Binding Theory, Semantic Interpretation and Context*

Josep Macià
Department of Linguistics and Philosophy, MIT, 20D-219.
77 Massachusetts Ave. Cambridge, MA 02139, USA
e-mail: josepmac@mit.edu

Received: January 1st 1996
Accepted: July 6th 1996

Abstract

There are several well-known counter-examples to the Binding Theory of Chomsky (1981, 1993) which have to do with Principles (B) and (C). These counter-examples have different degrees of acceptability. We can account for these counter-examples and explain their different degrees of acceptability by (i) Distinguishing different types of sentences among these examples; different kinds of sentences exemplify different phenomena and require different explanations. (ii) Viewing Binding Theory as a semantic theory about how to semantically interpret certain structural relationships between NPs. (iii) Modifying principles (B) and (C) so that they are not about the relation of having the same value, but rather about the relation of it being presupposed to have the same value in a given context. (iv) Identifying how the Binding Principles interact with other sources of semantic information in order to yield the interpretation for a sentence. These other semantic factors might provide information that conflicts with (B) or (C). The data show that when these other semantic factors determine some interpretation for a sentence in a clear enough way the speakers judge such an interpretation as at least partially acceptable.

Key words: formal semantics, binding theory, semantic interpretation.

Resum. Teoria del Lligam, interpretació semàntica i context

Hi ha tota una sèrie de coneguts contraexemples a la Teoria del Lligam de Chomsky (1981,1993) que tenen a veure amb els principis (B) i (C). Aquests contraexemples posseixen diferents graus d’acceptabilitat. Podem explicar aquests exemples i explicar a més per què unes frases són jutjades més acceptables gramaticalment que altres si fem...

* I would like to thank Manuel García-Carpintero, Howard Lasnik, Joe Lau, Alec Marantz, Susanna Pi, Gemma Riga, Joanna Rosselló Asensi, Robert Stalnaker and an anonymous CWPL reviewer for comments, discussions and other assistance. Special thanks are due to James Higginbotham for encouraging me to work on this topic, and to Irene Heim both for discussions and for the influence that her ideas on this topic have had on my own. I also benefited from the comments and reactions of the audiences at the I Inicontro Italo-Spagnolo di Filosofia del Linguaggio e Filosofia della Mente (Università di Venezia, June 3, 1994), and the Fourth Annual Harvard-MIT Graduate Philosophy Conference (Harvard University, March 10, 1996) where I presented most of the ideas in this paper.
el següent: (i) Distingir diferents tipus de frases dintre d'aquests exemples; diferents frases exemplifiquen diferents fenòmens i requereixen diferents explicacions. (ii) Considerar la Teoria del Lligam com una teoria semàntica sobre com interpretar semàticament determinades relacions estructurals entre SNs. (iii) Modificar els principis (B) i (C) de forma que no tractin sobre la relació \( x \te el mateix valor que y \), sinó sobre la relació es pressuposa (respecte a cert context) que \( x \) i \( y \) tenen el mateix valor. (iv) Identificar com els principis de la Teoria del Lligam interactuen amb altres fonts d'informació semàntica a fi de produir la interpretació d'una oració. Aquests altres factors semàntics poden proporcionar informació que estigui en conflicte amb (B) o (C). Les dades mostren que quan aquests altres factors semàntics determinen una interpretació per a una oració de forma suficientment clara els parlants consideren aquesta interpretació com a, si més no, parcialment acceptable.

Paraules clau: semàntica formal, teoria del lligam, interpretació semàntica.

### Table of Contents

1. Standard Binding Theory  
2. Counter-examples to Standard Binding Theory Principles (B) and (C)  
3. An Alternative Proposal  
4. The Disjointness Conditions and Other Sources of Semantic Information  
5. Identity and the Disjointness Conditions  
6. Some Comments  
7. Two-stages in the Interpretation of an Utterance  
8. Conclusion  

#### 1. Standard Binding Theory

1.1. Principles (B) and (C)

Consider the following examples, where two NPs in a sentence have the same index if and only if either they both refer to the same individual or one of them is a quantificational NP (like, e.g, \( Every \) \( boy \), or \( Somebody \)) and the other is bound to it (in the same sense that a logical variable is bound to a quantifier):

\[
\begin{align*}
(1) & \quad a. \text{ Kathrin}_i \text{ saw her}_j \\
 & \quad b. \text{ *Kathrin}_i \text{ saw her}_i \\
 & \quad c. \text{ She}_i \text{ saw her}_i \text{ father} \\
 & \quad d. \text{ Daniel}_i \text{ says that he}_i \text{ is tired} \\
 & \quad e. \text{ Daniel}_i \text{ says that he}_j \text{ is tired} \\
 & \quad f. \text{ He}_i \text{ says that Jason}_i \text{ is tired} \\
 & \quad g. \text{ He}_i \text{ says that Jason}_j \text{ is tired} \\
 & \quad h. \text{ His}_i \text{ father says that Jason}_i \text{ is tired} \\
 & \quad i. \text{ *(His father)}_i \text{ says that Richard}_i \text{ is tired} \\
 & \quad j. \text{ (His father)}_i \text{ says that he}_j \text{ is tired}
\end{align*}
\]
k. He, saw Robert,
l. *He, saw Robert,
m. His, mother saw Robert,

n. *Everybody, saw him
o. Everybody, saw him

p. Every thief, fears that he, gets caught
k. In Scarsdale, every single boy,’s mother sends the twerp, off to
   summer camp
r. *In Scarsdale, every boy, thought the twerp, would hate summer camp

We can explain all these data by appealing to conditions such as Chomsky’s
Principles (B) and (C) of the Binding Theory\(^1\) below (we will not go here, though,
into the detail of discussing how each sentence above is accounted for by those
principles):

**Principle (B):** Pronominals must be A-free in their Governing Category.

**Principle (C):** R-expressions must be A-free.

A succinct explanation of the terminology involved in the formulation of
Principles (B) and (C): The class of *Pronominals* includes the non-reflexive
pronouns (both possessive and non-possessive); *R-expressions* include proper
names, epithets, definite descriptions, quantificational NPs, and traces of w-h-
movement and Quantifier Raising. *Node X binds node Y* if they are co-indexed
and X c-commands Y. To be *A-free* is not to be bound by anything in an *A-posi-
tion* (complement and subject positions are A-positions, adjoined positions and
the specifier of CP are non A-positions). For the purposes of the present discussion
we can set aside some difficulties arising when trying to specify the notion of
**Governing Category**, and we can take the **Governing Category (GC)** of a node X
to be the smallest maximal projection containing both a subject and a lexical cate-
gory that governs X —or, more simply: the smallest NP or S that contains X.

Different sets of conditions can be formulated that, even if they are not equi-


\(^1\) In, for instance, Chomsky (1981:188).
formulated in terms of the relation of binding, which, in turn, appeals to the relation of 
\textit{sameness of indexes}.

Also notice that in order to account for the data in (1) we need to appeal not only to the Binding Principles but also to a principle such as «Referential NPs with the same index must refer to the same individual, referential NPs with different indexes must refer to different individuals; a NP is interpreted as bound by a quantificational NP if and only if they have the same index». We mentioned this principle above as in passing, but notice that it is crucial. Without it, Binding Theory would not explain, for instance, why (2a) is unacceptable if Begonya is coreferential with her. Even if the syntactic representation (2b)\textsuperscript{2} is ruled out by Principle (B), (2c) is not. If it were possible for two NPs with different indexes to have the same referent, then (2c) could give rise to an interpretation that (2a) does not have.

(2) a. Begonya likes her
b. Begonya\textsubscript{1} likes her\textsubscript{1}
c. Begonya\textsubscript{1} likes her\textsubscript{1}

This discussion shows also another point: a theory that tries to explain why (2a) cannot have a coreferential reading has to include, some way or other, a semantic component. The explanation can not be given in purely syntactic terms. It might be, as it is the case with Standard Binding Theory, that the bulk of the explanation is carried out by purely syntactic principles, and that the semantic component is straightforward. Still the semantic part, even if straightforward, is essential to explain what we want to explain.

It is possible to have a theory where the bulk of the explanation is at the semantic level. We can, for instance, formulate disjointness conditions which are completely analogous to Standard Binding Theory principles (B) and (C) but that operate at the semantic level (that is, they are conditions that directly constrain which interpretations are possible for a given LF-syntactic structure). Chomsky (1993) offers one set of such a semantic kind of disjointness conditions.\textsuperscript{3} Principles (B) and (C) are formulated there, roughly, in the following way:

\textbf{(B):} If \( \alpha \) is a pronominal, interpret it as distinct in reference from every c-commanding phrase in its GC.

\textbf{(C):} If \( \alpha \) is an r-expression, interpret it as distinct in reference from every c-commanding phrase.

(In order to account for cases like (1n)-(1r) where there is no co-referance involved, these principles should, of course, be modified —or at least the term «distinct reference» should be re-defined in such a way that it also applies to a relation

\textsuperscript{2} (2b) should be regarded as a partial specification of a tree. We do not explicitly specify the whole tree for simplicity of exposition. Analogously for (2c).

