

# The Information Structure of It-clefts, Wh-clefts and Reverse Wh-clefts in English<sup>1</sup>

Nancy Hedberg and Lorna Fadden  
Simon Fraser University

## 1. Introduction

In addition to expressing a proposition monoclausally as shown in (1), English permits a proposition to be split into two clauses and expressed in three types of cleft sentence, which we term 'it-clefts', 'wh-clefts' and 'reverse wh-clefts'. An example of each is shown in (2):

- (1) People are going to be voting on the economy, come November of next year.
- (2) a. It's the economy that people are going to be voting on, come November of next year. (IT-CLEFT)  
b. What people are going to be voting on, come November of next year, is the economy. (WH-CLEFT)  
c. The economy is what people are going to be voting on, come November of next year. (REVERSE WH-CLEFT)

Following Hedberg (1988, 1990, 2000) we label the phrase 'the economy' as the 'clefted constituent' in all three cases, and the phrase 'that people are going to be voting on, come November of next year' or 'what people are going to be voting on, come November of next year' as the 'cleft clause'.

Following Gundel and Fretheim (2004) and previous works by Gundel, we distinguish between referential and relational givenness in our treatment of information structure, although the two are linked in that relationally given expressions in general (i.e.

---

<sup>1</sup> This paper is a major revision of Hedberg (1988), which examined the discourse function of it-clefts, wh-clefts and reverse wh-clefts in a corpus of natural examples drawn from videotapes and transcripts of the McLaughlin Group broadcast by the Public Broadcasting Service in the United States during the years 1987-1988. For this paper, we collected new data from episodes of the McLaughlin Group aired in 2001-2002. We changed our database because we wanted more recent and currently relevant political discourse to discuss, and we wanted to use transcripts that are available on-line. The two studies differ in that instrumental means are used for the analysis of prosody in the current paper, whereas the prosody of the clefts in the 1988 paper was based on the author's intuitions. The major inspiration for this revision comes from Jeanette Gundel, who was a close advisor in the writing up of Hedberg (1988).

We thank Juan M. Sosa for his help with the prosodic analysis of our cleft examples, and thank Maite Taboada and Ron Zacharski for criticism and help with the argumentation. This research was supported by a Discovery Park Research Grant and a Small SSHRC Research Grant awarded to Nancy Hedberg and Juan Sosa by Simon Fraser University.

topics) have to attain a certain level of referential givenness (i.e. cognitive status--Gundel, Hedberg and Zacharski 1993), in order to function cognitively as ‘the peg on which the message is hung’ (Halliday 1970).<sup>2</sup> A topic is further constrained pragmatically in that it must evoke a relevant question for the comment to make a contribution ‘about’.

With regard to referential givenness, we assume that all clefts function analogously to definite descriptions, in that the cleft clause is presupposed and hence must be at least uniquely identifiable in cognitive status terms (c.f. Hedberg 2000 for it-clefts), while the clefted constituent presents an exhaustive element, a ‘kontrast’ in the terms of Vallduvi and Vilkuna (1998). These characteristics are at least part of what distinguish clefts from non-cleft sentences.

Our main claim with regard to the difference between the three types of clefts is that wh-clefts are more constrained in relational information structure than it-clefts or reverse wh-clefts. The initial element—the cleft clause—in a wh-cleft always presents the topic of the sentence, and the clefted constituent always presents the focus, with the comment being the identification of the variable in the topic with the focus. It-clefts and reverse wh-clefts are freer in their relational information structure, with the initial clefted constituent presenting either the topic or the focus, and the subsequent cleft clause presenting the focus or topic, respectively. We subsequently refer to these two possibilities as ‘topic-comment’ or ‘comment-topic’ clefts. We also claim that all-comment clefts are possible.

We rely primarily on a spoken corpus of 98 clefts from spontaneous televised political discussions, examining their discourse function in context and their prosody. We supplement this corpus with examples from a previously collected corpus of the televised political discussion program, some constructed examples and examples from a variety of written sources.

Although we claim that it-clefts and reverse wh-clefts are identical with regard to relational information structure, they do differ in distribution. The initial clefted constituent in an it-cleft is most likely to be a full NP while the initial clefted constituent in a reverse wh-cleft is most likely to be an anaphoric demonstrative pronoun. Also, question words can serve as the clefted constituent in it-clefts, but not in wh-clefts or reverse wh-clefts; and verb phrases and adjective phrases can serve as the clefted constituent in wh-clefts and reverse wh-clefts, but not in it-clefts (see Delin 1989 and Hedberg 1990 for discussion). We have only eight examples of it-clefts in our corpus compared to 65 examples of reverse wh-clefts, and thus concentrate primarily on the difference between the wh-clefts and reverse wh-clefts in the corpus discussion.

---

<sup>2</sup> Gundel (1985, 1988) suggests that a topic must be familiar to the hearer in order to serve as an address in memory to which the new information may be linked. In more recent work (e.g. Gundel 1999b) she acknowledges the possibility that a topic may be merely referential if the phrase expressing this topic is anchored to an activated element. Reinhart (1981) and Prince (1985) present examples which they claim are referential indefinite topics.

## 2. Method for Corpus Study

### 2.1 Data Collection

The data for the study were collected from 13 videotaped episodes of the McLaughlin Group, a half-hour, weekly televised PBS program in which a host, John McLaughlin, facilitates unscripted, spontaneous discussion on current political affairs with a panel of prominent journalists, almost all of whom are speakers of American English. There are significant advantages to deriving data from this particular source. First, transcripts available on the World Wide Web make it convenient to examine lengthy contexts, and second, sound recordings allow for prosodic examination of the data.

The episodes used in this study spanned from March 2001 to June 2002. From these video recordings, 98 cleft samples were put onto a CD rom using a Marantz CD Recorder.<sup>3</sup> These samples were then converted from analog to wav format using Gold Wave v.5.0. Pitch contour and primary and secondary stress were examined using Praat v.4.2.07. and Pitchworks v.5.0.

Once collected, the cleft samples were sorted into one of three categories: it-cleft, wh-cleft, and reverse wh-cleft. Table 1 shows the number of each cleft type.

Cleft Type	Number	Percent
It-cleft	8	8%
Reverse wh-cleft	65	66%
Wh-cleft	25	26%
<b>TOTAL</b>	<b>98</b>	<b>100%</b>

Table 1. Distribution of Cleft Types in McLaughlin Data

Hedberg (1988) examined 343 clefts in the previous McLaughlin corpus, finding only 12 (3%) it-clefts, a pattern similar to what we see here. However she found 182 (53%) wh-clefts and 149 (44%) reverse wh-clefts, which is different from the pattern observed in the current study where we have many more reverse wh-clefts. We have no explanation for this difference, except to note that the sample size was larger in Hedberg (1988).

In the following subsections we explain how the data were coded and analysed, and we provide examples of clefts with respect to their referential givenness structure (cognitive status), relational givenness structure (topic-comment or comment-topic status), and prosody.

