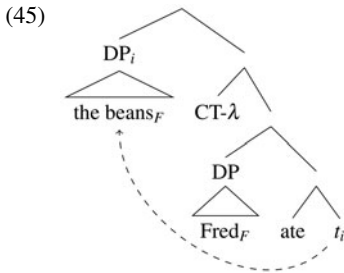
Topicalization, as in (46), then occurs when CT abstraction is overt, according to Constant.

(46)   A: What about the BEANS? Who ate THEM?
        B: The BEANS … Fred ate.
           L+H* L-H%

3.3   Summary

In this section, I have outlined some basic properties of CTs which will be useful in the discussion of SC subjects. In semantico-pragmatic terms, CTs are interpreted as a nested set of alternatives, which imply a complex discourse structure. That is, if an utterance without a CT indicates a question-answer move in the discourse (either asking a question and expecting a complete answer, or giving a complete answer to a question under discussion), then an utterance *with* a CT indicates a question-subquestion-answer strategy. In syntactic terms, a CT is a phrase which (*i*) has an F-marked constituent, and (*ii*) is generated in, or moves, often covertly, to the specifier of a a phrase projected by a CT operator (CT-λ). In English, the F-marking in a CT is typically realized as a rising pitch accent (L+H*) and the CT-λ operator, which cliticizes to the phrase in its specifier, is typically realized as a rising boundary tone (L-H%).

    Before continuing, I will pause here to consider the evidential status and utility of these three properties. First, I will consider the syntactic analysis of CTs that Constant (2014) develops. Evidence for this analysis, as with all syntactic analyses, is mostly indirect and inferential and, therefore, the analysis does not have very much diagnostic utility. That is, while Constant's syntactic analysis will help to assess the *theoretical* plausibility of my final analysis of SCs, it likely won't be particularly useful in my analysis of any particular examples. Next, I will consider the prosodic properties of English CTs — the rising pitch accent followed by a rising boundary tone. This seems to be a widely accepted and reliable diagnostic — when we find an L+H*L-H% contour in English we also find a CT. This endorsement of the prosodic diagnostic, however, comes with two reservations, the first being that the diagnostic is English-specific, and the second being that diagnostics tend to be one-way conditional inferences. Even if we assume that an L+H*L-H% contour implies a CT, it doesn't follow that the absence of an L+H*L-H% contour implies the absence of a CT. Finally, the discourse-pragmatic properties of CTs: a CT indicates a question-subquestion-answer strategy. While the general unruliness of discourse context severely limits its usefulness as a diagnostic, it does not share the shortcomings of our prosodic diagnostic. That is, it is not English-specific, and it seems to

allow a two way inference — if we wanted to identify CTs in an arbitrary language, discourse context would be the most reliable indicator by far. This paper, of course, is restricted to English, so prosodic cues will be useful diagnostics, but interaction with discourse context will be the ultimate indicator of the presence of a CT.

4.   THE CONTRASTIVE TOPIC REQUIREMENT ON SC SUBJECTS

I am now prepared to modify Mikkelsen's (2005) analysis of SCs so that it properly captures the indefinite restriction. Recall that Mikkelsen argued that SCs have a fixed information structure, with the postcopular DP being focus and the subject being topic, as shown in (47) below, and that for Mikkelsen, topicality requires discourse familiarity.

(47)   [My favourite singer]$_{\text{Top}}$ is [Ian]$_F$.

I propose that SC subjects must be *contrastive* topics, in the sense of Constant (2014), but must not be wholly F-marked. I will show, in the remainder of this section, that this addition to Mikkelsen's analysis effectively captures the indefinite restriction. Specifically, requiring SC subjects to properly contain an F-marked constituent will account for the fact that more complex/heavy indefinites (such as those in (4–7)) are more likely to be acceptable SC subjects, as well as for the fact that simple indefinites are almost never allowed as SC subjects.

(48)   **The contrastive topic requirement on specificational clauses**
A clause of the form *X* BE *Y* is a licit specificational clause iff

   a. $[\![X]\!]$ ($[\![Y]\!]$) is defined,

   b.   *Y* is an exhaustive focus, (Mikkelsen 2005)

   c.   *X* is a contrastive topic, and

   d.   An F-marked constituent is *properly* contained by *X*.
(i.e., *X* itself is not the F-marked constituent)

In the above definition, (48a) restricts the requirement to possible SCs, and (48b) incorporates Mikkelsen's observation of the fixed information structure of SCs. The final two parts of the requirement, (48c) and (48d) are what I will argue for in the following two sections.

I have framed this proposal as a condition on SCs in general, rather than one on indefinite subects of SCs, for reasons of parsimony. While indefinite subjects play an important role in the discussion that follows, I intersperse SCs with definite subjects for ease of demonstration.

4.1   A note on evidence for and against my proposal

Despite the fact that my analysis of the indefinite restriction is given in terms of discourse pragmatics, the evidence I adduce in its favour does not come from natural or even naturalistic discourse. This is to be expected, given the formal nature of the theories that form the basis of the proposed analysis. Take, for instance Büring's d-trees and the theory of discourse underlying them. This family of theories models a given

discourse as a series of questions starting at some universal superquestion such as (49) and ending at the answer to a polar question such as (50).

(49) What is the world like?

(50) Did Rahim eat a bagel?

If a model of this sort were applied to natural discourse, it would perform quite poorly, because natural discourse is nowhere near as methodical and mechanical as the model. Rather, evidence for or against any theoretical proposal is constructed to be demonstrably relevant to the proposal, where "relevant" has a precise meaning here. A piece of data X is relevant to a proposal P only insofar as P makes a definite prediction about the status of X.

Applying this requirement to my proposal, we can say that the relevant data is the set of specificational clauses for which speakers have clear judgments and which demonstrably do or do not meet all of the conditions listed in (48). A demonstrably relevant piece of data where prediction matches reality will be taken as corroborating evidence, while one where prediction and reality conflict will be taken as falsification. Demonstrably irrelevant data, of course, will be ignored. This leaves a class of examples whose (ir)relevance cannot be demonstrated, the data for which the proposal in question's predictions are unclear. Since virtually all progress in a field of empirical inquiry comes from moving data from this class into one of the other two classes, either through reanalysis of the data, or development of new theoretical proposals, this class of data is perhaps the most interesting class. However, since data which is not demonstrably relevant can neither corroborate nor falsify the particular claim of this paper, I will not address it beyond acknowledging that it exists and that its existence renders my proposal incomplete.

4.2   SC subjects must be contrastive topics

The first claim of my proposal that must be justified is that *contrastive* topic-hood, rather than givenness or aboutness topic-hood is the relevant notion for SC subjects. This claim can be further divided into three claims. First, a CT-Exh structure (Previously "CT-F structure", adapted to conform with Constant's analysis.) is a licit information structure for SCs. Second, SC subjects cannot be entirely discourse-given. Finally, SC subjects cannot be aboutness topics. In the following subsection, I will present evidence for each of these claims in turn. Following that, I will address the second claim of my proposal, that SC subjects cannot be wholly F-marked.

4.2.1   CT-Exh structure is compatible with SCs

English SCs are most naturally uttered with intonational stress on some part of their subject as shown in (51).

(51)   a.   A building on campus no-one LIKES is Robarts.

b.   A building on campus NO-ONE likes is Robarts.

c.   A building on CAMPUS no-one likes is Robarts.

d.   A building ON campus no-one likes is Robarts.

e.   A BUILDING on campus no-one likes is Robarts.

f.   * A building on campus no-one likes is Robarts.