\textsuperscript{3} There are several examples of earlier semantic formulations of disjointness conditions, which include the non-coreference rule of Reinhart (1983), and the Rule of Interpretation of Chomsky (1973).
such as the one between Everybody and him in (1n). With such a modification in place, this theory makes the same predictions as Standard Binding Theory with respect to what interpretations the different sentences can have).

2. Counter-examples to Standard Binding Theory Principles (B) and (C)

It has been observed in the literature that there are several different kinds of sentences that are counter-examples to the Standard Binding Theory Principles (B) and (C). In (3) we have some examples of sentences that pose a problem for principle (C):4

(3) a. Who is that woman over there? She is professor Rigau.
   b. Everyone has finally realized that Oscar is incompetent. Even he has finally realized that Oscar is incompetent.
   c. Mary, Betty and John have one thing in common: Mary admires John, Betty admires John, and he himself also admires John.
   d. What do you mean Oscar loves no one? He loves Oscar.
   e. The logic tutor while trying to explain the law of Universal Instantiation to a student tells him: Look fathead. If everyone loves Oscar, then certainly he himself must love Oscar.
   f. I think that woman talking on TV is Zelda. She says the same things that Zelda says in her book.

Each of these sentences can be interpreted so that the two italicized NPs have the same referent. If they are interpreted as having the same referent, then the two NPs must have the same index in the syntactic structure of the sentence. In each of these sentences, though, one NP c-commands the other, and, furthermore, the c-commanded NP is an R-expression. The syntactic structure, so, is undesirably ruled out by Principle (C) of the Standard Binding Theory.

Analogously, the sentences in (4) pose a problem for the Principle (B) of the Standard Binding Theory

(4) a. Mary, Betty and John have one thing in common: Mary admires John, Betty admires him, and John himself also admires him.
   b. At a Halloween party someone says: I think that the man with the devil costume is Joe. It is suspicious that he knows him so well.
   c. What do you mean no one loves Oscar? Oscar loves him.
   d. What do you mean no one loves Oscar? He loves him.

Some of these sentences (like (3a) or (4b)) are one hundred per cent acceptable, others are not perfect but are good enough as to pose a problem for a theory according to which they are simply ruled out.

4. Even if I do not specify the source of each of these sentences, most of them are adapted or directly taken from the references mentioned at the end; the same is true of the sentences in (4).
Heim (1993) offers some examples involving quantification which she claims are marginally acceptable and, to the extent that they are acceptable, also pose a problem for Principle (B), as for instance

(5) a. The logic tutor while trying to explain the law of Universal Instantiation tells his student: Look, if everyone loves Oscar then it surely follows that Oscar himself must love Oscar. And, of course, this doesn’t just hold for Oscar, but for any arbitrary man: if everyone loves a man, then that man himself loves him.

b. Somebody said that what he had in common with his siblings was that his sister admired him, his brother admired him, and he himself also admired him.

3. An alternative proposal

3.1. Other Proposals

There are several ways in which we could try to deal with these problematic examples: One possibility would be to claim that the disjointness conditions should not be constraints on the relation of ‘co-indexation’ or of ‘co-reference’ but rather they should constrain some other relation, and that when this is done the disjointness conditions will allow just those interpretations that the sentences can actually have. This is the approach taken in Higginbotham (1992). Of course the difficult task when opting for this view is to specify a suitable relation in terms of which we can formulate the disjointness conditions. Higginbotham introduces the relation of common reference, and formulates the disjointness conditions in terms of it. This relation is defined in terms of the intentions of the speaker.

Another possible approach is to accept principles (B) and (C) as they are in Standard Binding Theory, but to modify the semantic part of the theory. Recall that the principle on the interpretation of indexes requires that two NPs get assigned the same individual if and only if they have the same index (and it also constrains how to interpret those NPs that will have an interpretation analogous to bound variables). One possible modification to the principle is to require only that if two NPs have the same index then they get assigned the same individual, but allowing that NPs with different indexes might also get assigned the same individual. Fiengo & May (1994) takes this kind of approach. The difficult part for this approach is not to explain, for instance, why (3b) can have a co-referential interpretation, but rather why (2a) can not usually have it. Fiengo & May (1994) tries to solve this by appealing to pragmatic principles.

Another possibility is to think that in addition to disjointness conditions, we should also have some other principle which will allow us to make the desired distinction between, on the one hand, sentences like the ones in (3)-(4) and, on the other hand, sentences which are bad and should be ruled out. This is, very roughly, the line followed both by Tanya Reinhart (in, for instance, Reinhart (1983a,b) and Grodzinsky & Reinhart (1993)) and by Irene Heim (in Heim (1993)). The main difference between the two is this: Reinhart distinguishes, among the cases to which Standard Binding Theory applies, between the cases where there is binding.
and cases where there is merely co-reference, and proposes that the disjointness conditions should only constrain cases involving binding, whereas cases involving co-reference should be accounted for by another sort of principle (her Co-reference Rule), which is less restrictive than principles (B) and (C) of the Standard Binding Theory. Heim, on the other hand, proposes that the disjointness conditions apply to both the cases involving binding and the cases involving co-reference; the cases involving co-reference, though, will be in addition subject to another principle (her Co-determination Rule), which can redeem some sentences that the binding principles would rule out.

I think that none of these proposals is entirely satisfactory, though each one provides valuable insight into the phenomena. I will not comment on these or other theories here, though. I would rather like to propose here another way of looking at the phenomena we have been discussing. Of course, a complete defense of the proposal I will put forward would require a careful discussion of the alternatives. I intend to present such a discussion elsewhere.

3.2. Overview

An adequate theory should not only account for the contrast between the sentences which are not acceptable and those which are at least partially acceptable but it should also give an account of why the sentences in (3)-(5) have different degrees of acceptability, and specially, it should account for the difference between fully grammatical sentences like (3a), (3f) or (4b) and sentences which are regarded as grammatically awkward and only partially acceptable —like, for instance, (3c) or (4d).

The proposal defended here does not consist in trying to find a set of disjointness conditions with which none of the sentences in (3)-(5) is in conflict, but rather the disjointness conditions will be part of a more general account of what makes us regard some sentence (with some specific interpretation) as acceptable. This general account will explain why these sentences are acceptable even if some of them will be in conflict with the disjointness conditions we will provide.

In the present proposal I attempt to explore the idea in Chomsky (1993) of regarding the disjointness conditions as semantic principles about how to interpret the NPs in a sentence. 5

In summary, the present proposal is the following: The sentences that are fully acceptable can be divided in two groups: identity sentences, and non-identity sentences. I formulate in section 3.3. the disjointness conditions (in semantic terms) so that they do not constrain the relation of co-reference but rather the relation of presupposed co-reference —in this I follow Postal (1970) and Heim (1993). These disjointness conditions are then no longer in conflict with fully acceptable non-identity sentences like (3f) or (4d) (section 5.1.). On the other hand, I argue in section 5.2 that identity sentences like (the second clause of) (3a)

5. All the main ideas put forward in the following pages, though, could be incorporated, with some adjustments, into an account using a 'syntactic version' of the disjointness conditions.

6. Actually, once we take also into consideration quantificational NPs, the relation being constrained is that of presupposed co-valueness. These notions are explained in the next section (section 3.3).
are not constrained by the disjointness conditions, since one of the NPs is not an argument but a predicate.

One of the main contentions of this article is about what explains the (partial) acceptability of sentences like (3c): when interpreting an utterance of a sentence like (3c) the speakers use several sources of information, the disjointness conditions being just one of them. If the other sources determine an interpretation for the NPs in a clear enough way, the sentence will be able to be understood in accordance with that interpretation even if it conflicts with the information provided by the disjointness conditions. The existence of this conflict will explain that the sentences are not regarded as completely good (sections 4 and 6).

Finally, in section 7 I argue that we must distinguish two different levels in the interpretation of a sentence: one where only the lexical items and the grammatical structure of the sentence is taken into account, and a second level where contextual information is introduced so as to produce the full interpretation of the sentence.

3.3. Disjointness Conditions

In this section I propose a different formulation of the disjointness conditions. First I state the principles, then I explain the notions involved in their formulation.

I think that we should understand the disjointness conditions in the following way (I will refer to (B)' and (C)' below as the Disjointness Conditions):

**(B)′**: If a sentence whose LF representation is \( \ldots \alpha \ldots \beta \ldots \) (where \( \beta \) is a pronominal, and \( \alpha \) is an NP that c-commands \( \beta \) in its GC) is used in a context \( C \), it is not presupposed in \( C \) that \( \alpha \) and \( \beta \) are co-valued.

**(C)′**: If a sentence whose LF representation is \( \ldots \alpha \ldots \beta \ldots \) (where \( \beta \) is an R-expression, and \( \alpha \) is an NP that c-commands \( \beta \)) is used in a context \( C \), it is not presupposed in \( C \) that \( \alpha \) and \( \beta \) are co-valued.