### 2.2 Labelling the data

---

<sup>3</sup> One wh-cleft and one reverse wh-cleft were discarded because they were incomplete. There were two predicational th-clefts and one predicational it-cleft (c.f. Hedberg 2000) in the data, which we decided to not discuss in this paper because we didn't look for predicational wh-clefts or reverse wh-clefts. All of the clefts that we analyze are specificational.

### 2.2.1 Prosody

The first step in labelling our data was to identify the prosodic patterns in each sample. We primarily coded stress, but supplemented this with intonation coding in the examples especially relevant to the discussion. The example in (3) shows how stress was labelled.

- (3) Mr. Blankley: I think **that** kind of **LEADERSHIP** is *exactly* what we **need**. [11/3/01]

Primary stress is indicated in upper case bold, secondary stress is indicated in lower case bold italics, and other pitch accented words are indicated in regular lower case bold.

The ToBI transcription system (Beckman and Ayers-Elam 1997) was used to label intonation patterns when relevant. This system uses starred combinations of high and low tones to indicate pitch accents and (H\*, !H\*, L\*, L+H\*, L\*+H, H+!H\*). !H\* indicates a downstepped high tone. There are two phrase accents (L- and H-) and two boundary tones (H% and L%). A combination of a phrase accent followed by a boundary tone indicates the direction of pitch movement at the end of an intonational phrase, and a phrase accent alone indicates the end of an intermediate phrase. Examples (4) and (5) illustrate this system of coding.

- (4) Mr. O'Donnell: **THAT'S** what we **lack** in **Africa** now.  
H\* !H\* !H\* L-L% [4/14/01]
- (5) Mr. Blankley: **They're** the **ones** who were **pulling** the **STRINGS** with  
L+H\* !H\* H\* L- H\*L-  
**Yeltsin.**  
L+H\*L-L% [10/19/01]

When intonation is especially relevant to the following discussion, we will supplement the ToBI coding with pitch tracks to better illustrate how the example was pronounced.

### 2.2.2 Information Structure

The second step in coding our data was to label the clefted constituents and cleft clauses according to referential givenness structure and relational givenness structure.

For referential givenness we used ten categories derived loosely from the Givenness Hierarchy categories of Gundel et al. 1993, and classified the two major constituents of each cleft utterance as either directly having a certain cognitive status or as being inferrable from an element having one of the cognitive statuses.<sup>4</sup> We added the

---

<sup>4</sup> The cognitive statuses of Gundel et al. (1993) are arranged in a unidirectional entailment hierarchy, e.g. everything that is activated is familiar, but not everything familiar is activated. In the set of labels used here we labelled with the maximal obtaining status, so 'familiar', for example, means 'familiar but not activated.' We use

categories ‘cataphoric’ for constituents whose referents were not introduced until after the utterance of the cleft was completed, and ‘question-word’ for wh-words that function to express the quantificational ‘unknown’ portion of a wh-question. We divided the ten categories into the two supercategories of ‘discourse old’ elements or ‘discourse new’ elements, following Prince (1992). The resulting coding system is shown in Table 2.

Discourse Old	Activated
	Recently activated
	Inferrable from activated situation
	Inferrable from activated proposition
	Inferrable from recently activated proposition
Discourse New	Familiar
	Inferrable from familiar proposition
	Informative
	Cataphoric
	Question-word

Table 2. Referential Givenness Categories

For example, the cleft clause in (6) was labelled ‘activated’ because the proposition that the resources are going somewhere was immediately activated, whereas the cleft clause in (7) was labelled ‘recently activated’ because the proposition that someone says this is loose talk was activated many discourse turns away although in the same discussion.

- (6) Mr. McLaughlin: Don’t you think resources are really being redirected, for example, into security and into counterterrorism? **That’s where the resources are GOING**. Those are going to be the economic winners. [10/26/01]

the label ‘informative’ for a presupposed proposition that is unfamiliar to the listener, following Prince’s (1978) work on it-clefts. Hedberg (2000) would label these presuppositions as ‘uniquely identifiable but not familiar’ in the terms of Gundel et al. (1993). We also use the label ‘informative’ for the lowest cognitive statuses, ‘referential’ and ‘type identifiable.’ Note that our label ‘recently activated’ would be labeled ‘familiar but not activated’ by Gundel et al. (1993) since the information is no longer in short-term memory. We also treat inferrable propositions as falling under the same broad category as the proposition they are inferrable from, following Prince (1992) who treats NP referents inferrable from activated referents as ‘discourse old’. We thus here differ from Gundel et al. (1993, 2001) who treat NP referents and also cleft clauses as uniquely identifiable but not familiar when they are only ‘indirectly’ related to their antecedent. We do this, first of all, because so many cleft propositions in our corpus are only indirectly related to antecedent propositions (see Table 4 below), but especially because in this chapter we are trying to synthesize the work of Prince (1978, 1985) and Gundel (1985, 1988) on it-clefts and wh-clefts.

- (7) Ms. Clift: ... The administration is not going to be able to get away with this one, declaring an American-born – a Brooklyn-born boy an enemy –
- Mr. Barone: Well, he’s an adult.
- Ms. Clift: -- an adult – an enemy combatant, holding him indefinitely, without legal representation, without any charges, where even the deputy secretary of Defense, Paul Wolfowitz, has said this was nothing more than loose talk. The administration is really going to have to come up with some evidence here, or they’re going to have to let this guy go.
- (39 turns)
- Ms. Clift: Wolfowitz is an even bigger hawk than you are, Michael, and HE’s the one who says this is loose talk. [6/14/02]

For relational givenness we labelled each cleft for the topic-comment status of its two parts, distinguishing between discourse old and discourse new topics and comments, by using lower case ‘t’ or ‘c’ for the former and upper case ‘T’ or ‘C’ for the latter. We also allowed for the possibility that a cleft utterance be all comment. As will be discussed further below, the primary criterion for determining topic versus comment classification was that of prosody. We assumed that the primary stress of the cleft utterance falls on the comment. We also relied on the criterion of determining what the utterance as a whole seemed to be primarily ‘about’ in the extended context in which the example appeared, specifying this portion of the cleft as the topic. Thirdly, relational givenness correlated approximately, but not entirely, with referential givenness in that in most, but not all cases, the referentially more ‘given’ constituent in the cleft utterance ended up being labelled the ‘topic.’

Thus, examples (4) and (7) above with primary stress on the clefted constituent of the reverse wh-cleft were labelled comment-topic clefts, and examples (5) and (6) where primary stress fell on the cleft clause of the reverse wh-cleft were labelled topic-comment clefts.

One type of reverse wh-cleft that was difficult to classify relationally was labelled ‘tag’ due to the presence of a sentence final tag explicating the content of the subject. Many such tag examples appear to be a kind of blend of a reverse wh-cleft with a comment cleft clause in relation to the clefted constituent and a wh-cleft with a topical cleft clause in relation to the stressed tag. An example is shown in (8). Tag clefts will be discussed further in section 4.3.