English intonational stress is associated with informational prominence, and since, as Mikkelsen shows, DP2 position in SCs is necessarily focused, the intonational stress in the subjects of (51) cannot be primary focus.

More precisely, the intonational stress in SC subjects is a rising pitch accent (L+H*) generally followed by a low phrase tone (L-). To show this I recorded a native speaker of Canadian English[11] saying the SC in (52) in various three different discourse contexts.

(52)   A book I would recommend is *Barometer Rising*.

The first context, designed to target *book* for F-marking, is given in (53), with ellipses indicating the target sentence (52). The resulting intonational contour[12] is given in Figure 1.

(53)   A: I'm looking for a new TV show, can you recommend any?
       B: I don't really watch TV, but …
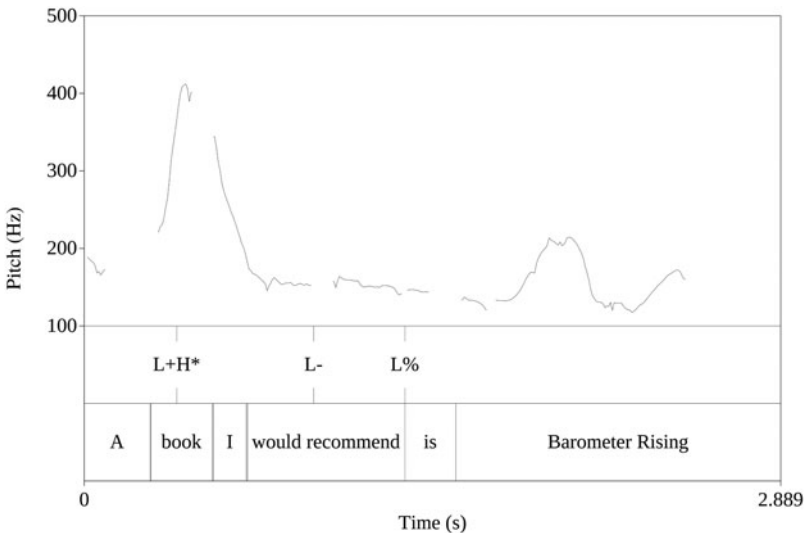
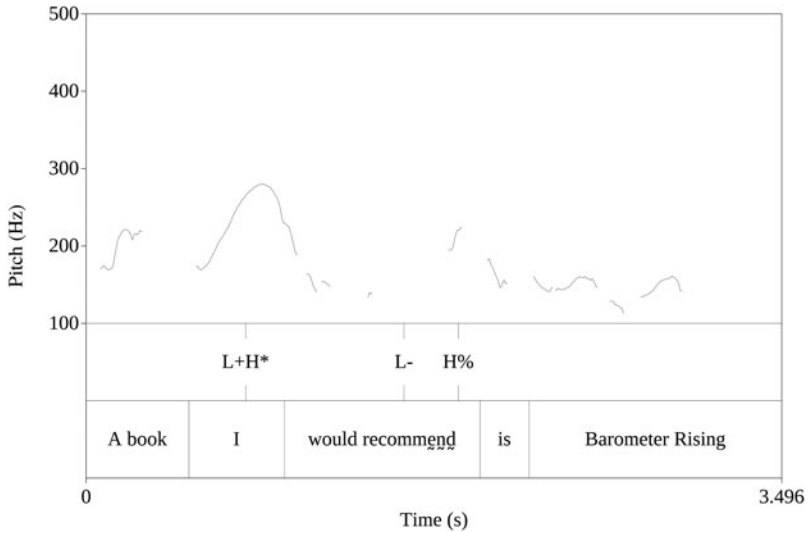**Figure 1:** The intonational contour of (52) in context (53)

The second context, designed to target the embedded subject *I* for F-marking, is given in (56). The resulting intonational contour is given in Figure 2.

(54)   A: I'm looking for a new book. Everyone's telling me to read *Handmaid's Tale*. What do you think?
       B: I didn't really like it. …

---

[11]The speaker is Shawna Edward, a local actor and comedian.

[12]The pitch contour analysis was performed in Praat (Boersma and Weenink 2019). Letters with tildes below them indicate creaky voice.
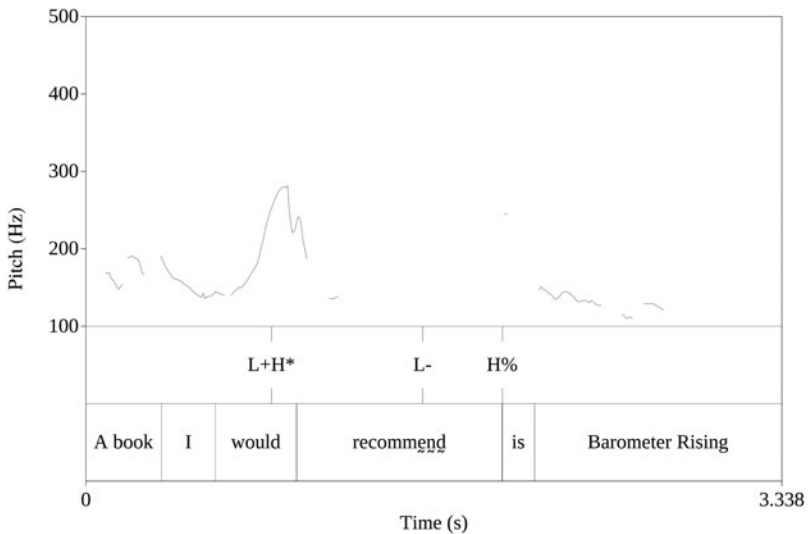
Figure 2: The intonational countour of (52) in context (54)

The third context, designed to target *would* for F-marking, is given in (55). The resulting intonational contour is given in Figure 3.

(55) A: (Hands B a list of books) Which of these would you recommend?
       B: I wouldn't recommend any of those. …

**Figure 3:** The intonational contour of (52) in context (55)

In all of these, we see the characteristic L+H* pitch accent followed by a L-phrase tone of a CT constituent. There is some variation in the presence of a H% boundary tone, that Constant (2014) associates with a cliticized CT-λ head. In the contour shown in Figure 3, we see no rising pitch at the right edge of the SC subject, and furthermore, a second recorded speaker, whose context-contour pairs are given in an appendix (5.2), shows a complete lack of H% tones on SC subjects. I will set this variation aside for now, addressing it briefly in section 5.1. Boundary tones aside, though, the intonational contour associated with SC subjects seems to be that of a CT constituent.

Pragmatically, CT-Exh structures are characterized by association with a complex discourse strategy of a question and subquestion. SCs can indeed be associated with a question-subquestion strategy. Consider the example in (56).

(56) (Not many people like the Athletic Centre.)
     A building on campus NO ONE likes is Robarts.

If DP2 is F-marked as an Exh, and the stressed constituent *no one* is CT F-marked, then we can use Büring's (2003) CT-value formation procedure to construct the d-tree associated with it.

(57) CT-value formation:
     step 1: What's a building on campus no one likes?
     step 2: $\left\{\begin{array}{l} \text{What's a building on campus no one likes?} \\ \text{What's a building on campus someone likes?} \\ \dots \\ \text{What's a building on campus everyone likes?} \end{array}\right\}$

(58)

                         What is a building on campus who likes?