Let's explain what we mean by the notions of 'being presupposed', 'context' and 'being covalued' that we use in stating the Disjointness Conditions.

'being covalued with': NPs can have different sorts of semantic value. Corresponding to these different sorts of semantic value, there are different possible relations among NPs in a sentence.

If two NPs refer to the same individual, we would say that they are on the relation of common reference. If two NPs are in the same relation that his and Every boy are in (6) when we understand the sentence so that every boy loves his own mother —where one NP acts like a variable bound by another NP—, we would say that the latter links the former.

(6) Every boy loves his mother.

---

Our characterization of the distinction between the two sorts of relation is quite uncommitted, but it suffices for the present purposes of defining the relation of 'being covalued', since, as we will presently see, the two former relations get combined in the definition of the latter.\(^8\)

We define 'to be covalued with' as the transitive closure of the relation 'to link, to be linked to, or to be in common reference with'.\(^9\) That is, \(NP\ \alpha\ \text{is co-valued with}\ NP\ \beta\ \text{iff}\ \alpha\ \text{belongs to any set}\ s\ \text{which is such that:}

\[
\begin{align*}
(1) & \quad \beta\ \text{belongs to}\ s \text{ and} \\
(2) & \quad \text{if } x\ \text{belongs to}\ s\ \text{ then any } y\ \text{ such that } x\ \text{links } y\ \text{ or } y\ \text{links } x \\
& \quad \text{or } x\ \text{is in common reference with } y\ \text{ also belongs to}\ s.
\end{align*}
\]

So, for instance, if in (7) we were to interpret both he and him as linked to Every boy, then he and him would be co-valued (even though the two pronouns are not themselves in a linking relation or in a common reference relation), and so such an interpretation would violate the Disjointness Conditions:

(7) Every boy thinks that he loves him.

'Presupposition' and 'context': The notions of 'presupposition' and 'context' that we use in stating our disjointness conditions are the same as in Stalnaker (1973, 1974): linguistic communication always takes place on the basis of a background of common beliefs and assumptions, or context. Using the possible worlds framework we can identify a context in which some instance of linguistic communication takes place with a set of possible worlds: those worlds that as far as the participants in the conversation can tell could be the actual world. The context set consists of those worlds that could be the actual world according to what the participants in the conversation believe, and believe that the others believe, and believe that the others believe that they believe and so on. (If we wanted to make this characterization more precise we should take into consideration that what is relevant is not only what the participants believe but also what they pretend to believe).

A proposition \(p\) is presupposed in a context \(C\) if \(p\) is true in each world in \(C\). A sentence \(s\) carries the presupposition that \(p\) if it would be infelicitous to assert \(s\) in any context where it is not presupposed that \(p\).

The individual that a referential NP picks up at a world in the context \(w\), is the individual that the NP would refer to if \(w\) were the actual world. A sentence is true in a world in the context \(w\), if the sentence would be true if \(w\) were the actual world.

Let's consider one example. Suppose I utter the sentence He is an artist while pointing to a man who is in front of us. Since all the participants in the conversa-

\(^8\) For our present purposes, for instance, we do not need to characterize the two relations in a way which is precise enough as to decide whether in (i) (when it is interpreted as meaning that Albert loves his own brother) Albert and his are in common reference, one links the other, or the sentence is ambiguous with respect to which of the two sort of relations holds.

\(^9\) This notion of 'co-valueness' is a very close relative of the notion of 'codetermination' in Heim (1993).
tion will believe (and believe that the others believe, etc.) that the man is in front of us and that I uttered He while pointing at him, it will be part of the context that the man is in front of us and that he is the one I am referring to by He. That is, for each world in the context w He will pick up that man in w. Let’s suppose further that we are unsure whether the man is Jim Harris. That means that there will be some worlds in the context where the man in front of us is Jim Harris, but there will also be some worlds in the context where Jim Harris is someone other than the man in front of us.\textsuperscript{10} He will pick up Jim Harris in those worlds where Harris is the man in front of us, but will not pick up Jim Harris in those worlds where Jim Harris is someone else. This agrees with the intuitive idea that, if we do not know whether the man is or not Jim Harris, then we are unsure as well as to whether He refers to Jim Harris or to someone else.

We are finally in the position to understand the central notion in our formulation of the disjointness conditions: Two NPs are presupposed to be co-valued if they are co-valued in each possible world in the context. So if \( \alpha \) and \( \beta \) are referential NPs then it is not presupposed that \( \alpha \) and \( \beta \) are co-valued if there is at least one world \( w \) in the context set such that: the individual that \( \alpha \) picks up in \( w \) is not the same as the individual \( \beta \) picks up in \( w \) (i.e., if \( w \) were the actual world then \( \alpha \) and \( \beta \) would not refer to the same individual). If a NP \( \alpha \) is understood as linked by another NP \( \beta \), this fact is not dependent on the context in the way that the specific value of a referential NP depends on the context, and \( \alpha \) will be linked to \( \beta \) in each world in the context, and so it will be presupposed that they are co-valued. So, if two NPs are not presupposed to be co-valued, then it cannot be that they are interpreted so that one of them links the other.

To end section 3, let’s point out that the disjointness conditions formulated in this section incorporate three kinds of ideas: First, the Disjointness Conditions are semantic principles that specify what the semantic significance of certain structural relations among NPs is. They are not syntactic principles that rule out certain syntactic representations. Second, the relation the Disjointness Conditions are about is the relation of presupposed co-valueness (as opposed to the relation of actually being co-valued).\textsuperscript{11} Third, the principles apply both to sentences containing quantificational NPs, and to sentences containing only referential NPs.

\textsuperscript{10} If I were to utter He is Jim Harris, then this sentence would express the necessary true proposition in those worlds where the man is Jim Harris and the necessary false proposition in those worlds where Jim Harris is someone else. Not knowing which worlds in the context agree with how the actual world is, we would not know whether the utterance was necessarily true or necessarily false. Still the utterance would be informative since it would indicate to anyone who accepted it, that those worlds where the man is not Jim Harris (and where the utterance expresses the necessarily false proposition) are not compatible with what we take to be the case, and so should no longer be regarded as part of the context. This picture helps clarify how it is possible that an utterance expresses a necessary proposition but it is nevertheless informative. For a more detailed explanation see Stalnaker (1979).

\textsuperscript{11} I take the idea of appealing to what it is presupposed rather than to what is actually the case in trying to deal with the phenomena related to the disjointness conditions from Heim (1982, 1993). Heim, in turn, credits Postal (1970) for this idea. The main difference between Heim’s formulation of the disjointness conditions and mine is that she states her Binding Principles as purely syntactic principles that rule out certain co-indexations among NPs.
4. The Disjointness Conditions and Other Sources of Semantic Information

In this section I explain what I think accounts for the acceptability of sentences like (3b) or (4a) which are somewhat awkward but which are good enough as to pose a problem to Standard Binding Theory. The explanation in this section will make no essential use of the fact that we have formulated the Disjointness Conditions in terms of the relation of presupposed co-valueness. The importance of so formulating the Disjointness Conditions will become apparent in the next section (section 5) when we discuss those sentences that, like (3f) and (4b) are completely good.

We will introduce the idea we want to put forward in this section by means of an analogy.

First, though, a comment regarding the numbers between square brackets that will appear after some of the sentences from this point on: they are the mean grade that the sentence obtained on a 0-7 scale according to the grammaticality judgments provided by a number of speakers.12 We will make use of these data at several points in this paper. Even though the specific mean grade for a certain sentence is not in itself very significant (different speakers may, for instance, have used different criteria regarding how good a sentence should be in order to be given, say, a 5), the relative grade of a sentence with respect to other sentences is, I think, very significant. I think that it is very difficult to use reliable data on the relative acceptability of different sentences without having recourse to some methodology of the kind I have used.13

4.1. Analogy

I think it would help to explain what the present proposal is if we consider an analogy. Imagine that an English speaker is asked to determine whether the sentences in (8) below are acceptable, and that she is told that an index m is meant to indicate that we are considering a reading of the sentence where the expression with the subindex m refers to some individual who is a male.14

(8) a. She, is wearing a nice dress
    b. I didn't mean to hurt her,

The speaker would say that, given that subindexation, (8a,b) are unacceptable since these sentences cannot be used with she or her referring to a man.

12. I submitted two questionnaires to 11 English speakers, asking for their linguistic judgements on a total of 36 sentences. The speakers were asked to give a grade to each sentence according to how good or acceptable they felt the sentence to be, 7 being the grade for a perfectly good sentence, and 0 the grade for a completely unacceptable one. They were asked to evaluate the sentence without thinking of any special context where the sentence might be uttered, unless such an special or uncommon context was explicitly provided in the questionnaire.