- (8) Mr. Blankley: **That’s** what *I’m* **TOLD**, that they got it from the **SENATE side**. [10/19/01]

In the next section, we turn to the results of coding our corpus data for referential givenness, syntactic type and relational givenness of the two major parts of the three types of clefts. We are especially interested in which of these three coding criteria best explains the data and the language system generating the data.

### 3. Corpus Study Findings

**3.1 Referential Givenness.** Table 3 shows the referential givenness coding of the clefted constituent for all three types of clefts. It can be seen that that reverse wh-clefts and wh-clefts are almost mirror images of each other. Roughly, when the clefted constituent is activated, the reverse wh-cleft form occurs. When the clefted constituent is familiar or informative, the wh-cleft form occurs. When the clefted constituent is recently activated, either form is chosen.

	IT	REV	WH	TOTAL
Activated	3	58	-	<b>61</b>
Recently activated	1	4	2	<b>7</b>
<i>Total Discourse Old</i>	<i>4</i>	<i>62</i>	<i>2</i>	<b>68</b>
Familiar	2	-	7	<b>9</b>
Informative	-	-	16	<b>16</b>
Cataphoric	-	3	-	<b>3</b>
WH	2	-	-	<b>2</b>
<i>Total Discourse New</i>	<i>4</i>	<i>3</i>	<i>23</i>	<b>30</b>
<b>GRAND TOTAL</b>	<b>8</b>	<b>65</b>	<b>25</b>	<b>98</b>

Table 3. Referential Givenness Status of the Clefted Constituent

Table 4 shows the referential givenness coding of the cleft clause in all three types of clefts. It can be seen that it-cleft clauses are evenly divided between discourse old and discourse new information, while wh-cleft clauses are predominately discourse old (88%). Reverse wh-cleft clauses are also predominately discourse old (71%) but not to quite the same extent.

	IT	REV	WH	TOTAL
Activated	1	3	-	<b>4</b>
Inferrable from activated situation	-	10	4	<b>14</b>
Inferrable from activated proposition	1	19	11	<b>31</b>
Recently activated	1	7	4	<b>22</b>
Inferrable from recently activated proposition	1	7	3	<b>11</b>
<i>Total Discourse Old</i>	<i>4</i>	<i>46</i>	<i>22</i>	<b>72</b>
Familiar	1	7	1	<b>9</b>
Inferrable from familiar proposition	3	4	1	<b>8</b>
Informative	-	8	1	<b>9</b>
<i>Total Discourse New</i>	<i>4</i>	<i>19</i>	<i>3</i>	<b>26</b>
<b>GRAND TOTAL</b>	<b>8</b>	<b>65</b>	<b>25</b>	<b>98</b>

Table 4. Referential Givenness Status of the Cleft clause

Although it is not entirely clear from the preceding two tables, which don't compare the referential givenness status of the clefted constituent compared to the cleft clause in each cleft directly due to the need for readability, a constraint on constituent order in terms of referential givenness can be extracted from the full set of data if we allow inferrable or indirectly anaphoric cleft clause information to count as (at least) one level less given than the corresponding directly anaphoric information. This constraint is that referentially more given information (e.g. activated information) tends to precede referentially newer information (e.g. inferrable from activated information).<sup>5</sup>

An exception to this constraint is the cataphoric reverse wh-clefted constituents, which we classified as discourse new, but which precede more given cleft clauses. Since these clefted constituents are proximal demonstrative pronouns and are thus one word pronominal forms, perhaps it is this syntactic 'lightness' that best accounts for the word order pattern found in the data. In the next section we explore this possibility more carefully.

### 3.2 Syntactic Type of Clefted Constituent

Table 5 shows the syntactic type of the clefted constituent in all three types of cleft. It can be seen that except for full NP clefted constituents, reverse wh-clefts and canonical wh-clefts are again in complementary distribution. Reverse wh-clefted constituents tend to be short and anaphoric or cataphoric, whereas canonical wh-clefted constituents tend to be longer verb phrase and clausal constituents. The it-clefts in our data either had full NP or question word clefted constituents, both of which are also short.

	IT	REV	WH	TOTAL
Question word	2	-	-	<b>2</b>
Relative pronoun	-	8	-	<b>8</b>
Personal pronoun	-	4	-	<b>4</b>
<i>that</i>	-	44	-	<b>44</b>
<i>this</i>	-	3	-	<b>3</b>
Full NP	6	5	7	<b>18</b>
Adverb	-	1	-	<b>1</b>
VP	-	-	6	<b>6</b>
Clause	-	-	11	<b>11</b>
Gerund	-	-	1	<b>1</b>

<sup>5</sup> Thus 'inferrable from activated proposition' or 'inferrable from activated situation' would count as (at least) one level less given than 'activated', and so on, with both being classified as 'discourse old'. This system follows Prince (1992) in treating entity-level 'inferrables' as 'discourse old' and also follows Hedberg (1990) in treating 'indirectly activated' cleft clauses in the same general category as 'activated' ones. As mentioned in footnote 4, Gundel et al. (1993, 2001) don't follow this system for either NP or cleft clause referents, instead classifying the indirect referents as only 'uniquely identifiable' (i.e. as discourse new). If the Gundel et al. system is followed, the 'given' before 'new' linear order for clefts of course falls out directly.



<b>TOTAL</b>	<b>8</b>	<b>65</b>	<b>25</b>	<b>98</b>
--------------	----------	-----------	-----------	-----------

Table 5. Syntactic Type of the Clefted Constituent

A similar pattern of results has been found in other corpus studies. For example, Delin (1989) found that 102 out of 160 spoken and written reverse wh-clefts had *that* as the clefted constituent and 43 had *this*, while none had clauses or verb phrases. In addition, 52 out 162 wh-clefted constituents were clausal, 38 were VPs, and none were demonstrative pronouns. 158 out of 245 it-clefted constituents were NPs.

It should be noted that wh-clefts can always be flipped and turned into reverse wh-clefts without degradation in grammaticality, although sometimes minor changes need to be made such as adding a complementizer or changing the position of an adverb. However, reverse wh-clefts with anaphoric pronominal clefted constituents can never be flipped without extreme awkwardness, unless the clefted constituent is highly stressed<sup>6</sup> (and some further minor changes are made in some cases). Compare (9b) and (10b):

- (9) a. Mr. Zuckerman: What I think you have to **appreciate** about this **program** is that **Pat does set the standard for CIVILITY**. [5/24/02]
- b. That Pat does set the standard for civility is what I think you have to appreciate about this program.
- (10) a. Mr. Blankley: It's got to be standardized, and **THAT'S** what the **President's proposal** will **do**. [10/26/01]
- b. ?? It's got to be standardized and what the President's proposal will do is that.