            What is a building on campus no one likes?   ...

   Is Robarts a building on campus no one likes?   ...

                         |
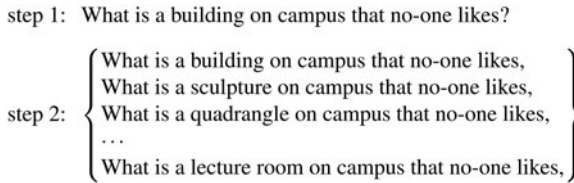
   A building on campus no one likes is Robarts.
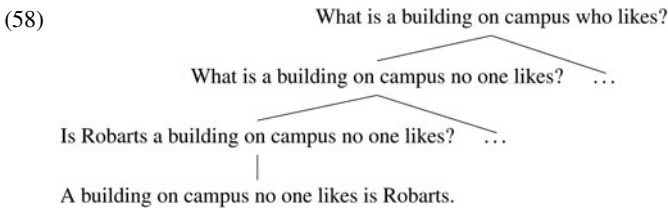
Similarly, we can see that the felicity conditions on the accent placement in SC subjects match the those of the canonical CT-F structures demonstrated in (39) and (41). So, the SCs in question need to imply a question and subquestion to which they provide a (partial) answer, and this question-subquestion-answer sequence must be congruent with the QUD. For instance, consider the infelicitous discourses in (59) and (60).

(59) Everyone likes Hart House
     # A BUILDING on campus no-one likes is Robarts.

(60) A: What's a building on campus no one likes?
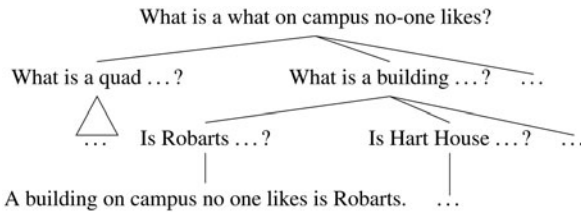   B: # A building on campus everyone LIKES is Hart House.

If we assume that the SCs in these examples represent CT-Exh structures, then we can easily explain their infelicty. Assuming the stressed element, *building*, in (59) is

F-marked within a CT, and the DP2 *Robarts* is the exhaustive focus, then we can represent the CT-Value generation and resulting d-tree in (61) below.
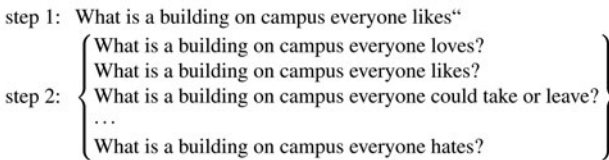
(61)   a. CT-Value Formation

step 1:   What is a building on campus that no-one likes?

step 2: $\left\{\begin{array}{l}\text{What is a building on campus that no-one likes,}\\ \text{What is a sculpture on campus that no-one likes,}\\ \text{What is a quadrangle on campus that no-one likes,}\\ \ldots\\ \text{What is a lecture room on campus that no-one likes,}\end{array}\right\}$

b. D-Tree

What is a what on campus no-one likes?

What is a quad ...?   What is a building ...?   ...

...   Is Robarts ...?   Is Hart House ...?   ...

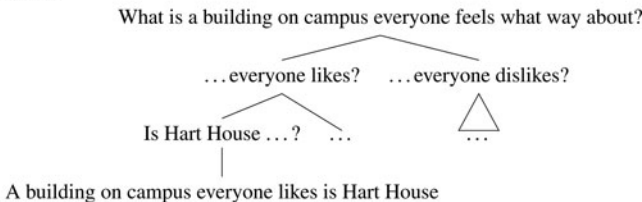A building on campus no one likes is Robarts.   ...

Note that the prior discourse, *Everyone likes Hart House*, is nowhere to be found in the d-tree generated by the SC. Therefore, the SC is infelicitous because its d-tree is incongruous with the the discourse it is embedded in.

Similar remarks apply to the SC in (60) whose proposed CT-Value and d-tree are represented in (62).

(62)   a. CT-Value formation

step 1:   What is a building on campus everyone likes"

step 2: $\left\{\begin{array}{l}\text{What is a building on campus everyone loves?}\\ \text{What is a building on campus everyone likes?}\\ \text{What is a building on campus everyone could take or leave?}\\ \ldots\\ \text{What is a building on campus everyone hates?}\end{array}\right\}$

b. D-Tree

What is a building on campus everyone feels what way about?

...everyone likes?   ...everyone dislikes?

Is Hart House ...?   ...   ...

A building on campus everyone likes is Hart House

As with the previous example, the prior discourse is not in the d-tree generated by this SC, rendering the SC infelicitous in the context. So, the intonational stress in SC subjects is consistent with a CT-Exh structure.

4.2.2   SC subjects are not wholly givenness topics

If Mikkelsen (2005) is correct, and SC subjects are necessarily topics which must contain given material, then we would expect that a maximally given DP is the

ideal SC subject. As (63a) demonstrates, however, maximally given DPs are not good
SC subjects, but SC subjects that are minimally contrastive are acceptable.[13]

(63) Many philosophers have written about the mind-body problem.

    a. # A philosopher who has written about the mind-body problem is Chomsky.

    b. A modern philosopher who has written about the mind-body problem is Chomsky.

    The entire content of the subject in (63a) has already been introduced in the dis-
course, meaning it is all discourse-given. One might, however, object that, according
to Heim's (1982) Novelty Condition, an indefinite DP must introduce a new dis-
course referent, meaning that the subject of (63a) is not wholly given. Even if this
were the case, the new dicourse referent would be included in the set of philosophers
who have written about the mind-body problem, meaning that it could still be con-
sidered discourse-given. So, it seems that SC subjects are not wholly givenness
topics.

4.2.3   SC subjects are not wholly aboutness topics

Reinhart (1981) argues that the important notion associated with topic-hood is about-
ness rather than givenness. If we wish to retain Mikkelsen's (2005) analysis, the
natural move would be to claim that licit SC subjects are characterized by aboutness.
Aboutness is diagnosable by a paraphrasing test.

(64) **Reinhart's test for aboutness**
    If sentence S is about constituent X, then S is paraphrasable by the sentence *They said
    about* X, *that* S′, where S′ is derived by replacing X in S with a proform.

    As (65) shows, when the entire SC subject is the aboutness topic, as diagnosed
by Reinhart's test, it is interpreted *de re*, rendering the copular clause equative rather
than specificational. Conversely, when the subject is not entirely the aboutness topic,
it is interpreted *de dicto*, rendering the clause specificational.

(65) **Background:** David Bowie = John's favourite singer.
    (Mary said that) John's favourite singer is Iggy Pop. (Equative/Specificational)

    a. Mary said of John's favourite singer that {he/?it}'s Iggy Pop.
        (Equative/*Specificational)
        (=Mary said David Bowie is Iggy Pop)

    b. Mary said of singers that John's favourite (one) is Iggy Pop.
        (*Equative/Specificational)
        (≠Mary said David Bowie is Iggy Pop)

---

[13]The infelicity is not due to a constraint on repeating indefinites. Consider the following:

(i) Many philosophers have written about the mind-body problem.

    a. #A philosopher who has written about the mind-body problem is Chomsky.

    b. A philosopher who has written about the mind-body problem lives next door.

c. Mary said of John that his favourite singer is Iggy Pop.
   (*Equative/Specificational)
   (≠Mary said David Bowie is Iggy Pop)

d. Mary said of people's favourite singers that John's is Iggy Pop.
   (*Equative/Specificational)
   (≠Mary said David Bowie is Iggy Pop)

In the above examples, Mary's claim that John's favourite singer is Iggy Pop is invariably false, but varies in the exact claim being made. In the case that *John's favourite singer* is understood *de re*, Mary is wrongly identifying David Bowie as Iggy Pop. When *John's favourite singer* is understood *de dicto*, Mary is wrongly specifying the singer that John prefers above all other singers to be Iggy Pop.