13. For a very interesting discussion and criticism of the methodology employed in linguistics see Schutze (1995).

14. Notice that here, unlike what was the case in previous examples, the index is not part of the syntactic representation of the sentence, but just a device to indicate to the person giving linguistic judgments what is the interpretation we want her to consider.
If we place the sentence in an appropriate context, through, the speaker might say that the sentences are, at least, partially acceptable. For instance, if we postulate that the sentence (8a) is uttered in a situation in which a male friend has put on a woman's dress: even if all the people involved in the conversation are well aware that the speaker is talking about a man, the use of *she* to refer to him is acceptable. Similarly for (8b): if a rather male chauvinistic speaker wants to suggest that some man who was offended by the speaker and is crying is too sensitive, he might use (8b) where *her* will be understood by all the participants to refer to the man who is crying. In (8a,b) the use of *She* and *her* carries the information that the speaker is referring to a (human) female. This information, though, might be neutralized by some extra information provided by the context in which the sentence is uttered.

What I want to suggest is that some of the examples involving violations of Principles (B) and (C) are, in part, similar to those above involving the use of *She* and *her* above: An English speaker is asked to evaluate some sentence with certain intended pattern of coreference, like, for instance, *Oscar loves him* with *Oscar* being co-referential with *him*. The sentence is declared as unacceptable since the arrangement of the NPs carries the information that they cannot be assumed to corefer, that is, carries some information that contradicts what we have stipulated is the intended interpretation we are considering. If we place the sentence in a suitable context, though, for instance in (4c) [What do you mean no one loves Oscar? *Oscar loves him*], the information provided by the different semantic factors that act on that context might overcome the information provided by the arrangement of the two NPs in the last clause, and the person providing linguistic judgements will say that it is possible to understand the sentence with coreference in that particular situation, i.e., that the sentence (with the interpretation involving coreference) is at least partially acceptable.

4.2. The Interaction of Different Semantic Factors

We could try to account for the data we have considered in section 2 in the following way: instead of trying to provide some principles that forbid certain sort of relation (be it co-indexation [as in Standard Binding Theory], sameness of sense or codetermination [as in Heim(1993)], common reference [as in Higginbotham (1992)], or referential dependency [as in Evans(1980)]) when certain structural relation among NPs obtains, I think we could rather look at the disjointness conditions as describing what information certain structural relations among NPs introduce. That is, instead of looking at it from the restrictive side (what can not be the case and will be ruled out), we could look at it from the side of what it is contributed (what information is introduced by certain structural relations). This allows us to see the disjointness conditions as part of a more general picture: that of the different semantic factors that give information about the interpretation of the NPs in a sentence. Then, I content, we have a better and more natural way of explaining the grammaticality judgements of the speakers about some of the problematic sentences: When interpreting the NPs in a sentence we have, on the one hand, the information provided by (B)' and (C)' (on the basis of the structural relations among the NPs); on the other hand, we have the information provided by other
aspects of the sentence or the discourse, and by the context. These two sources of
information usually concur, but sometimes they might be in conflict. If they are in
conflict and if the second sort of information is strong and unequivocal enough it
might overcome the information provided by the Disjointness Conditions. Then the
speakers will judge that the interpretation of the sentence induced by the second
source of information is possible. Nevertheless, and because of the conflict with the
Disjointness Conditions, the sentence will be judged as somewhat awkward or
only partially acceptable.

Sentence (9) exemplifies this point. Consider (9) (which is the same as (4a)):

(9) Mary, Betty and John have one thing in common: Mary admires John,
Betty admires him, and John himself also admires him. [3.8]

The structural arrangement of John himself and him in the last clause carries the
information that we cannot understand John himself and him so that it will be
presupposed that they are the same individual; on the other hand, other factors force
that we realize that the individual the two NPs will refer to is the same. The final
result is that we can understand the sentence with the coreferential interpretation
of John himself and him. Because of the conflict between the information that
different aspects of the sentence convey, though, the sentence has a somewhat odd
character (its mean grade is [3.8]). The factors that force that both John himself
and him are taken to refer to John are: John himself refers naturally to John
—the same person named John that has been referred to in the previous clauses—;
this is reinforced by the fact that we are expecting to be told what Mary, Betty and John
have in common, and in the previous two clauses the subject NP refers to, respec-
tively, Mary and Betty, and so we expect the subject NP in the last clause to
refer to John; in addition the use of himself also forces John himself to refer
to the same male named John that the previous clauses refer to, since the use of
himself carries the presupposition that the person in question has already been
mentioned. The use of also in the last clause introduces the presupposition that a
property that is being attributed in the last clause has already been considered
(the last clause can be seen as involving the attribution of either one of two
properties: that of 'admiring him', and that of 'being admired by John'); the
repetition in the previous clauses of the attribution of the property of admiring John,
makes that property completely salient in order to be the property that also carries
the presupposition about (as opposed to the property of being admired by John); the
repetition in the previous clauses of the attribution of the property of admiring John,
now, if also in the last clause introduces the presupposition that the last clause is
attributing the property of admiring John, that means that him in the VP of the last
clause should refer to John. Finally, the fact that it has been announced at the
beginning that we would be told what Mary, Betty and John have in common, and
that we have already been told that what Mary and Betty both do is to admire John,
makes us to expect that in the last clause we will be told two things: first (as we
already mentioned above) what John does, and, second, that what he does is the
same that Mary and Betty do: to admire John. That means, again, that both John
and him will have to refer to John. All these factors force us to interpret the
sentence so that we realize that both John himself and him refer to the same indi-
vidual. That is, all these factors together counteract the information conveyed by
the Disjointness Conditions on the basis of the structural relation between the two NPs.

Notice that if the sentence were (10)

(10) Susan admires John, and he admires him.

the sentence would be quite acceptable but the last clause would not be understood so that the two NPs refer both to John but rather it would probably be taken to mean that John admires some man or boy who, it would be assumed, the speaker is treating as having already been introduced in the conversation. If asked whether the last part of (10) can mean that John admires himself, the subject offering her linguistic judgements would have to say that, if nothing else is added, it cannot. If forced to give a grade to the sentence ‘when there is coreference’ it would be a very low grade ((10) —with the addition of italics to indicate intended coreference— obtained a mean grade below 1).

When a speaker is asked to evaluate a sentence like the last clause of (10) (with no particular context) coreference is deemed unacceptable (since the only factor that gives information about the relation between the two NPs is the structural arrangement of the two NPs, and the information that this arrangement gives is that they cannot be taken to corefer); but if a discourse or context is added which provides additional information about who the two NPs refer to, the speaker might judge that coreference is possible, and so that a co-referential reading of the sentence is, at least, marginally acceptable. Something similar was the case with respect to she and the possibility of referring to a male.

Notice that from the account of the (partial) acceptability of (9) that the have provided, it trivially follows an explanation of why if we modify (9) so as to suppress some of the features that we have said help the sentence to be regarded as good the sentence becomes less good. For instance if we omit himself or also as in (11a) the sentence becomes less good; it becomes even worse if, in addition, we do not mention one of the women as in (11b). Also, if we omit to announce that we are going to tell what it is that the three people have in common, as in (11c), the sentence is certainly worse than (9).

(11) a. Mary, Betty and John have one thing in common: Mary admires John, Betty admires him, and John admires him.

b. Mary and John have one thing in common: Mary admires John, and John admires him.

c. Mary admires John, Betty admires him, and John himself also admires him.

[2.2]

15. Unfortunately, the linguistic judgements questionnaire I use did not include the totality of the sentences we will use in our discussion. In particular, it did not include sentences (11a) and (11b). I will have to hope that the intuitions of the reader with respect to these two sentences coincide with my own —which are the ones expressed in the text.
We will further develop the idea introduced in this section in section 6, where we will comment on some of the other sentences in (3)-(5) that are regarded as acceptable for the same reasons as (9) is.

5. Identity and the Disjointness Conditions

In this section we will focus on those sentences analogous to those in (3)-(4) that are completely good ((3a), (3f) and (4b)). We will see that their interpretation does not involve any conflict with the Disjointness Conditions.

5.1. Non-presupposed identity

Under our formulation of the disjointness conditions it is easy to explain why the last part of (4b) and of (3f) are completely acceptable sentences and, unlike some of the other sentences in (3), (4) and (5), they do not have any flavour of oddness or of being only partially good. Consider, for instance, (4b) (repeated here as (12)):

(12) At a Halloween party someone says: I think that the man with the devil costume is Joe. It is suspicious that he knows him so well.

The reason why it is completely good even if he and him in the last clause actually refer both to Joe is this: when uttering the last part of (12) it is still an open question whether the man in the devil costume (who is the one he refers to) is or is not Joe (who is the one him refers to). Whether they are the same or not is precisely what is being discussed. Putting it in terms of possible worlds: in some worlds in the context, Joe is the man that is wearing the devil costume at the Halloween party, but in some other worlds in the context someone other than Joe is the one who is wearing the devil costume. That means that we interpret he and him so that they refer to the same individual in those worlds in the context where Joe is the one wearing the devil costume, but they refer to different individuals in those worlds where someone else is wearing the costume. So, we interpret the two NPs so that it is not presupposed that they refer to the same individual (and so it is not presupposed that they are co-valued). There is no problem, then, in interpreting the last part of (12) in accordance to (B)'.