The actual pattern of results found in the data between reverse wh-clefts and canonical wh-clefts, then, could be attributed to a syntactic constraint that short constituents should precede potentially long constituents. However, wh-clefts with relatively short clefted constituents are grammatical and even occur in the data, as shown in (11), so a short-before-long syntactic constraint does not account for the language system or even the data.

- (11) Mr. Zuckerman: So let me just say, what they **resent** more than anything else is the **success** of **AMERICA**.

<sup>6</sup> Note that this example is also fine if the clefted constituent is modified with *just* or *exactly*. (We thank Maite Taboada for this observation.) We suggest that such focus particles contribute to the interpretation of the clefted constituent as a stressed focus, and that this explains the improvement.

Furthermore, the reverse wh-clefts with cataphoric clefted constituents in the data can easily be flipped, even though the final clefted constituent would be short. Compare (12a) and (12b):

- (12) a. Mr. McLaughlin: THIS is what he is talking about. He's talking about bringing *homeland* security **ABROAD**.  
[5/24/02]
- b. What he's talking about is THIS. He's talking about bringing homeland security abroad.

In the following subsection we will examine the relational givenness status of the different parts of each type of cleft exhibited in the corpus data and relate these to the referential givenness status, and in the remainder of the paper will argue that relational information structure explains several aspects of the distribution of the three types of clefts. Specifically we will argue that wh-clefts are exclusively topic-comment structures, whereas reverse wh-clefts and it-clefts can be either topic-comment or comment-topic structures.<sup>7</sup>

### 3.3 Relational Givenness

Table 6 presents the results of the relational givenness coding of the data in the corpus. Recall that lower case indicates discourse old information and capital letters indicate discourse new information. Cataphoric elements were classified as discourse new.

	IT	REV	WH	TOTAL
tc	1	27	-	<b>28</b>
tC	1	16	23	<b>40</b>
Tc	-	1	-	<b>1</b>
TC	-	-	1	<b>1</b>
ct	1	10	-	<b>11</b>
Ct	2	1	-	<b>3</b>
all C	3	-	1	<b>4</b>
tag	-	10	-	<b>10</b>
<b>TOTAL</b>	<b>8</b>	<b>65</b>	<b>25</b>	<b>98</b>

Table 6. Relational Givenness Structure

<sup>7</sup> We ignore all-comment clefts here as we don't have enough examples in our data to draw firm conclusions. We would predict that wh-clefts can't be all-comment (which goes against our first coding of one example as discussed below) but that it-clefts and reverse wh-clefts can be all-comment. We find evidence in our data only for the possibility of all-comment interpretations of it-clefts, not reverse wh-clefts.

Discourse-new final topics don't occur in the data. This is predicted by the interaction of the two principles formulated in Gundel (1988): the Given Before New Principle ("State what is given before what is new in relation to it"), and the First Things First Principle ("Provide the most important information first"). The Given Before New Principle implies that topics should precede comments, all things being equal, and the First Things First Principle implies that new topics and foci should come first since both are crucially important to how the message is understood. Thus, a new topic before a comment (TC) follows both constraints and should thus be possible, or maximally optimal. Furthermore, tC and Ct (where the topic is activated or 'given') each follow one constraint and thus are predicted to be possible, or relatively optimal. However, CT violates both constraints and is thus predicted to be impossible, or non-optimal. We see the prediction followed in our data, that final topics are always discourse old, thus supporting Gundel's hypothesis concerning the interaction between the two principles.

In our data, it-clefts exhibit the most variation. We illustrate the three major types with a Ct, a tc and an all-C it-cleft in (13)-(15). In (13), primary stress falls on *ACLU*. The clefted constituent was coded as 'familiar' and the cleft clause was coded as 'inferrable from a recently activated proposition'. The topic is more strongly activated than the comment. The discussion was about United States authorities holding a terrorist suspect for five weeks without legal representation and what type of public reception this action would incur.

- (13) Ms. Clift:        And it's *not* only the **ACLU** that's going to be at the **barricades on this one**, but conservative Republicans who don't want the government taking their guns are not going to like the way this case is handled, either. [6/14/02]

In (14), primary stress falls on the cleft clause. In this case, immediate discussion was about China's forcing down an American military plane and detaining the people on board. It had been suggested several turns previously that the United States was behaving in a Cold War fashion by flying such reconnaissance missions in the first place, and Pat Buchanan was objecting to this statement. We coded the clefted constituent as 'activated' and the cleft clause as 'recently activated'. Again the topic is more strongly activated than the comment.

- (14) Mr. Buchanan:        It is the ***Communist Chinese*** who are behaving as a **Cold War POWER** right now. [4/6/01]

We coded (15) as an 'all-comment' it-cleft because both parts of the cleft were familiar but discourse new. Ms. Clift was talking to Pat Buchanan and reminding him of President Nixon's actions in normalizing relations with China. If there is a topic expressed in this utterance, it is Pat Buchanan himself, the referent of *your*.

- (15) Ms. Clift:        Well, I think it was **your PRESIDENT, your past EMPLOYER**, who opened the **door** for us to China. [4/6/01]

Turning to reverse wh-clefts, we classified 44 (80%) of them as topic-comment clefts, as opposed to 11 (20%) as comment-topic reverse wh-clefts. This pattern is similar to the findings of Hedberg (1988) who also found more topic-comment reverse wh-clefts (66% as opposed to 34%). Both findings verify the claim made in Heycock and Kroch 2002 that in reverse wh-clefts, the clefted constituent typically expresses the topic and the cleft clause the focus, whereas in canonical wh-clefts, the clefted constituent necessarily expresses the focus.

Topic-comment as opposed to comment-topic reverse wh-clefts were illustrated above in (6) and (10). In (6), both parts of the reverse wh-cleft are activated, but we perceived primary stress to fall on the cleft clause, so we coded it as a topic-comment reverse wh-cleft. In (10), we classified the clefted constituent as ‘activated’ since it referred to the state-of-affairs introduced by the immediately preceding clause, and we classified the cleft clause as ‘inferred from an activated proposition’ since the president’s proposal to privatize the airline security system had been discussed, and it can be inferred that this proposal will do something. In this case primary stress fell on the clefted constituent even though it was more strongly activated than the cleft clause. We therefore classified this as a ct cleft.

Furthermore, in (12), the clefted constituent is cataphoric and its referent is thus new to the discourse, while the cleft clause is inferable from the speech situation in which Sam Nunn talking was just played on videotape. Primary stress falls on the clefted constituent, which is less activated than the cleft clause, and we classified this example as Ct.