Several commenters have suggested the possibility that it is the pronominal subject of (65) that forces its equative reading. That is, pronouns are inherently referential, and since the subject of the copular clause in (65) is a pronoun, and therefore referential, that copular clause cannot be specificational. While I am not prepared to concede this point, even if it were true, we are left with (65d), which cannot be captured by this claim. If pronominal subjects forced equative readings, the reverse could not be true, as most SCs with full (definite) DP subjects are ambiguous with equative readings. If we were to apply this hypothesis to (65d) it would be non-predictive, so we would need a further explanation for the fact that specificational readings are forced when only part of the subject is an aboutness topic as in (65d).

So, absent any compelling argument otherwise, it seems that while some part of an SC subject can be an aboutness topic, the entire subject DP cannot be the aboutness topic.

4.2.4 Summary

Since SC subjects are compatible with F-marking and cannot be givenness or aboutness topics, it is reasonable to assume that SC subjects are CTs.

4.3 SC subjects cannot be entirely F-marked

The second claim of my proposal is that SC subjects cannot be entirely F-marked constituents. That is, if the entirety of the SC subject is new/contrastive, the SC is unacceptable. So, the B utterance in (66) is infelicitous because the entire SC subject is new/contrastive material.

(66) A: Tell me about your home university?
      B: # A BUILDING on campus no-one likes is Robarts.

This hypothesis — that SC subjects must contain but not be an F-marked constituent — in fact predicts the fact that simple indefinites cannot be SC subjects. Consider the unacceptable SC *A doctor is Mary*. The subject *a doctor* must contain an F-marked constituent, in this case *doctor*. Since the indefinite article is standardly assumed to be semantically vacuous, it does not encode any particular

information. Therefore, F-marking on the nominal is equivalent to F-marking on the entire DP.

More precisely, the analysis of the indefinite restriction proposed in this paper predicts that certain apparently simple indefinites and simple definites can be SC subjects, provided they can be analyzed as CTs which properfly contain an F-marked constituent. For instance, simple indefinites with the determiner-like elements *one* and *another* can be SC subjects if those deteminer-like elements are F-marked as in (67).

(67) a.   * A doctor$_{CT}$ is Mary.

    b.   One$_{CT}$ doctor is Mary.

    c.   Another$_{CT}$ doctor is Mary.

The elements *one* and *another* can be F-marked in a way that the indefinite article cannot, because they encode enough semantic material to generate alternatives.

Consider *another* first. Following Heim et al. (1991), I take the meaning of *other* to include two crucial parts: anaphoricity and distinctness. Consider the sentence in (68).

(68) Alice met with another student.

This sentence presupposes that there is a previously mentioned student (anaphoricity) and asserts that the student Alice met with is distinct from the presupposed antecedent (distinctness). As we can see from (69), the anaphoricity projects when embedded or negated, as shown by the fact that the (*i*) examples are infelicitous, but the distinctness does not, as shown by the fact that the (*ii*) examples are judged as felicitous.

(69) a. Alice didn't meet with another student

      i.   # … she never met with any student.

      ii.   … it was the same student.

    b. If Alice met with another student, she would have told us.

      i.   # She didn't tell us because she hadn't met with a student previous to this one.

      ii.   She didn't tell us because it was the same student.

    c. Alice probably met with another student.

      i.   # but she might not have met with a student previous to this one.

      ii.   but it might have been the same student.

    d. Johan thought that Alice met with another student.

      i.   # He was wrong. She hadn't met with a student previous to this one.

      ii.   He was wrong. It was the same student.

The SC in (67c) then, is roughly paraphrasable as *A doctor* [*OTHER than x*] *is Mary*, where the value of *x*) is resolved contextually. Assuming that *other* is CT-marked in

(67c) and, following Heim et al. (1991), that *other* is a three-place predicate,[14] we can calculate the SC's CT-value.[15] If we calculate the CT-value of (67c) given this understanding of its semantics, we can see that its acceptability is expected under my proposal.

(70) a. i. $[\![\text{ANOTHER}_{CT}\ \text{doctor is Mary}_F]\!]^f$
$= \{doctor(x) \wedge other(x)(\bigwedge doctor)(y)|x \in D_e\}\ (y\ \text{is a doctor})$
(Who is another doctor?)

ii. $[\![\text{ANOTHER}_{CT}\ \text{doctor is Mary}_F]\!]^{ct}$
$= \{\{doctor(x) \wedge P(x)(y)(\bigwedge doctor)|x \in D_e\}\ |P \in D_{\langle e,\langle e,\langle e,t\rangle\rangle\rangle}\}$
($\approx$ Who is a doctor?)

b. Molly$_i$ is a doctor.
Another$_i$ doctor is Mary.

c.
Who is a doctor?

Is Molly a doctor?   Is Mary a(nother) doctor?   . . .

Molly$_i$ is a doctor.   **Another$_i$ doctor is Mary.**

So, *ANOTHER doctor* contains both new/contrastive information, in *other* and given/presupposed material in *doctor*, thus it is a licit SC subject.

The SC in (67b) shows the inverse felicity conditions, it requires that doctors have been discussed but none have been named as demonstrated in (71).

(71) a. Let me tell you about doctors.
One doctor is Mary.

b. Molly is a doctor.
#One doctor is Mary.

If *one* is merely the stressed pronunciation of *a/an*, then the account I have proposed would likely require serious revision. Fortunately, there are good reasons to believe that *one* and *a/an* are distinct lexical items. First, it is unlikely that *one* is the stressed version of *a/an*, since *a/an* has another stressed version pronounced [ej]/[æn], which usually marks a contrast of definiteness.

---

[14]Heim et al. (1991), discussing the reciprocals *each other* and *one another* give the following denotation for *other*: z is an atomic part of y, a plural individual, and z is distinct from x.

(i) $[\![other]\!] = \lambda x \lambda y \lambda z(x \cdot \Pi y \wedge z \neq x)$

If we were to translate this directly into the example under discussion (*Another doctor is Mary.*), x would be the contextually given doctor, y would be the plural individual *doctor* and z would be *Mary*. So the SC roughly means that x is a doctor, Mary is not x, and Mary is a doctor.

[15]There may be good reason to question the particulars of both of these assumptions. There is also good reason to believe that the particulars of these assumptions are irrelevant to the discussion at hand.

(72) A: Are you the professor?
   B: I'm [ej] professor.

Also, Kayne (2015) presents several pieces of evidence that *one* is lexically distinct from *a/an*. While *a/an NP* can be interpreted as generic, *one NP* cannot

(73) a. A spider has eight legs and many eyes. (generic/specific)   (Kayne 2015)

      b. One spider has eight legs and many eyes. (*generic/specific)

He also notes that the syntactic distribution of *a/an* differs from *one* as shown below.