We could not interpret «He knows him very well» in (12) with He referring to whoever is the man in the devil costume, and him referring to Joe if it were not an open question whether the two individuals are the same or not. We can see this in (13), where the previous discourse has been modified so as to try to make clear that when the utterance of the last clause takes place it is presupposed that Joe is the man with the devil costume.

(13) A: Do you have the list of who is each person in the party?
B: Yes.
A: Could you tell me who is the man with the devil costume?
B: The man with the devil costume is Joe.
A: Oh, I see. It is suspicious that he knows him so well.
We cannot understand the last clause so that he and him both refer to that individual that we have already established that is both Joe and the man with the devil costume.  

5.2. Identity Sentences

Regarding sentences with a so called equative use of the copula like for instance the last part of (14) (which is the same as (3a))

(14) Who is that woman over there? She is professor Rigau.

one might think that the explanation for why they are completely good sentences is the same that we just gave for (12). However this is not really the case.

5.2.1. A Problem. It is true that usually when we utter a sentence such as the last clause of (14) we are not presupposing that the two NPs refer to the same individual. Nevertheless it still might be perfectly fine to utter an identity sentence in a situation where according to what the participants assume, the two NPs will refer to same individual. For instance, if Diana, Pilar and Colin are in a room and they all know each other (and know that they know each other, etc), Diana's utterance of (15) is still good:

(15) He is Colin.

It might be a silly and pointless utterance, but it is still perfectly good from a grammatical point of view.

5.2.2. Heim's Reply. One reply given to this objection by several authors (in Heim (1988,1993), Lasnik (1990), and in some way also in Higginbotham (1992))

16. We can make the sentence good, or at least much better, by changing the tense of the discourse to past, and by replacing «It's suspicious that» in the last clause by an expression such as «This is why», «This explains that» or «No wonder then that». I do not think this poses a problem for the claim that what makes the last part of (8) good is that we are not presupposing that the individuals that he and him refer to are the same. The use of past tense and of expressions like «This is the reason why...» facilitates interpreting the sentence with respect to the context as it was before the previous sentence had been uttered. We can also see this in a text like (i):

(i) A: I wonder why Tom did not come to the party with his wife.
   B: Tom has never been married
   A: This explains why he did not come with his wife.

The use of his wife in the last clause requires that it is not presupposed that Tom is not married. This, though, is exactly the information that has been introduced in the context by B. Nevertheless, the use of «this explains why» makes easier to understand that what follows does not take for granted the information that this (in «this explains why») refers to.

More could be said about how expressions like «this is the reason why...» affect what the participants will be able to understand is the context they should use to evaluate a particular utterance. Examining this any further, though, is beyond the scope of this paper.
consists in claiming that by the very act of uttering a sentence that asserts certain identity the speaker indicates that we are not presupposing that identity (and so that the hearers should assume that the identity was not presupposed, even if they thought it was).\(^{17}\) If the context is assumed to have been such that it was not presupposed that the two individuals were the same, then the identity sentence is not in conflict with our disjointness conditions.

The reason why the speaker indicates we are not presupposing the identity when uttering the identity sentence is that it is a general principle governing conversation that what is asserted cannot be presupposed. This principle follows from Grice's Cooperative Principle, and, in particular, from the maxim of Relation.\(^{18}\)

5.2.3. Objection-1 to Heim's Reply. I do not think we should be satisfied with this explanation of why identity statements are perfectly good sentences even when the identity is already presupposed. One way of realizing that this explanation cannot be satisfactory is by noticing the following: The principle that what is said cannot be presupposed is just a principle about rational interchange of information. As with any other principle derived from the Cooperative Principle, it is possible, at the price of acting silly, not to act in accordance with the principle. I can say to you «This man is Higginbotham», and you might reply «Yes certainly, he is Higginbotham», and I might go on to say «He is Higginbotham. He is Higginbotham». My utterance would be silly, but completely acceptable from a grammatical point of view. The reason why we feel that my utterance would be silly is that it would violate the maxim of Relation ('Be relevant'). So, if in the situation described we would realize that the maxim of Relation is not operating, we should not expect the principle «what is said cannot be presupposed» (that gets its justification from that maxim) to apply either. Nevertheless, the sentence is good. So, it is not the operation of the principle «what is said cannot be presupposed» that explains why in the situation described the last utterance of «He is Higginbotham» is still completely correct.

5.2.4. Objection-2 to Heim's Reply. The reasoning just given shows that it cannot be that the explanation for why identity sentences are correct in contexts where the identity is already presupposed makes essential use of the maxim of Relation, because there are contexts where the maxim of Relation is violated but where an identity sentence stating some identity that is already presupposed is still acceptable. Now, a more direct way of objecting to what we have termed Heim's Reply is to indicate where exactly there is a problem in the argumentation given as part of the reply. In order to do so, it will help to make more explicit what the argument that appeals to the principle 'what is said cannot be presupposed' and that I am objecting to is.

17. This process of reacting to an utterance of the speaker by regarding the context as having been different from what it actually was is what David Lewis calls accommodation. See Lewis (1979).
The argument can be formulated in the following way: Let NP1 and NP2 be two referential noun phrases, and C a context where 'NP1 is NP2' is uttered and where the referents of NP1 and of NP2 are presupposed to be the same; let \( \text{val}(X,w) \) denote the individual that the referential expression X picks up in the world w, then, by assumption, we have (I) and (II):

(I) 'NP1 is NP2' is uttered in C
(II) NP1, NP2 and C are such that: \( \forall w \in C \ (\text{val}(NP1,w) = \text{val}(NP2,w)) \)

Given (II) we have that what 'NP1 is NP2' says is true in each world in C, and so that (III) holds:

(III) What 'NP1 is NP2' (when uttered in C) says is already presupposed in C.

By Grice's Cooperation Principle and, in particular, by the maxim of Relation, we have (IV):

(IV) When the speaker utters 'NP1 is NP2' in C, she is not saying something that is already presupposed in C.

So, there is a conflict between (I), (III) and (IV). It is a general fact that conflicts involving the Cooperation Principle can be solved by accommodating (i.e., by assuming that the context was different from what it actually was). In the present situation the conflict is resolved by assuming (V):

(V) The context where 'NP1 is NP2' was uttered was not C, but rather a context C' which is such that \( \exists w \in C' (\text{val}(NP1,w) \neq \text{val}(NP2,w)) \)

Finally, given (V) we have (VI):

(VI) The context where 'NP1 is NP2' is taken to have been uttered is such that there is no conflict between the Disjointness Conditions and the interpretation of the utterance in that context.

Now that we have the argument laid out in detail we can see that there is a problem in (III). (III) assumes that the sentence 'NP1 is NP2' has an interpretation before the conclusion (VI) which establishes that the sentence is not in a conflict with the Disjointness Conditions has been reached. Premise (III) could be analyzed into two sub-premises (for concreteness we assume that what a sentence says is a proposition, that a proposition is a set a possible worlds, and that identity statements express the diagonal proposition in the sense of Stalnaker (1979) —these assumptions are not at all essential to the point I want to make, though):

19. Recall that a context C presupposes that p if for each w \( C, p \) is true in w.
20. We could take (IIIa) and (IIIb) simply to be

(III) a. 'NP1 is NP2' when uttered in C says that p.
b. C presupposes that p.
(III)  

(a) An utterance of 'NP1 is NP2' in C expresses the proposition:

\{w ∈ C : \text{val}(NP1, w) = \text{val}(NP2, w)\}

(b) C presupposes the proposition \{w ∈ C : \text{val}(NP1, w) = \text{val}(NP2, w)\}

(since from (II) it follows that C = \{w ∈ C : \text{val}(NP1, w) = \text{val}(NP2, w)\},

and so, in particular, C is included in \{w ∈ C : \text{val}(NP1, w) = \text{val}(NP2, w)\}.

(IIIa) is completely unjustified, unless we regard the problem we are trying to solve as already solved—that is, unless we assume that there is no problem in establishing that, given certain context, the identity sentence has certain interpretation even though the Disjointness Conditions prescribe that the sentence does not have that interpretation given that context. (Notice, incidently, that disjointness conditions are not taken at all into consideration in the argument (I)-(V) that establishes conclusion (VI)).

In particular, under the assumptions accepted by Heim (where the disjointness conditions are syntactic constrains ruling out certain syntactic structures) (IIIa) is certainly not true. Any syntactic representation of 'NP1 is NP2' that would give rise in C to the interpretation that (IIIa) attributes to the sentence would already be ruled out at the syntactic level by the Principles (B) or (C) of the Binding Theory, and so would not express any proposition.

5.2.5. Another Reply. Maybe behind what I called Heim's reply lurks another argument that does not appeal to the principle 'what is asserted can not be presupposed'. It is the following: Suppose that a context C is such that a use of the pronoun he would naturally pick up Collin in each world in C, and so would a use of the NP Collin. If someone utters He is Collin in C, what would be the natural way of interpreting the two NPs conflicts with what the Disjointness Conditions prescribe. The hearers, assuming that the speaker is rational and is trying to abide by the Cooperative Principle, will try to find a way of escaping the conclusion that the speaker's utterance is not interpretable. There is one way of doing so: to accommodate, that is, to assume that the context was not C but rather C', where C' is such that for some world in C' the two NPs pick up different individuals. So the hearers will accommodate, and if accommodation takes place then there is no problem in interpreting the sentence in accordance with the disjointness conditions. So, the fact that accommodation will take place explains that an utterance of He is Collin is grammatical even if uttered in a context where the identity was already presupposed.