There were five examples of reverse wh-clefts with informative relative clauses. One example is shown in (16):

- (16) Mr. Blankley: ...I agree its going to be a fight, because we heard Senator Daschle say another silly statement, he’d rather invest in Social Security than Nasdaq. Over any relevant period of time, that’s idiocy, because the Nasdaq’s gone up in the past, it’ll go up in the future. He’s just trying to key into the news cycle, which is what the Democrats are going to try to do while the markets aren’t doing well. Over any long period of time, the stock market goes up well, over the history of the country, and **that’s why privatization makes SENSE**. [5/4/01]

Here the discussion was about privatizing Social Security, and Mr. Blankley was voicing for the first time his opinion that privatization makes sense for the reason that he articulates.

The possibility of a cleft presupposition being informative was first introduced in Prince (1978) for it-clefts. Her most famous example is shown in (17):

- (17) ## It was just about 50 years ago that Henry Ford gave us the weekend. On September 25, 1926, in a somewhat shocking move for that time, he decided to establish a 40-hour work week, giving his employees two days off instead of one. [Philadelphia Bulletin, 1/3/76, p. 3L]

The double crosshatch indicates that the example occurred at the beginning of the article. She concludes that the presupposition ‘that Henry Ford gave us the weekend (at some time)’ is presented as new information to (most) readers, although a known fact historically. A similar explanation can be inferred from Mr. Blankley’s reverse wh-cleft in (16): he is trying to establish that his opinion has a basis in fact, which is why he uses a cleft to express this presupposition. Prince denies the possibility of performing this informative function to wh-cleft presuppositions, concluding instead that wh-clefts mark the information as “assumed/assumable to be in the hearer’s consciousness, or given” [p. 904].

Turning now to wh-clefts in our data, all but two of the wh-cleft examples were classified as having a ‘discourse-old’ cleft clause, in accordance with Prince’s (1978) conclusion about wh-cleft presuppositions being ‘given’. (Note that the labels in Table 6 specify the linear order of the two parts of the different kinds of clefts, so that it is the cleft clause whose status is reported first in the case of wh-clefts as opposed to the other two kinds of cleft.) All of the wh-clefts had discourse-new final comments. A typical example is shown in (15):

- (18) Mr. McLaughlin: I think what he should be *congratulated* for is, under **those** circumstances, it was an *unerring PITCH*. [11/3/01].

Here we classified the cleft clause as ‘inferred from an activated proposition’ since the participants were all approving of President Bush’s throwing out the first pitch in a World Series baseball game soon after the terrorist attack on 9/11. We classified the clefted constituent as ‘familiar’. The topic here is more activated than the comment.

There were quite a few cases of informative wh-clefted constituents as well, as in (19), where ‘political coverage for officials’ is completely new to the participants and the cleft clause had been recently activated.

- (19) Ms. Clift: What **these warnings** have **achieved** is political *coverage* for **OFFICIALS**. [11/3/01]

There was one example, shown in (20), in which we coded the cleft clause as familiar but not activated:

- (20) Mr. Zuckerman: What I think you have to **appreciate** about this **program** is that *Pat does set* the **standard** for **CIVILITY**. [5/24/02]

Here the proposition ‘that you have to appreciate something about this program’ had not yet come up during the discussion. It is interesting to speculate as to whether the clausal content here should perhaps rather be classified as ‘inferred from the discourse situation’—and thus discourse old and more-or-less classifiable as ‘in the hearer’s consciousness’ as Prince’s (1978) constraint on wh-cleft clauses would predict—or as ‘familiar’ and thus ‘discourse new’ as we coded it. We suggest that the content of the cleft clause is indeed discourse new but is a relevant new topic, and that it is this relational as opposed to referential givenness that licenses use of the wh-cleft here.

There was also one case of a wh-cleft that we classified as ‘all-comment’ since both parts of the cleft were informative. This example is shown in (21):

- (21) Mr. McLaughlin: All right, let me ask you this question. You know that the president, President Bush, is preparing his own peace package proposal. Do you know anything about the elements of that?
- Ms. Clift: Well, what’s been *leaked* to the New York Times
- Mr. McLaughlin: What do you know?
- Ms. Clift: —is that he will favor—he favors a Palestinian STATE.  
[10/19/01],

Here, it is clearly referentially new information that something was leaked to the New York Times as well as new information that President Bush favors a Palestinian state, so we classified it as all comment. If topics need not be familiar (as suggested, but not advocated, in Gundel 1985), we can classify this example as a topic-comment cleft like the other 24 examples. As a relevant new topic, we can then perhaps conclude that it fits the requirement that the presupposition of a wh-cleft be *appropriately* in the hearer’s consciousness although not necessarily in fact already in that consciousness. In other words the hearer can anticipate that this is an appropriate way to continue the discourse. We can thus agree with Prince (1978) that the cleft clause in an it-cleft does not need to meet the condition on wh-cleft clauses, and extend this conclusion to reverse wh-clefts.

We thus agree with Prince (1978) that wh-cleft clauses but not it-cleft clauses must express given information, but disagree with her on which sense of givenness such clauses must conform to. We claim that wh-cleft clauses are always *relationally* given, i.e. topics; although we dispute that they need to be *referentially* given, i.e. activated, or ‘Chafe-given’ (Prince 1985). Thus, we believe that the distinction between referential and relational givenness as articulated in Gundel and Fretheim (2004), as well as previous works by Gundel, is crucial to the analysis of clefts.

In the following section, we further argue that wh-clefts are always topic-comment structures, unlike it-clefts and reverse wh-clefts which can be either topic-comment or comment-topic (or all comment). We use examples from both the current McLaughlin corpus and the previous one, as well as other attested and constructed examples.

#### 4. Evidence for the Relational Information Structure of Clefts

In this section we present a variety of arguments in favor of our hypothesis concerning the relational information structure of cleft sentences. To review, this hypothesis is that cleft sentences in English typically split the expression of a proposition into two parts, one part representing the topic and one part representing the comment of the utterance. The initial cleft clause in a wh-cleft always presents the topic of the utterance, with the following clefted constituent always presenting the comment; whereas it-clefts and reverse wh-clefts allow either mapping of cleft parts onto the two information structure categories. We start with three arguments based on examples from our corpus studies, follow these with two new arguments based on examples from outside the corpora and on

constructed examples, and then close with an argument from the literature that relies on linguistic intuitions associated with constructed examples.

#### 4.1 Personal pronoun clefted constituents

There were four examples of reverse wh-clefts in the data with personal pronouns as subjects. We analyzed all of these as topic-comment reverse wh-clefts. The clearest example, shown in (22), is one with an unstressed pronominal *it* as subject. It is clear that the main sentence stress falls at the end of the cleft clause in this example since the clefted constituent is completely unstressed. It is also clear that the utterance is ‘about’ countries expelling their own citizens. Furthermore, the presupposition of the cleft clause, that the U.S. did something to their Japanese citizens during World War II is familiar to the listeners but has not been mentioned in the discourse. It thus can function easier as a comment than a topic.