(74) a.   i. too long a book   (Kayne 2015)

           ii. * too long one book

      b.   i. a few books

           ii. * one few books

      c.   i. * They're selling a-drawer desks in the back of the store.

           ii. They're selling one-drawer desks in the back of the store.

Kayne argues that *one* is a complex determiner composed of *a/an* and a *singular classifier*, with the syntactic structure given below in (75) Since the locus of CT-marking is not the indefinite article, it must be the *singular classifier*, which means that the classifier ought to be contentful enough to generate alternatives.

(75)
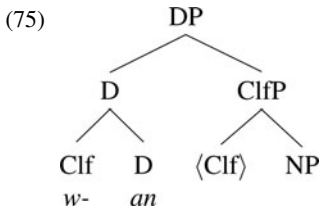
The licit SC *One doctor is Mary* would, by hypothesis, have the following CT-F structure.

(76) **CT**: $[\![w-]\!]$
      Focus: Mary
      given/presupposed: doctor/doctors/a doctor

Note that the licitness of (6) can be explained if *one*, or rather the hypothesized *w-* classifier, is F-marked within a CT. We can see that this, in fact, fits with the context in which the SC in (6) is found, reproduced in (77).

(77) Among the best potential witnesses on the subject of Iraq's actual nuclear ca- pabilities are the men and women who worked in the Iraqi weapons industry and for the National Monitoring Directorate, the agency set up by Saddam to work with the United Nations and I.A.E.A. inspectors. Many of the most senior weapons-industry officials, even those who voluntarily surrendered to U.S. forces, are being held in captivity at the Baghdad airport and other places, away from reporters. Their families have been told little by Amer- ican authorities. Desperate for information, they have been calling friends and other contacts in America for help.

> **One Iraqi émigré who has heard from the scientists' families** is Shakir al Kha Fagi, who left Iraq as a young man and runs a successful business in the Detroit area.[16]

The first paragraph is about Iraqi weapons scientists and introduces the group of Iraqi émigrés who have been contacted by families of these scientists. The second paragraph introduces a particular member of that group. As I discussed above, this discourse pivots from group to individual member is naturally achieved by an F-marked *one* (*ONE Iraqi émigré …*). Since DP2, in this case *Shakir al Kha Fagi*, is the Exh of the SC, the F-marking in the SC subject must be indicative of a CT.

If this is the correct analysis of F-marked *one*, then the singular classifier must be able to generate alternatives. The question is, what counts as an alternative to *one*. A proper answer to that question would require an in-depth study of the semantics and pragmatics of *one*, which is beyond the scope of this paper.

An anonymous reviewer suggested (78) as a possible counterexample to the CT condition on SC subjects.

(78) Many young people are turning away from technology; **an example is Sam**, who replaced her iPhone with a flip-phone this year.

In this example, the bolded clause is an SC with an apparent simple indefinite subject. On the surface, this seems to invalidate the generalization in (48), for two reasons. First, the subject *an example* is, in a sense, anaphoric to the preceding discourse, and therefore given/presupposed rather than new/contrastive. Second, the subject seems to be a simple indefinite, and therefore, any F-marking would constitute F-marking of the entire constituent. These arguments, however, do not hold water.

First, in the natural intonational contour of (78) example bears a L + H* pitch accent, as can be seen in Figure 4 This suggests that *example* is new/contrastive, rather than given/presupposed. Which brings us to the second argument: *Example* being F-marked amounts to the entire DP being F-marked. When we consider the nature of the noun *example*, however, we can see that the apparent simplicity of the DP *an example* is just that: apparent.

Example is a relational noun, meaning there are no example-things in the world the way there are dog-things, red-things, or courage-things. If an entity is an example, it is an example *of* something. In this case, Sam is an example of a young person who is turning away from technology. So, if we assume that the DP *an example* includes an elided complement, two things are explained: First, the fact that we seem to have a licit SC with a simple indefinite subject, and second, the fact that *an example* seems to be discourse anaphoric. So, the SC in (78) is properly analyzed as in (79).

(79) [An exAMpleF Ø PP] is SamExh.

Note that if we replace *example* with an non-relational noun, (78) becomes unacceptable.

---

[16]Seymore M. Hersh "The Stovepipe", *The New Yorker*, Oct 27, 2003: 86 cited by Mikkelsen (2005: 118)
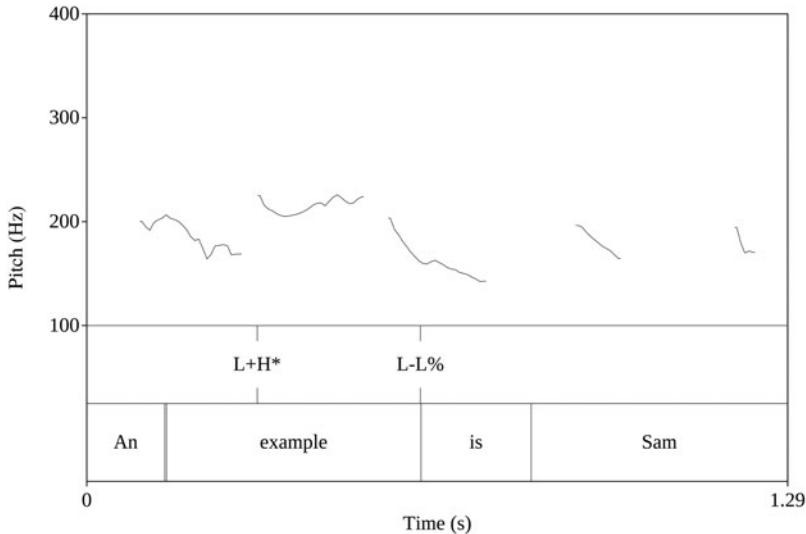
Figure 4: A Natural pitch contour of *an example is Sam* in (78)

(80) Many young people are turning away from technology; **#a luddite is Sam**, who replaced her iPhone with a flip-phone this year.

Thus, (78) is not a counterexample to the analysis of SC subjects offered in this paper.

Heycock (2010) and Béjar and Kahnemuyipour (2013) discuss a particular reading of SCs with simple definite subjects, called "the Poirot reading" which is shown below in (81).

(81) And Poirot pointed at the Major and said "For a long time now we have been trying to establish the identity of the murderer. But now I know...
... The murderer is you"

At first blush, this seems to be a counterexample to my proposal. In this context, the existence and relevance *the murderer* is entirely given/presupposed, while the fact that the identity of the murderer is Poirot's addressee seems to be new/contrastive. This would mean that no part of the subject is F-marked, which should render the clause unacceptable.

If we consider the context carefully, we can see that this is not the entire story. The sentence *The murderer is you* would occur at the culmination of a murder mystery at which point many properties of *the murderer* have been gleaned from the evidence. The only relevant "property" left is *the murderer*'s identity. So, what is given is the existence, salience and uniqueness of some murderer and several of *the murderer*'s properties. What is new/contrastive is the identity of *the murderer*, and that that identity is Poirot's addressee.

Consider the following alternative discourse:

(82) We already know the following: The murderer is 6 feet tall. The murder has dark hair.
      The murderer walks with a limp. From this I have deduced that
      … The murderer is you.