I think, though, that this argumentation is not satisfactory. If the reasoning in the previous paragraph were correct then accommodation should also be an option when trying to interpret other sentences that also involve a conflict with the disjointness conditions. Consider the contrast between (16) and (17):

(16) This woman is Pilar. She is Pilar.

21. Heim (1993) focuses her discussion only on phenomena related to principle (B). What she says, though, is easily and naturally extendable to phenomena related to principle (C) as well.
(17) This woman is Pilar. She knows Pilar.

(17) is quite bad (if She is taken to refer to Pilar), whereas (16) is fully acceptable. If accommodation is possible when interpreting the second clause of (16), then it should also be possible when interpreting the second clause of (17); and if the fact that accommodation is possible is what explains that (16) is good then (17) should be predicted to be also perfectly good; but (17) is not good.

Similarly, we cannot explain why identity sentences of the same form as (15) are completely good by appealing to an explanation like the one we gave in section 4 to account for the acceptability of sentences like (4a) or (4c). Sentences like (4a) or (4c) are only partially acceptable. As pointed out in section 4.2, it is precisely the fact that the interpretation of these sentences is in conflict with what the Disjointness Conditions prescribe that explains why they are only partially acceptable. Identity sentences even when uttered in contexts as the one described for (15) are perfectly good, though. So, it cannot be that the interpretation of identity sentences involves a resolution of a conflict similar to the one involved in the interpretation of (4a) or (4c). So, the explanation of the complete acceptability of identity sentences can not be along the same lines as the one we gave for (4a) or (4c).

5.2.6. A feasible solution. I think that a promising way of trying to account for the complete grammaticality of identity sentences is in terms of the special syntactic character of sentences containing the verb to be. Several works have pointed out and tried to explain the special syntactic properties of the so called copular sentences. In particular Heggie (1988) argues that the predicate of copular sentences is not the verb to be, but rather one of the NPs. That means that there is only one argument in identity sentences: the NP that is not the predicate. If we assume that the Disjointness Conditions apply only to arguments, then the Disjointness Conditions do not provide any information about how copular sentences

22. Recall that the explanation was that the interpretation of the sentence involves a conflict between the Disjointness Conditions and other kinds of information. In sentences like (4a) and (4c) the information opposing the Disjointness Conditions is strong enough as to prevail and make the sentence interpretable.


24. According to Heggie, this claim requires some qualifications.

25. Argument and predicate are technical notions in linguistic theory. They are usually characterized in terms of the so called Theta-theory. One alternative way of characterizing the notions of argument and predicate is in terms of the type of semantic values they can have. We will not commit ourselves here to any particular characterization of the two notions. Among many other alternative ways, one of the simplest characterizations would be the following:

We define a predicate of degree-n inductively as follows: A predicate of degree-1 is an expression whose semantic value is a function from the set of individuals to the set of truth values; a predicate of degree-n+1 is an expression whose semantic value is a function from the set of individuals to the set of predicates of degree-n. X is a predicate if, for some n, X is a predicate of degree-n. X is an argument, if X can be functionally combined with a predicate of degree-1 to yield a truth value. (Notice that under this definition not only expressions whose semantic value is an individual can be arguments; if the semantic value of a quantificational expression is a function from predicates of degree-1 to truth values, then quantificational expressions are also arguments.)
like (15) should be interpreted. The interpretation of identity sentences does not, then, involve any conflict with the Disjointness Conditions.

Discussing the different arguments that show the special character of copular sentences and that try to prove that the predicate of copular sentences is one of the NPs, as well as examining the possible evidence against this view is by itself a lengthy topic. The reader is referred to the discussion in Heggie (1988) and the other works mentioned in footnote 23. Here we will restrict ourselves to mentioning one kind of data that gives plausibility to the view, without entering in any further details.

In Catalan the clitic el (l') corresponds to an argument position; the clitic ho can correspond to predicates but cannot correspond to any personal NP in an argument position. This is illustrated in (18). In (18b) the clitic el stands for the argument el Joan in (18a), while it is not possible for ho to stand for that argument:26

(18) a. Aquell home estima el Joan.
    that man loves (the) John

b. Aquell home l'ho estima.
    that man him/it loves

In (19b) the clitic ho stands for the predicate molt feliç 'very happy' in (19a), whereas el can not stand for that predicate

(19) a. El Joan és molt feliç.
    (the) John is very happy

b. El Joan ho/l'és.
    (the) John him/it is

In contrast with (18a-b), the argument clitic el in (20b) can not be made to stand for the NP el Joan in the identity sentence (20a), but the predicate clitic ho can:27

(20) a. Aquest home és el Joan.
    this man is (the) John

b. Aquest home *l'/ho és.
    this man him/it is

26. Heggie(1987) and Longobardi(l985) describe similar facts to (18)-(20) for French and Italian, respectively.
27. According to Fabra (1956), the standard normative text for the Catalan language, the judgements for (20b) should rather be

(i). Aquest home l'/ho és.

I have collected the linguistic judgements for these sentences from six Catalan speakers. None agrees with what Fabra (1956) prescribes; five of them agree with the data as I presented above; one of them — maybe under the normative influence of Fabra (1956) — judges that the NP in (20a) can pronominalize as el or ho.
The data (18)-(20) suggest that the NP *el Joan* in (20a) is not an argument but a predicate. If the Disjointness Conditions apply only to argument NPs, then the Disjointness Conditions are irrelevant for the interpretation of (20a). In particular, there would be no conflict with the Disjointness Conditions if we interpreted the two NPs in (20a) in such a way that they determine the same individual in each world in the context.\footnote{I use the word «determine» in an ambiguous way. The two NPs will 'determine' the individual in a different way, since, presumably, they will have a different type of semantic value. If the semantic value of the argument NP *Aquest home* 'this man' is an individual, then the semantic value of the predicate NP *el Joan* 'John' might, for instance, be a set of individuals—actually, a singleton set, which also determines an individual.}

6. Some comments

6.1. The Disjointness Conditions and the Cooperative Principle

That our Disjointness Conditions (B)' and (C)' be concerned with what it is presupposed to be the case, rather than what is actually the case is what allows our proposal to deal with examples like (12). It could seem, though, that this very same feature gives rise to wrong predictions regarding the use of sentences like (21) in certain contexts.

(21) He loves him.

Usually when someone utters (21) the context will contain information about who each NP refers to (say, we are talking about John's feelings for Paul, or we are talking about who loves Clinton and at that point Henry comes into the room, etc). If this contextual information implies that *He* and *him* refer to different individuals (and so that, for each world in the context, the individual that *He* picks up is distinct from the individual that *him* picks up), then we will interpret the sentence with the two NPs referring to two different individuals. This is in agreement with what the Disjointness Conditions prescribe (the Disjointness Conditions prescribe that it not be presupposed that the NPs refer to the same individual, and the contextual information agrees with that since this information will determine that the individuals that the two NPs refer to are presupposed not to be the same).\footnote{Compare: (I) Not to presuppose that p; (II) to presuppose that not-p. (I) does not imply (II), though (II) implies (I) (assuming consistency), so to satisfy (II) is one way of satisfying (I).}

If we hear someone utter (21) and we have to interpret it without being able to use any previous context (say we have just joined an ongoing conversation), though, we will also think that the individuals referred to by *He* and *him* are distinct, not just that it is not presupposed that they are the same. This might seem to be a problem for the formulation of the disjointness conditions that I am defending for the following reason.

If (21) is uttered in a context that contains no information about who the referent of the two NPs is, then it seems that the hearers, in order to interpret the NPs, will have to rely solely on the lexical information and the information
provided by the Disjointness Conditions. The lexical information is just that the individuals the NPs refer to are male and are not the speaker or the hearers though they are somewhat salient at that point in the conversation. The information provided by the Disjointness Conditions is that either (I) it is an open question whether the two NPs refer to the same individual (i.e., in some worlds in the context they pick up the same individual, but in some other worlds they do not), or (II) the two NPs refer to two different individuals (i.e., in each world in the context they pick up different individuals). The Disjointness Conditions do not determine which of (I) or (III) is the case.

So, using solely the Disjointness Conditions (and the lexical information) we could not conclude (II). But, as pointed out above, (II) is what we do conclude when we hear (21) in a context that does not provide any information about the reference of the NPs. This seems to suggest that there is a problem for my formulation of the disjointness conditions and that they should be amended. If what we conclude when we hear (21) in a context that does not provide information on the reference of the NPs is that they refer to two different individuals, and it seems that all the information we can use in such a context is provided by the disjointness conditions (and lexical semantics), then it seems that what the disjointness conditions should prescribe is that (when the relevant syntactic relation obtains) the interpretation of the two NPs must be presupposed to be distinct, not just not presupposed to be the same. (Adopting this modification would, of course, be problematic in other respects since, for instance, we would no longer be able to account for sentences like (12)).