- (22) Mr. Barone: ... So I think that would be okay. I don't think a country should go expelling its own citizens
- ...
- Mr. Blankley: It is repugnant. However, it may be the wave of the future—not just in Israel, but perhaps in Europe, as well. As long as the West fears Islamist terrorism, there will be advocates for removing every possible suspect, the innocent along with the guilty. Whether it will ever become government policy, I don't know, but it will be argued more and more by people who are more and more fearful.
- Ms. Clift: It's what we did to the JAPANESE. [5/24/02]

Another example of a reverse wh-cleft with an unstressed pronominal subject was reported in Hedberg (1988) and is shown in (23).

- (23) Mr. McLaughlin: When will Dole exit the presidential race?
- Mr. Kondracke: ...I said he was going to be an inactive candidate, but still in the race, and not declaring that it's over, --not concede to George Bush—
- ...
- Mr. Germond: We're saying—you weren't in the real world in those days, John, but it was what Ed MUSKIE did in 1972. He suspended his campaign; remained a candidate for awhile. [McLaughlin Group, 3/11/88]

Here again primary sentence stress appears on the cleft clause, the utterance is most clearly ‘about’ candidates remaining in the race with an inactive campaign, and the presupposition that Ed Muskie did something in 1972 is familiar but discourse new.

Hedberg (1988) also reports an example of a purportedly topic-comment reverse wh-cleft with a stressed personal pronoun in subject position. This example, shown in

(24), was not videotaped, so the stress pattern was assigned purely on the basis of the author's intuition:

- (24) Mr. McLaughlin: Number two, is it not true that Nancy Reagan is always right?... This is not a bash Reagan session. I just want to know whether or not her instincts are invariably correct? I ask you.
- Mr. Kondracke: No, they're not. I mean she was the one who wanted to keep Reagan from appearing anywhere in PUBLIC. That was—
- Mr. McLaughlin: But she also has the concern about the man's health. She—he lived through an assassination attempt, remember that.  
[McLaughlin Group, 3/6/87]

This example is especially interesting because the analysis of it as having a topic-comment structure was disputed in Lambrecht (2001: 481-482). Lambrecht argues that the clefted constituent in any kind of cleft always expresses the focus. He discusses example (24) because it 'raises analytical problems.' He concedes that Mr. Kondracke's reply is 'clearly ABOUT Nancy Reagan in the sense that it is meant to answer [Mr. McLaughlin's] request for information about the First Lady' and that the aboutness relation can't be construed as going the other way around.

However, Lambrecht argues against the topic-comment analysis of (24) for two reasons. First, he disputes the import of the prosodic analysis that assigns secondary stress to the subject *she* and primary stress to *public* at the end of the cleft clause. Citing Ladd (1996), he suggests that the relative prominence of the latter accent may be due to the fact that the final accent in sentences with more than one accent tends to be the strongest. Secondly, and for him more importantly, he says, "if the speaker had intended for the [reverse wh-cleft] proposition to be interpreted as a comment about Nancy Reagan, there would have been no reason not to use the canonical version *she wanted to keep Reagan from appearing anywhere in PUBLIC*. In this version, the predicate phrase would necessarily be construed as attributing a hitherto unknown property to the subject referent."

With regard to this second criticism, we would reply that Lambrecht has a different definition of topic and focus (comment) from us. For him, topic is the complement of 'predicate focus' and is associated apparently exclusively with the subject of a simple NP + VP clausal structure. He analyzes clefts as encoding 'argument focus' structures, where narrow focus falls on a single argument of a proposition. He terms the complement of the argument focus a 'presupposition' and doesn't associate the function of 'topic' per se with presuppositions, although presuppositions can be 'topic-presupposed'. For us, however, the notion of topic is broader. Either a presupposition or the complement of a presupposition can be a topic. Most importantly, contrary to Lambrecht's assumption that the complement of a topic expresses a 'hitherto unknown property', the reason a cleft is chosen in this example is precisely that the property attributed to the topic, Nancy Reagan, is *not* new, but rather is presupposed—in this case, hearer old but discourse new (Prince 1992).

With regard to the first criticism, we dispute Lambrecht's conclusion. A final accent in a sentence with more than one accent is not necessarily the strongest one. We



perceive the final accent in comment-topic reverse wh-clefts as weaker than the initial accent on the clefted constituent, and this relative weakness may even have a basis in acoustic fact. Topic-comment reverse wh-clefts can be contrasted with this case. The difference can be seen in pitch tracks derived from the utterance of (25), which we analyze as a topic-comment reverse wh-cleft, compared to the utterance of (26), which we analyze as comment-topic reverse wh-cleft.

- (25) Mr. McLaughlin: Exit: On a probability scale of zero to 10, zero meaning zero probability, 10 meaning metaphysical certitude, rank the probability that Osama bin Laden will go nuclear. That includes two presumptions: One, that he has the bomb, and two, that he will use the bomb. Zero to 10.  
Mr. Lowry: I think it's a two. It's very likely that we'll kill him before he gets his hands on anything.  
Mr. McLaughlin: But you give it a two, not a one.  
Mr. Lowry: You can't totally rule it out. And a two is still a big threat, and we have to take the precautions.  
Mr. McLaughlin: **That's what I'm **SAYING.**** [11/9/01]
- (26) Mr. McLaughlin: All right. I want to go back to Nunn because his testimony on March the 7th was unexcelled. Nunn—no question about it—and Lugar are the most authoritative figures in this area. Listen to Nunn. Small point.  
Sam Nunn (co-chairman and CEO, Nuclear Threat Initiative): (From videotape.) In protecting America from nuclear terrorism, an ounce of prevention is worth a megaton of consequence management.  
Mr. McLaughlin: **THIS is what he's talking about.** He's talking about bringing homeland security abroad. That's where it has to go. And it has to go to Russia because Russia is the terrorist's gold mine.... [5/24/02]

The pitch track of (25) is shown in Figure 1 and the pitch track of (26) is shown in Figure 2. It can be seen that the final pitch accent in Figure 1 on the cleft clause is higher than the pitch accent on the pitch accent on the subject demonstrative pronoun, whereas the opposite pattern appears in Figure 2. This difference in relative height correlates with the perceived difference in prominence, although it is perfectly possible for a final accent which is lower than an initial accent to be perceived as the most prominent.