In the discourse leading up to *The murderer is you*, we can see that *the murderer* is only used referentially. The culminating accusation shifts the usage of *the murderer* to that of a predicate. For the purposes of this paper, I will assume that shifting *the murderer* from e to $\langle e, t \rangle$ is accomplished by an IDENT operator (cf. Partee 1987).[17] The SC in (82) and (81), then, has the following CT-F structure:

(83) The murderer is you.
      Focus: *you*
      CT: IDENT
      Given/presupposed: *the murderer*

In fact, this sort of context allows us to generate a (near) minimal pair if we change the SC to a negative SC. So, a similar context to that in (82) will allow a negative SC, as demonstrated in (84).

(84) We already know the following: The murderer isn't 6 feet tall. The murder doesn't
      have dark hair. The murderer doesn't walk with a limp. From this I have deduced that
      … The murderer isn't you.

As in the previous example, the context here is a discussion of properties that hold (or don't hold) of the murderer, and the SC marks a shift into talking about the identitiy of the murderer. Contrast this with (85), where no such shift occurs.

(85) We already know the following: Mrs Bridges isn't the murderer. Mr Devizes isn't the
      murderer. Colonel Mustard isn't the murderer. …
      … #The murderer isn't you.

Here, the context is precisely a discussion of the identity of the murderer and, therefore, when the SC is uttered, the IDENT operator cannot be F-marked, as it is not contrastive. In fact, only the pronoun *you* is contrastive here, and it serves as exhaustive focus. The infelicity of the SC in (85), then, is predicted by the proposed CT condition in (48).[18]

---

[17]Note that the IDENT operator is not paraphrasable as *the identity of*. Rather, a better paraphrase might be *the entity that matches the description*. This can be seen in the case of a predicational copular clause, which is the canonical example of an IDENT operator for Partee.

   (i) John is the murderer.
   a. *John is the identity of the murderer.
   b. John is the entity that matches the description *the murderer.*

[18]An anonymous reviewer suggests the following version of the Poirot context:

   (i) We already know the following: Mrs Bridges isn't the murderer. Mr Devizes isn't the murderer.
       Colonel Mustard isn't the murderer. From this I have arrived at the only possible conclusion…
       …The murderer is you.

Note that the context is populated by negative clauses, while the SC is affirmative. This modicum of contrast seems to be sufficient to make the SC acceptable. This analysis, however, raises two issues for the proposal argued for in this paper. The first issue is that

Another example of SCs with simple definite subjects, provided by an anonymous reviewer, is given in (86).

(86)   A: Who are the members of the Ramanu Trio?
       B: The PIanist is Zach Pitts, The DRUmmer is Monica Brown, and the SINGer is Sue Listman.

This, however, is not a counterexample. Taking the first SC, *the PIanist is Zach Pitts*, as an example, we can see that *Zach Pitts*, being the Exh of the sentence, is new/contrastive, as is *pianist*, being F-marked. To see what is given/presupposed, consider what B is actually asserting by uttering *the PIanist is Zach Pitts*. Speaker B is asserting that Zach Pitts is the pianist *in the Ramanu Trio*. This sugests one of two possible things: One possibility is that there is actually an elided PP in the SC subject in question, in which case, the SC subject is not a simple definite. The other possibility is that the content of the PP comes from the context by way of the definite determiner, in which case the determiner cannot be F-marked. In either case, F-marking of the nouns in (86), does not entail F-marking of the entire SC subject.

As stated, the condition that indefinites can be SC subjects if they contain but do not comprise an F-marked constituent is a necessary condition, but not a sufficient one. For instance, consider the example of an over-informative answer to a polar question in (87) (adapted from Mikkelsen 2008).

(87)   A: Is Eve the undergraduate advisor?
       B: i.   No, Eve is the GRAduate advisor.
           ii. # No, the GRAduate advisor is Eve.

If the condition discussed above was the only condition on SCs, we would expect an SC to be a felicitous answer to A's question in (87) . B's F-marking of *GRAduate* implies the superquestion *who is which advisor?*, suggesting that the DP *the GRAduate advisor* is a CT. However, the SC is infelicitous due to the other DP *Eve*. Recall that in an SC, DP2 must be Exh; however, in (87), *Eve* seems to be an aboutness topic, as the examples in (88) indicate.

(88)   a. Mary asked about Eve$_i$ if she$_i$ was the undergraduate advisor.

       b. * Mary asked about Eve$_i$ if the undergraduate advisor was her.

If A's question is reformulated, however, the judgements of B's responses are reversed, as shown in (89).

_____

affirmation is "marked" not by an affirmative morpheme, but by the absence of the negative morpheme. This would seem to commit me to the implausible proposition that nothingness can be contrastively marked. One could get around this issue by postulating a null affirmative morpheme that, along with the negative morpheme, forms the category POLARITY.
Setting aside its stipulative nature, the proposal of such a morpheme raises another issue for the CT condition. According to the condition as I have proposed it, the SC subject itself must contain an F-marked constituent. A polarity morpheme, however, would be part of the clausal spine, and therefore not contained in the subject. The examples (i) and (85), it seems, require further investigation which I set aside for now.

(89) A: Is the undergraduate advisor Eve?
     B: i. # No, Eve is the GRAduate advisor.
         ii.   No, the GRAduate advisor is Eve.

In this case, *Eve* is the Exh, and *the GRAduate advisor* is the CT with *GRAduate* being F-marked. The SC response to A's question in (89) now meets all of the conditions on SCs as expressed in (48).

4.4 Repetitive SC Subjects

An intuitive prediction of my proposal is that an SC subject cannot be a reiteration of preceding discourse. An anonymous reviewr, however, has identified a class of SCs which seem to be violations of this prediction. SCs of this class, as demonstrated in (90), are answers to *wh*-questions and have subjects which entirely reiterate part of the questions they answer.

(90) A: Who is your favourite author?
     B: My favourite author is Kerouac.

The fact that these SCs are answers to *wh*-questions is crucial to their acceptability. Note that if the SC in (90) follows a different type of utterance and has a reiterative subject, it is infelicitous, as shown in (91).

(91) My favourite author died in 1969.
     #My favourite author is Kerouac.

Assuming we take the answer to a *wh*-question to be a copy of that question with the *wh*-element replaced by a non-*wh*-element, then the topic-focus structure of the question should be carried over into the answer. It already seems to be the case that the exhaustive focus of the answer falls on the element that replaces the *wh*-element. Furthermore, since topic and focus are standardly taken to be syntactically represented (Rizzi 1997), it stands to reason that they would be copied over from question to answer. So, for instance, if (90) was part of a conversation in which A was trying to discover how B felt about which authors, the adjective *favourite* would be contrastive in both A's question and B's answer.[19] Therefore, the SC subject in (90) can be reiterative and contrastive, provided the contrast exists in the preceding *wh*-question.

4.5 Summary

In this section I have presented evidence that the restriction on indefinite SC subjects comes from a requirement that an SC subject be a CT that properly contains an F-marked constituent. I first showed that *contrastive* rather than aboutness or givenness topic-hood is the source of the restriction. I then argued that the ban on simple indefinite SC subjects is neatly predicted if the SC subject itself, rather than a proper part of it, is banned from being the F-marked constituent. In the next section, I will

---

[19]Given the fact that *favourite* intuitively implies a comparison class, it seems reasonable to assume that it, along with other words with similar meanings, is inherently contrastive in some sense. In what sense these types of words are contrastive and how that inherent contrast interacts with F-marking are questions for further research.

discuss the prospects of relating this discourse-pragmatic account of SCs to their syntax.