I think, though, that the suggested objection above overlooks one kind of information that will be available to the hearers when (21) is uttered even if the context does not include any information about who the NPs refer to: that the speaker is conforming to Grice's Cooperative Principle. Barring information to the contrary, we will always assume that the Cooperative Principle and in particular, the Gricean maxims, are respected. And we will rely on this assumption when trying to interpret what the speaker says.

If we hear (21) in a context that does not contain information about who the referent of the two pronouns is, but that, otherwise, is a non-special context, we will assume that the speaker knows who she is talking about. So she knows who the person that she is referring to by He is, and who the person that she is referring to by him is, and so she knows whether they are the same or not. If she knew she was talking about one single person, then she would not talk as if the question whether there is one person or two was open, since doing so would go against the maxim of Quantity ('be as informative as is required'). So, since she is treating the question as open, we can conclude that she is not talking about one single person but two.

Of course there can be special contexts (as, for instance, in (12)) in which it is clear that the speaker is unsure about the identity of the individuals she is talking about, or where she has some good reason to act as if she is unsure. This is not, though, what is usually the case. Barring information to the contrary, we will assume that the conversation does not take place in any of these special contexts, and so we will assume that the speaker knows who she is talking about, and that she is open about it. So, assuming Disjointness Condition (B) and the Gricean
maxims, it follows, as desired, that in non-special contexts the hearers will assume that when the speaker utters (21) she is talking about two individuals.

Notice that in a special context where it is reasonable to believe that Gricean maxims do not apply, like, for instance, when an oracle says something, or in the statement of some puzzle in, say, the Sunday edition of a newspaper, we would not conclude from the use of (21) that the individual referred to by \textit{He} is not the individual referred to by \textit{him}, but rather only that it might or it might not be the same.

6.2. Sources of information for the interpretation of NPs

As we have pointed out, there are several sources of information that are used in interpreting the NPs in a sentence, besides the information provided by the Disjointness Conditions on the basis of the structural relation among the NPs. In this subsection we will examine some of the ways of generating the information that, as we have seen, in sentences like the ones in (3)-(5) can conflict with and prevail over the information provided by the Disjointness Conditions.

6.2.1. Presuppositions. The most common way of generating such information is by making it clear that the clause in which certain NP \textit{X} appears expresses some property which is the same as certain property that has been introduced before. The property involves certain individual or certain pattern of linking. So \textit{X} must refer to that individual or be subject to that particular pattern of linking. One way of having this effect is by repeating several times the attribution of a property, so as to create a pattern that will make the listener expect that the next attribution of a property will fit the same pattern (this is illustrated, for instance, in (3c)). Another way of having this effect is by having mentioned the application of the property as a general case and then making clear that we are considering a particular application of that general case (this is illustrated in (3e) and (5a)). One specially good way of having this effect, though, is by using some device which introduces a presupposition. We have already commented (with respect to (3c)) on the effect that the presence of \textit{also} can have. The word \textit{even} (that appears in (3b)) has a similar effect to that of \textit{also}. The sentences in (3)-(5) illustrate, though, other ways of introducing presuppositions besides including some specific word. Consider, for instance, (22):

\begin{equation}
\text{(22) What do you mean no one loves Oscar? He loves him. [2.5]}
\end{equation}

\textit{He} clearly refers to Oscar, which is the only salient individual when the pronoun is uttered. In a more common discourse \textit{him} would be taken as referring to some other individual different from Oscar (because of Disjointness Condition (B)' and the Cooperation Principle). But a sequence of the form \textit{What do you mean \textit{a}? \textit{b}} carries the presupposition that \textit{b} implies \textit{no-a}. In the case of (22) this means that \textit{He loves him} implies that it is not the case that no one loves Oscar. The most simple and likely way for \textit{He loves him} to have that implication is if \textit{He loves him} itself is an attribution of the property of loving Oscar. If this is so, then \textit{him} must refer to Oscar, Disjointness Condition (B)' notwithstanding.
Another way of introducing presuppositions is by stressing some word. Whatever exactly the presupposition induced by stressing an NP is, it includes that the property being attributed to the individual determined by the NP, has already been considered. So in (23)

(23) What do you mean John loves no one? He loves JOHN. [5.7]

the stress on John carries the presupposition that the property of being loved by the subject of the last clause has already being considered. There are two properties that have been considered in the first clause: the property of being loved by John and the property of loving no one. The latter, though, cannot be the one that the stressing of John carries the presupposition about since in the last clause —even before determining who the NPs refer to— it is clear that we are not attributing the property of loving no one. So the presupposition brought about by stressing John, is that, in the last clause, we are attributing the property of being loved by John, that means that He must refer to John, Disjointness Condition (C)' notwithstanding. There is also, of course, the presupposition brought about by the structure What do you mean α? β which adds its effect to the stressing of John.

6.2.2. Demonstration. Another way of generating information that can override the information provided by the Disjointness Conditions is by using an NP demonstratively. (24a,b) illustrate this:

(24) a. He refuted HIM [points to the person] [3.1]
   b. A: John saw Peter.
      B: No, John saw HIM [points to John] [5.3]

In, for instance, (24a) He refers to whoever is most salient individual at that point in the conversation; because of Disjointness Condition (B)', without demonstration him would be assumed not to refer to the same individual; the demonstration, though, forces it to refer to that same individual.

Using an NP demonstratively is a very clear way of indicating what the NP refers to; this is why the intended reference can be communicated even if the Disjointness Conditions are providing opposing information. Notice that most of the sentences considered in the previous subsection involved several devices that together were able to quite successfully override (B)' or (C)' (24a) shows that pointing is effective enough as to have that effect on its own. Still the combination of demonstration with other devices, as in (24b), makes the overruling of (B)' clearer and the sentence better.

6.2.3. Pronominals versus Referential Expressions. A pronoun is very strongly dependent on the context to determine what it refers to. This is not so for proper names and definite descriptions. They are to an extend dependent on the context (there are many people named John and many presidents), but not as much as pronouns are. Proper names and definite descriptions are able, to a good extend, to determine their referent 'on their own'.
It is usually said that 'Principle (C) violations' are less strong than 'Principle (B) violations'. Under my proposal we can explain what motivates this claim without having to accept the queer idea that there is a different degree of prescriptive associated to each of the Disjointness Conditions.\footnote{As we see in the next paragraph, the proponent of a different force associated with each disjointness condition would also have the problem of explaining why some kinds of sentences violating (C) are as bad as their analogues which violate (B).}

Notice that the claim does not seem to be true of sentences violating (C) that include a pronoun but do not contain any indication at all about who the referent of the pronoun is. So, for instance, I think that without any particular previous context, it is as hard and unlikely to understand that the NPs in (25b,c) (principle (C) configurations) refer to the same individual as it is in (25a) (a principle (B) configuration):

\begin{enumerate}
\item He admires him.
\item He thinks that Paul is crazy.
\item He admires Paul.
\end{enumerate}

I think that what motivates the claim that 'violations of (C) are not as bad as violations of (B)' are contrasts like (26a-b):

\begin{enumerate}
\item I talked to Paul today. He admires him.
\item I talked to Paul today. He admires Paul.
\end{enumerate}

It seems that coreference in the last clause of (26b), even if it is still quite bad, is not as bad as in the last clause of (26a). I think we can explain why in the following way: both in (26a) and (26b) the pronoun He would naturally tend to pick up Paul as its referent. In (26a), nothing indicates what the referent of him ought to be; Disjointness Condition (B) indicates that it can NOT be assumed to the same individual He refers to; so, because of the effect of (B), him can not be understood as also referring to Paul. In the second clause of (26b), unlike what is the case with respect to him in (26a), Paul would naturally tend by itself to refer to the Paul that is mentioned in the first clause; because of the effect of (C), though, the most likely way of taking an utterance of (26b) (without any other relevant context), would probably be as Paul in the second clause referring to a person named Paul distinct from the Paul mentioned in the first sentence. In any case, even if the presence of a proper name, which by itself gives clear indication of what its most likely referent is, is not sufficient to neutralize and prevail over the effect of (C), it is enough to diminish it. This makes it easier to understand the last part of (26b) as involving coreference than to understand the last part of (26a) as involving coreference; this, in turn, is what makes us feel that a coreferential interpretation in (26b) is not as bad as in (26a).
6.3. Corroborating Data

The grade that was given to (the correference reading of) the sentences in (27) is exactly what should be expected if the proposal under consideration and the observations in the previous sections are correct:\footnote{31}

(27) a. What do you mean no one loves Oscar? He loves Oscar. [2.9]
    b. What do you mean no one loves Oscar? He loves Oscar. [4.2]
    c. What do you mean no one loves Oscar? He loves him. [2.5]
    d. What do you mean no one loves Oscar? He loves him. [3.6]

Sentence (27a) is better than (27c), and (27b) better than (27d), in accordance to the fact that proper names can determine its referent with much less dependence from the context than a pronoun (and so are less affected by the opposing effect of (B)' or (C)'); (27b) is better than (27a) and (27d) better than (27c), as should be expected if stress introduces the presupposition that the property attributed to the subject of the second clause is a property that has already been considered in the previous clause, and if this presupposition helps to make clearer what the referent of the last NP is, and this, in turn, makes a difference on how acceptable the sentence is. Finally, the sentences in (27) seem to suggest that those devices that can make correference in a sentence more or less acceptable, add to each other: (27d) is better than (27c), and (27b) is even better than (27d). ((27d) includes the use of the presupposition generating structure What do you mean \( \alpha \)? \( \beta \) and the use of stress, whereas (27c) does not contain stress; (27b) in addition to the devices in (27d), also includes the use of a proper name).