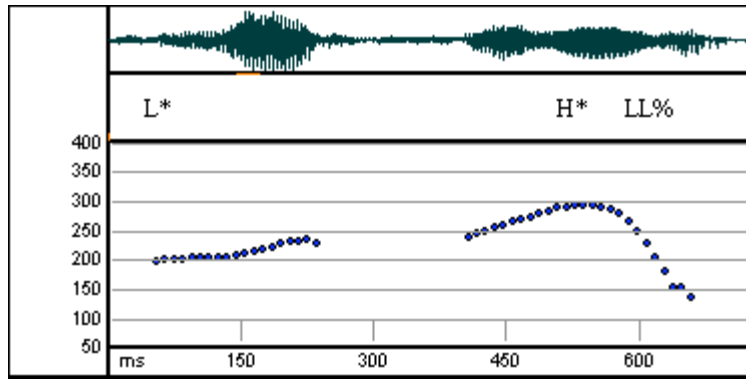


Figure 1. **That's** what I'm **SAYING**.  
L\* H\* LL%

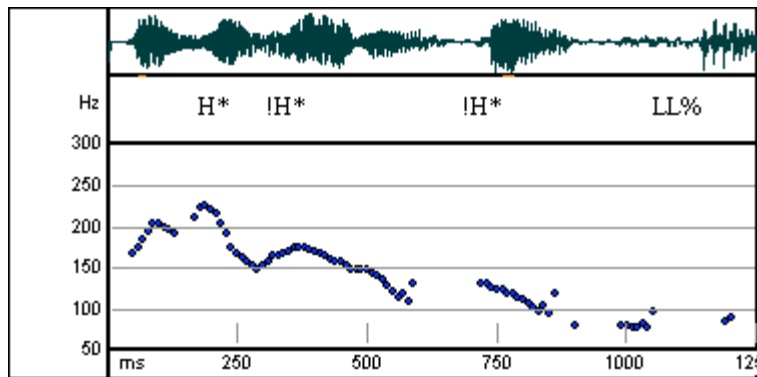


Figure 2. **THIS** is **what** he's **talking** about.  
H\* !H\* !H\* L-L%

We thus dispute Lambrecht's claim that the final pitch accent in a reverse wh-cleft utterance is always perceived as more prominent than an earlier one, and this leaves us free to posit a distinction between topic-comment and comment-topic reverse wh-clefts.

#### 4.2 Relative clause reverse wh-clefts

The relative pronoun (or deleted NP in older theories) has sometimes been argued to be the topic of a relative clause, for example by Kuno (1972), Gundel (1974), and Bresnan and Mchombo (1987). Gundel (1974) argues that relative clause extraction can take place 'asymmetrically' out of conjunction structures such as that in (24a) for the same reason that the topicalization and left dislocation can take place asymmetrically out of conjunction structures in (24b) and (24c):

- (27) a. The guitar which Jim saved \$50 and bought was a Yamaha.  
b. This guitar I've sung folksongs and accompanied myself on all my life.

- c. This guitar, I've sung folksongs and accompanied myself on it all my life.<sup>8</sup>

In all three cases 'the guitar' is the topic of both conjuncts, thus motivating an exception to the 'Coordinate Structure Constraint' of Ross (1967).

Since relative pronouns can occur as clefted constituents in reverse wh-clefts, it is consistent with the above arguments to conclude that such reverse wh-clefts are topic-comment clefts. We had eight in the McLaughlin Group data, of which two are shown in (28) and (29).

- (28) Ms. Clift: So NASA does have the difficulty of balancing the incredible cost of this, which is **why** I would **TAKE** Mr. Tito's check. Good idea. [5/4/01]
- (29) Mr. O'Donnell: He make almost no mistakes, and probably no mistakes since September 11th, which is when the **REAL** campaign, both quiet and active, was **really** going on. [11/10/01]

In both (28) and (29) the relative clause is intuitively 'about' the antecedent of the relative pronoun. And in (29), which has a noun phrase as antecedent, the 'topic tests' of Gundel (1974) verify that the period of time since September 11th is the topic of the relative clause. (The discourse is about the period of time during which the electoral race for mayor of New York was most relevant.) For example, the speaker could have stopped after his first clause in (29), and then continued as in (30a). He also could have stopped after his first clause, and then answered another participant's question in (30bi) with (30bii).

- (30) a. As for the period of time since September 11th, that is when the real campaign, both quiet and active, was really going on.
- b. i. What about the period of time since September 11th?
- ii. That is when the real campaign, both quiet and active, was really going on.

The main sentence (clause) accent appears on the cleft clause in both (28) and (29), which is what we would expect if the cleft clause is functioning as the comment in these examples.

### 4.3. Right-dislocated clefted constituents

Left- and right-dislocated constituents are widely assumed to function as topics of the utterances they appear in (Gundel 1974, 1985, 1988, Lambrecht 1994, inter alia). In our data, there is one case of an apparent right-dislocated clefted constituent in a reverse wh-cleft, shown in (31).

- (31) Mr. Zuckerman: You've got to advance the argument. That's why you've got to give me a chance to talk. So let me just say, what they

---

<sup>8</sup> This example comes from Ross (1967).

resent more than anything else is the success of America.  
 We are the success -  
 Mr. Buchanan: That's not why they're **KILLING** us, because we  
**succeeded.** [5/24/02]

Gundel (1985) reports that right-dislocated constituents express activated topics, and Lambrecht (1994) reports that right-dislocated constituents ('antitopics' in his terms) must be accessible in the discourse context and unaccented. In (31) the 'because' clause is coreferential with the subject demonstrative pronoun and is at least indirectly activated. It is not unaccented, however, but the accent is a secondary one.

Another example, from the earlier corpus, is shown in (32):

(32) Mr. Germond: The market reacted this week to—  
 Mr. Novak: Oh, it went up and it went down.  
 Mr. Germond: That's what the market **DOES**, it goes *up*, it goes *down*.  
 [11/20/87]

However, there were several examples of reverse wh-clefts which on the surface look like right-dislocation constructions, but upon examination of the context prove not to be. One such example is shown in (33).

(33) Mr. McLaughlin: Did Hastert actually hear that it was weapons-grade anthrax?  
 Mr Blankley: That's what *I'm* **TOLD**, that they got it from the **SENATE side.**  
 Mr. McLaughlin: Do you know whom he got it from?  
 Mr. Blankley: I don't know who it was, but I have to --  
 Mr. McLaughlin: What senator was putting out that story? [10/20/01]

Here, the right-attached clause is not coreferential with any proposition actually activated in the context and is not inferrable from any proposition in the context. It thus does not have the characteristics of a right-dislocated constituent, and does not function as a topic. Instead it functions as a parenthetical second comment, further elaborating on the activated proposition denoted by the demonstrative pronoun subject of the reverse wh-cleft. The utterance as a whole seems to be a syntactic blend composed of a reverse wh-cleft expressing the proposition 'I'm told that Hastert actually heard it was weapons-grade anthrax' and a wh-cleft expressing the proposition 'I'm told that they got it from the Senate side.'

It is difficult to come any conclusion about the topic-comment organization of a reverse wh-cleft like this since the cleft clause seems to be simultaneously acting like a comment relative to the previously established proposition and the topic relative to the information in the right-attached tag.

A pitch track for (33) is shown in Figure 3. The demonstrative subject and the word *Senate* in the tag are the words with the highest pitch, the latter perhaps in part because it is contrastive ('Senate' as opposed to 'House'). The word *I'm* is also contrastive, but auditorily *told* seems to be the word with primary stress.