5. RESIDUAL ISSUES AND CONCLUSION

Before concluding, I will address an unresolved issue which future research could take up. Specifically, I will address the syntactic analysis of SCs based on Constant's analysis of CT that my discourse-pragmatic analysis suggests.

5.1 A syntactic analysis

Constant (2014: 124) proposes that the CT-λ head occupies a position in the left periphery. Specifically, he proposes that it occupies one of the Top positions first hypothesized by Rizzi (1997). The CT constituent, then, occupies the specifier position of the phrase projected by CT-λ, having moved there from a lower position, often covertly. It is, I believe, quite reasonable to hypothesize that the overt SC subject postion is, in fact, this [Spec, CT-λ] position. Assuming the copula surfaces at least as high as T, an SC can be represented as in 92.

(92) [$_{TopP}$ [$_{DP}$ An example$_F$ Ø$_{PP}$] CT-λ [TP is Sam.]]

If the SC surfaces as (98), then we would expect the CT-λ to cliticize to the DP and be pronounced as a H% tone (following Constant 2014). However, as I discussed in section 4.2.1, there is some variation in the boundary tone associated with the SC subject. Specifically, some SC subjects lack H% boundary tones. This seems to run counter to Constant's phonetic diagnostic for CTs, but can be explained if we assume that the copula optionally raises to CT-λ. In this case, one of two things will occur: either the COP + CT-λ amalgam will cliticize to the subject DP in th form of =*'s*, or it will fail to cliticize, and surface as an independent word.

(93) [$_{TopP}$[$_{DP}$ An example$_F$ Ø$_{PP}$] is+CT-λ [$_{TP}$ Sam.]]
    a. An example's Sam.
    b. An example is Sam.

In either case, the CT-λ head will not surface as an H% boundary tone.

5.1.1 *Difficulties in expressing the CT requirement syntactically*

Since the CT requirement, as defined in (48), is a restriction on a particular syntactic structure, it should be expressible in syntactic terms. According to the analysis presented above, an SC is characterized by a copular clause embedded below a TopP headed by CT-λ. This head then triggers the movement of a predicative DP, if some part of that DP is F-marked. Adapting this analysis to reflect the CT condition, however, is problematic. To demonstrate this, I will assume that a CT feature on the predicative DP triggers/licenses SCs (at least with indefinite subjects). Consider the SC in (94), below.

(94) [$_{DP}$A figure [$_{PP}$in the history [$_{PP}$of generative$_F$ grammar]]] is Eric Lenneberg.

In this case the F-marking is on an adjective in a PP, which is embedded in a PP in the SC subject, rather than the SC subject itself. If we assume that this F-marking

behaves like classic F-marking, then it ought to project in the manner that Selkirk (1996) describes.

(95) **Focus Projection** (Selkirk 1996)

    a. F-marking of the *head* of a phrase licenses the F-marking of the phrase.

    b. F-marking of an *internal argument* of a head licenses the F-marking of the head.

Crucially, according to Selkirk, non-arguments do not project focus, so F-marking of *generative* in (94) would not project to the entire subject.

Suppose, however, the CT condition is satisfied by Agree. It is still not clear that this could account for the SC in (94), as the F-marked constituent is contained in a strong island (i.e., a complex NP). In standard theories, Agree has the same structural requirements as movement, so we expect it to obey strong island constraints, rendering the F-marked constituent *generative* inaccessible to Agree.

It seems, then, that more work will be required to express the CT condition syntactically.

5.2   Concluding remarks

In this paper I have presented a pragmatic account of the restriction on indefinite SC subjects. According to this account, SC subjects must be a CT, but must not be wholly F-marked. I have shown how this captures the fact that simple indefinites cannot be SC subjects. If this account is correct, and the indefinite restriction is due to pragmatic rather than semantic constraints, then the restriction can no longer be adduced as evidence against an inversion analysis of SCs. This is not to say that I have presented evidence in favour of the inversion analysis. Rather, I have striven to present my account of the restriction in neutral terms with respect to this debate.

That said, a couple of comments regarding the dabate are called for. I have proposed that SCs have a rather rigid information structure (DP1 is CT, DP2 is Exh), and any theory of SCs must account for that. Since current syntactic theories of information structure tend to involve movement to Topic or Focus projections, and inversion accounts of SCs necessarily involve movement, the latter seem to be more naturally suited to explaining SCs. It may turn out, however, that other facts militate against an inversion analysis. In this case, our theory must account for the rigid CT-Exh structure of SCs in some way. This, however, is a topic for another paper.

ADDITIONAL SC PITCH CONTOURS

In addition to the recordings presented in section 4.2.1, I recorded another native speaker of Canadian English uttering SCs. Unlike the speaker in section 4.2.1, who is an actor and comedian, this speaker is a linguistics graduate student. The target utterance was (96), and each context was designed to target a distinct CT-Exh structure.

(96) A book I would recommend is *Aspects*.

The first context, given in (97), is designed to target *book* for F-marking. The resulting pitch contour is given in Figure 5.
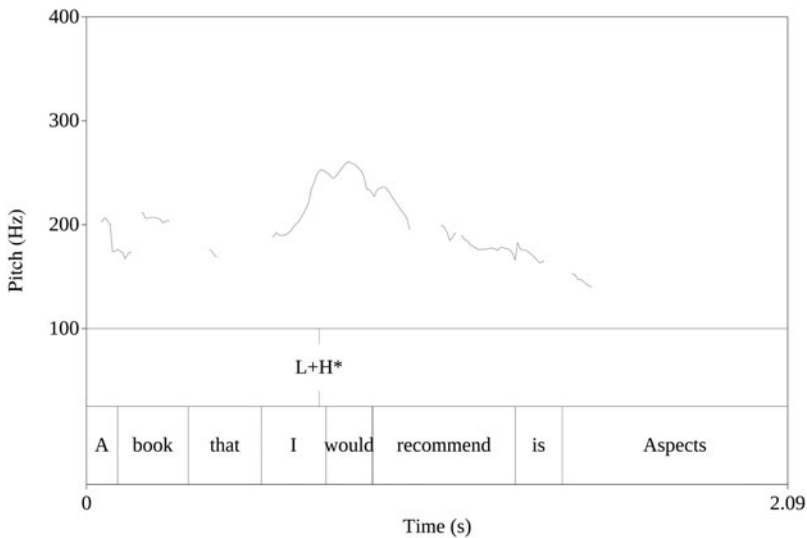
(97)   A: Do you know any good papers on syntactic theory?
       B: No, but …
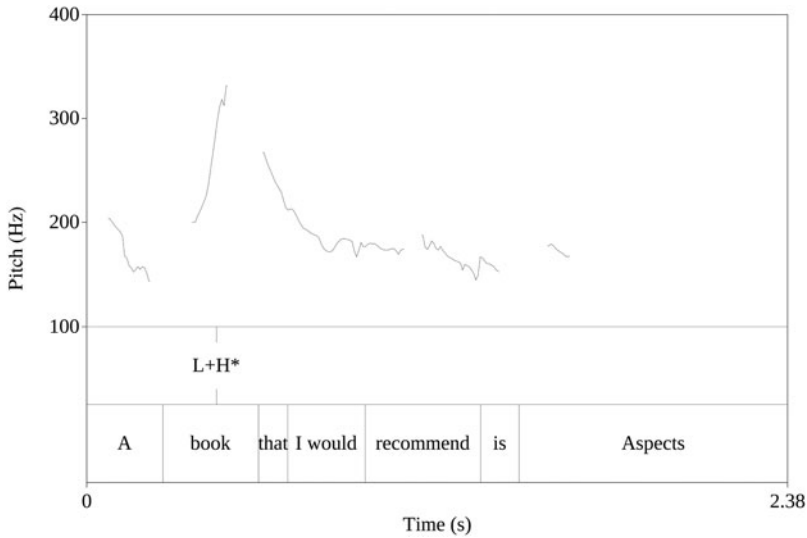
Figure 5: The intonational contour of (96) in context (97)

The second context, given in (98), is designed to target *I* for F-marking. The resulting pitch contour is given in Figure 6.