7. Two-stages in the interpretation of an utterance

The proposal I present in this paper is not yet complete. We have seen that the basis for explaining the acceptability of sentences like (3b-e) is the fact that the interpretation of these sentences involves an interaction between, on the one hand, the information provided by the Disjointness Conditions and, on the other hand, other information provided by other aspects of the sentence, or by the context. We will see in this section that it is necessary to be more specific about what is involved in this interaction in order to satisfactorily account for the data. In particular we will see that we need to take into consideration the existence of two stages in the interpretation of a sentence. That something must be missing in the proposal as it stands is shown by (28):

(28) I refuted me. [0]

31. I do not pretend that the results of my linguistic questionnaire provide a completely reliable measure of the acceptability of the different sentences. More tests would be needed to support the results I obtained. Still, I think it worth mentioning how well the grades obtained by the sentences in (27) fit with the proposal I am defending.
It is completely clear that the pronouns *I* and *me* in (28) should refer to the same person, the speaker. So it would seem that, according to the proposal I am defending, the effect of Disjointness Condition (B)' in (28) should be overridden by the opposing and unequivocal information about the reference of the pronouns that the very pronouns provide. And so, (28) should be, at least, partially acceptable. (28), though, is completely bad.

It is clear that the process of interpretation of a word, sentence or piece of discourse involves different aspects. For example, part of the interpretation of an utterance of the word *she* will consist in recognizing, for instance, that the word has been used referentially and to interpret the word as an expression that will potentially refer to some human female that is in some way salient at the moment of the utterance. This is part of the interpretation of the utterance but it is not all there is to interpreting the utterance. Another aspect of the interpretation is to determine which specific individual the word refers to (for instance, to determine that it is Delia that the speaker refers to by his use of *she*).

In order to satisfactorily account for the phenomena related to the Disjointness Conditions we must take into consideration the existence in the process of semantic interpretation of the two stages suggested in the previous paragraph: in the first stage only structural, lexical and general semantic information is used, and a partial interpretation is produced. For instance, a partial interpretation of a referential use of a pronoun like *he* would identify it as a referential expression that picks up a male human who is in some way salient.32 The partial interpretation for a sentence like *He loves him* would be the pseudo-proposition that a not-yet determined male human loves a not-yet determined male human who it is not assumed to be the same as the first one.33 This partial interpretation is produced on the sole basis of the lexical information of the words and the semantic rules—including the Disjointness Conditions (which depend, in turn, on the structural relationship between the NPs).

In the second stage the information from other parts of the discourse and from the context is brought in and the complete interpretation is produced. For instance,
in the second stage of the interpretation of an utterance of *he*, the contextual information that, say, John is a salient human male and the individual that is most relevant for the topic that is currently discussed is used in order to determine, on the basis of the partial interpretation of *he* produced in the first stage, that the speaker referred to John. In the second stage of the interpretation of *He loves him*, discourse and contextual information (like for instance that the topic under discussion is James' feelings for John) is applied to the first stage's partial interpretation in order to produce the final interpretation of the sentence (for instance that James loves John).

It might be that the contextual information that is used in the second stage is in conflict with some aspect of the partial interpretation produced in the first stage. The observations about how the conflict might get resolved made in the previous sections apply here: the existence of the conflict might make the sentence to be regarded as bad, or, if the opposing contextual and discourse information brought in at the second stage is clear and strong enough, the sentence will be interpreted in accordance with the contextual and discourse information and will be regarded as (at least partially) acceptable.

What we just said about the possibility of resolving conflicts between the product of the first stage and the contextual information brought in at the second stage is perfectly compatible with things working differently when the conflict is internal to the first stage. It does not seem unreasonable to suppose that when there is a conflict in what the semantic rules and the lexical information provided by the different items prescribe, then no coherent partial interpretation is possible, and the sentence is regarded as unacceptable.

This would explain the complete unacceptability of *I refuted me*. The lexical information associated with *I* determines that the word refers to the speaker, and so does the lexical information associated with *me*. So, it is part of the partial interpretation of the first stage that the two pronouns should be understood as referring to the same individual, whichever the actual world turns out to be. On the other hand, given the structural relationship between the pronouns, Disjointness Condition (B)' prescribes that the referent of the two pronouns can not be assumed to be the same. The conflict prevents a coherent representation at the first stage, and this makes the sentence unacceptable.

By referring to the two stages as 'first' and 'second' I am, of course, not claiming that this is a temporal description of empirical processes. The first stage is previous to the second just in the logical sense that the second stage presupposes the first, since the partial interpretation which is the product of the first stage is one of the elements used in the second stage to provide the final interpretation. It might very well be that, when an actual evaluation of the sentence takes place, processes corresponding to the second stage take place at the same time as processes in the first, in the same way in which semantic processes might take place before a logically previous process (for instance a syntactic process like, say, determining which was the first word of the sentence uttered —as when we have not clearly heard what was uttered) has been completed.

Someone might object to my having regarded the evaluation of presuppositions as pertaining to the second stage where contextual information is taking into account. It might be claimed that, since presuppositions are introduced by some
specific words or forms of sentences, they should be regarded as part of what is evaluated at the first stage when the lexical information, the structure of the sentence and the semantic rules are taken into account. If presuppositions are evaluated at the first stage then, for many of the examples we have considered, the conflict between different sources of information would already arise at the first stage. So, there would be no justification for my claim that the conflict of information that explains the (partial) acceptability of sentences like (3b-e) is of a different sort from the conflict of information that makes (28) bad.

We can see that this objection is not correct by taking into account one of the points made by Saul Kripke in Kripke (1990): some presuppositions involve an anaphoric element, analogous to some uses of pronouns. If someone says «Paul also went to the movies» the sentence involves an anaphoric reference to certain proposition, like for instance, that Paul went to the library, or that Arthur went to the movies. We say that some presuppositions involve an anaphoric element in the following sense: the full interpretation of the sentence carrying the presupposition requires the identification of one proposition that has certain characteristics and that is in some way salient. The full interpretation of a referential use of the pronoun he requires the identification of an individual who is male and that is in some way salient (either by having being just mentioned or in some other way). Analogously, the full interpretation of an utterance of the sentence «Paul also went to the movies» requires the identification of a proposition with certain characteristics (it has to involve the attribution of a property to Paul, or one attribution of the property of going to the movies) and which is salient in some way — either by having been recently expressed or in some other way (for instance, by the fact that the speakers are watching a video where they can see Paul going to the library). In the same way in which the determination of the referent of a pronoun is carried out in the second stage of the interpretation where contextual information is brought in, it also belongs to the second stage the determination of what the particular proposition is that is being anaphorically referred to by a sentence carrying a presupposition which arises because of the presence of words like also or even, or the stressing of some word. Similarly for the presupposition that plays a role in interpreting the NPs in β in structures like What do you mean α? β, which also depends on identifying a proposition (the one expressed by α) that has been previously expressed.

8. Conclusion

There are two main ideas we have argued for in this paper. First, the disjointness conditions should not be formulated in terms of the relation of actually having the same value, but rather, they should take into consideration what the speakers believe to be the case. Following Postal (1970) and Heim (1988, 1993), we formulate the disjointness conditions in terms of the relation of it being part of what the speakers are assuming, or context, that the expressions have the same value. This requires using a background semantic theory that enables us to make the desired distinctions. Stalnaker (1979) provides one such adequate semantic framework. Second, in order to account for the judgements of the speakers regarding the acceptability of different sentences we do not need to provide a set of disjointness conditions that allows exactly those sentences and interpretations of sentences that the
speakers declare as acceptable. Their judgements do not directly reflect their (implicit) knowledge of the disjointness conditions. The data seems to be best explained if we take into consideration that the disjointness conditions are just one among several sources of semantic information that speakers use in trying to come up with an interpretation for a sentence. The data supports the idea that a sentence will be regarded as at least partially acceptable when understood in a way that conflicts with what the disjointness conditions prescribe if the sources of semantic information other than the disjointness conditions determine such an interpretation in a clear enough way. Viewing things this way enables us not only to account for the sentences that have been offered as counter-examples to Standard Binding Theory, but also to explain their different degrees of acceptability.

References