(98)   A: What's the best book on syntactic theory?
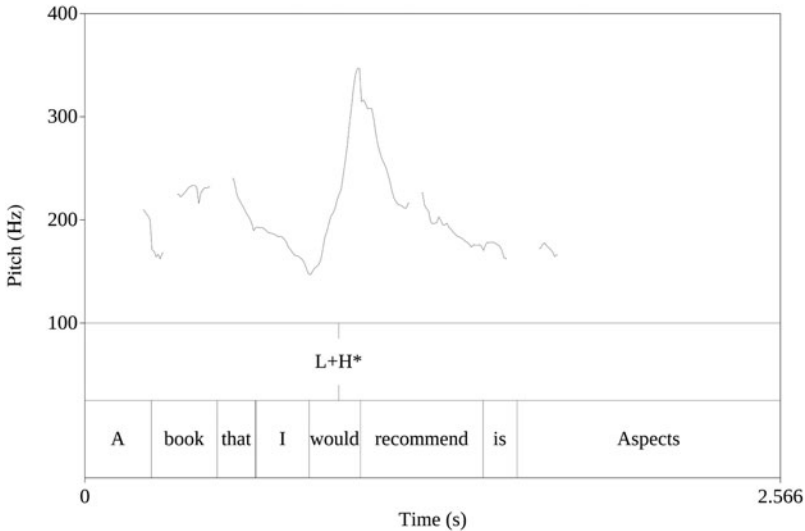   B: Most people recommend *Syntactic Structures*, but …

Figure 6: The intonational contour of (96) in context (98)

The third context, given in (99), is designed to target *would* for F-marking. The resulting pitch contour is given in Figure 7.

(99) A: Which of these books on syntactic theory would you recommend?
      B: I wouldn't recommend any of those, but …

Figure 7: The intonational contour of (96) in context (99)

REFERENCES

Béjar, Susana and Arsalan Kahnemuyipour. 2013. Agreement in copular clauses embedded in modal contexts. In *Proceedings of the 2013 annual conference of the Canadian Linguistic Association*, ed. Shan Luo. Victoria, BC: CLA. <homes.chass.utoronto.ca/~cla-acl/actes2013/Bejar_and_Kahnemuyipour-2013.pdf>.

Birner, B. J. 1994. Information status and word order: An analysis of English inversion. *Language* 20(2): 233–259.

Birner, B. J. 1996. *The discourse function of inversion in English*. New York: Garland.

Boersma, Paul and David Weenink. 2019. Praat: doing phonetics by computer. <www.praat.org>.

Büring, Daniel. 2003. On d-trees, beans, and b-accents. *Linguistics and philosophy* 26(5): 511–545.

Büring, Daniel. 2016. (Contrastive) topic. In *The handbook of information structure*, ed. Caroline Féry and Shinichiro Ishihara. Oxford: Oxford University Press.

Constant, Noah. 2014. *Contrastive topic: Meanings and realizations*. Doctoral dissertation, University of Massachusetts Amherst.

Delacruz, Enrique B. 1976. Factives and proposition level constructions in Montague grammar. In *Montague grammar,* ed. Barbara Partee, 177–199. New York: Academic Press.

Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: MIT Press.

Halliday, Michael A.K. 1967. Notes on transitivity and theme in English: Part 2. *Journal of linguistics* 3(2): 199–244.

Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of language* 10(1): 41–53.

Heggie, Lorie. 1988. *The syntax of copular constructions*. Doctoral dissertation, University of Southern California, Los Angeles.

Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*. Doctoral dissertation, University of Massachusetts.

Heim, Irene, Howard Lasnik, and Robert May. 1991. Reciprocity and plurality. *Linguistic Inquiry* 22(1): 63–101.

Heycock, Caroline. 1994. The internal structure of small clauses. In *Proceedings of NELS 25*, ed. Jill Beckman, 223–238. Amherst.

Heycock, Caroline. 2010. Variability and variation in agreement in copular clauses: Evidence from Faroese. Talk presented at CGSW 25. Tromsø, Norway.

Heycock, Caroline. 2012. Specification, equation, and agreement in copular sentences. *The Canadian Journal of Linguistics/La revue canadienne de linguistique* 57(2): 209–240.

Heycock, Caroline and Anthony Kroch. 1999. Pseudocleft connectedness: Implications for the lf interface level. *Linguistics Inquiry* 30(3): 365–397.

Higgins, Francis Roger. 1973. *The pseudo-cleft construction in English*. Doctoral dissertation, Massachusetts Institute of Technology.

de Hoop, Helen. 1992. *Case configuration and Noun Phrase interpretation*. Doctoral dissertation, University of Groningen.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Kayne, Richard S. 2015. English *one* and *ones* as complex determiners. <linguistics.as.nyu.edu/docs/IO/2652/Kayne0615One.pdf>.

Lenerz, Jürgen. 1977. *Zur Abfolge nominaler Satzglieder im Deutschen*. Tübingen: Gunter Narr.

Mikkelsen, Line. 2004. *Specifying who: On the structure, meaning, and use of specificational copular clauses*. Doctoral dissertation, University of California Santa Cruz.

Mikkelsen, Line. 2005. *Copular clauses: Specification, predication and equation*. Philadelphia: John Benjamins.

Mikkelsen, Line. 2008. Specification under discussion. In *Annual meeting of the Berkeley Linguistics Society*, vol. 34, 473–482.

Milsark, Gary L. 1974. *Existential sentences in english*. Doctoral dissertation, Massachusetts Institute of Technology.

Moro, Andrea. 1997. *The raising of predicates: Predicative noun phrases and the theory of clause structure*. Cambridge: Cambridge University Press.

Partee, Barbara. 1987. Noun phrase interpretation and type-shifting principles. In *Studies in discourse representation theory and the theory of generalized quantifiers,* ed. Jeroen Groenendijk, Dick de Jongh, and Martin Stokhof, 115–143. Dordrecht: Foris.

Pierrehumbert, Janet and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication,* ed. Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, 271–311. Cambridge: MIT Press.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27(1): 53–94.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar,* ed. Liliane Haegeman, 281–337. Dordrecht: Springer.

Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6): 1–69.

Romero, Maribel. 2005. Concealed questions and specificational subjects. *Linguistics and Philosophy* 28(6): 687–737.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural language semantics* 1(1): 75–116.

Selkirk, Elisabeth. 1996. Sentence prosody: Intonation, stress, and phrasing. In *The handbook of phonological theory*, ed. J.A. Goldsmith. Cambridge, Mass: Wiley-Blackwell.

Williams, Edwin. 1997. The asymmetry of predication. In *Texas linguistic forum 38 - the syntax and semantics of predication,* ed. Ralph C. Blight and Michelle J. Moosally, 323–333. Austin: Department of Linguistics, The University of Texas at Austin.