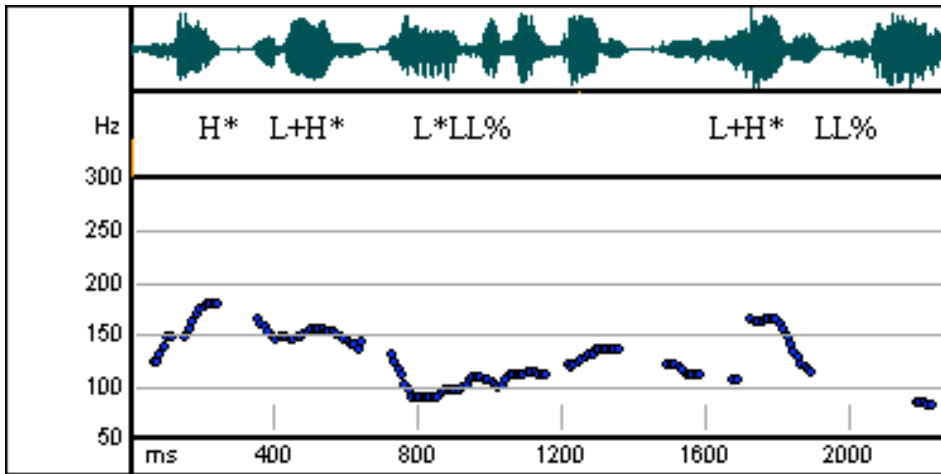


Figure 3. **That's** what *I'm* **TOLD**, that they got it from the **SENATE** side.

H\*                    L+H\*   L\*LL%                    L+H\*   LL%

The more clear right-dislocation constructions in (31) and (32), however, do support the conclusion that the initial clefted constituent, which is coreferential with the right-attached clause and is activated in the discourse context, is the topic of these reverse wh-clefts.

#### 4.4 *Also* clefts

It-clefts have long been said to be associated with a uniqueness or exhaustiveness condition on the clefted constituent, and this condition restricts the focus particles that can modify them. Thus, Horn (1969) makes the claim that ‘clefting, like *only*, specifies uniqueness, while *even* and *also* presuppose non-uniqueness and thus cannot be clefted. He gives the examples in (34).

- (34) a. It's only Muriel who voted for Hubert.  
 b. \*It's also Muriel who voted for Hubert.  
 c. \*It's even Muriel who voted for Hubert.

In arguing that an it-clefted constituent constitutes an ‘identificational focus’, É. Kiss (1998) reiterates this claim, giving the judgments shown in (35):

- (35) a. ?It was also John that Mary invited to her birthday party.  
 b. \*It was even John that Mary invited to her birthday party.

She specifies (35a) with a question mark instead of an asterisk because she identifies a context in which (35a) could be used appropriately. This context is shown in (36):

- (36) A: Bill danced with Mary.  
 B: No, it was Sam that danced with Mary.

C: It was also John that danced with her.

É. Kiss says, “B identifies Sam as the member of the set present at the party who danced with Mary, excluding the rest of the men. C adds John to the man identified by B, excluding everybody but Sam and John.” She says that “a cleft *also*-phrase appears to be acceptable precisely in a context where it can be understood to identify a member of a relevant set in addition to one or more members identified previously as such for which the predicate holds, with the rest of the set still excluded.” Note that the cleft in (36C) is a comment-topic cleft.

The claims of Horn and É. Kiss that *also*-clefts are impossible except in É. Kiss’s special context are contradicted by examples of *also*-clefts, such as those in (37) and (38):

(37) It was the President, in a rare departure from the diplomacy of caution, who initiated the successful Panama invasion. It was also Bush who came up with the ideas of having an early, informal Malta summit with Gorbachev and a second round of troop cuts in Europe after the fall of the Berlin wall. But it was Baker who subtly turned the Malta summit from the informal, ‘putting our fee up’ chat initially envisaged by the President into a platform for the United States to demonstrate through a 16-point initiative that it was prepared to help Gorbachev. [M. Dowd and T.L. Friedman, ‘The Fabulous Bush and Baker Boys, The New York Times Magazine, 5/6/90, p. 64].

(38) Rough location work is nothing new for Sheen. When he was young, the family travelled to location with his father, actor Martin Sheen. They spent 16 months in the Philippines.... Ten years later, Charlie Sheen found himself back in the Philippines...as the star of... Platoon.... The terrain and environmental elements...are very similar to Vietnam, including 120 degree heat by 8 a.m., blood-thirsty bugs and an impenetrable jungle....

It was also location work that gave Sheen his first acting break. He was nine and his dad was filming The Execution of Private Slovik.... [Jane Ammerson, ‘Intensity fuels Charlie Sheen’s On-screen Presence’, Northwest Airlines Magazine, July 1990, p. 69]

It-clefts such as these were analyzed in Hedberg (1990) as topic-comment clefts since it is clear that the main sentence stress would fall on the cleft clause (on *acting*) if they were spoken aloud, and additional information ‘about’ the activated clefted constituent is added by the cleft clause. The scope of *also* seems to be wider than just the clefted constituent in these examples, e.g. applying to *it was Bush* instead of just *Bush* in (37). In support of this hypothesis, the position of *also* can be changed to sentence-initial position with a felicitous result.

The it-cleft in (38) can be paraphrased by a reverse wh-cleft as shown in (39a), but not by a wh-cleft without awkwardness, as shown in (39b). This awkwardness seems to be the same as the awkwardness associated with a wh-cleft with a distal demonstrative clefted constituent. In both (39a) and (39b), primary stress must be placed on *acting*.

(39) a. Location work was *also* what gave Sheen his first **ACTING** break.

- b. ??What gave Sheen his first **ACTING** break was *also* location work.

If primary stress is placed on the clefted constituent, an it-cleft or reverse wh-cleft can only be used in É Kiss's special context, and in the wh-cleft ungrammaticality results, or at least a semantic clash between the exhaustivity contributed by the cleft construction and the non-exhaustivity contributed by the additive particle *also*:

- (40) a. ?It was *also* **LOCATION** work that gave Sheen his first acting break.  
 b. ?**LOCATION** work was *also* what gave Sheen his first acting break.  
 c. \*What gave Sheen his first acting break was *also* **LOCATION** work.

Thus, we can conclude that clefted constituents can be modified by *also* only if the clefted constituent represents the topic of the utterance, and that it-clefts and reverse wh-clefts can readily be used in this way but wh-clefts are more resistant.<sup>9</sup>

#### 4.5 Vice-versa clefts

It-clefts and reverse wh-clefts but not wh-clefts can also be used in a type of cleft first noticed in Ball and Prince 1978, which they termed 'vice-versa clefts'. Two examples of vice-versa it-clefts from Hedberg 1990 are shown in (41) and (42):

- (41) Anna: So, what's the case you're working on?  
 Robert: Nothing I need bother you with now. It's *you* who called **ME**,  
 remember? [General Hospital, ABC, 6/21/89]
- (42) The photon gives up part of its energy to the electron, and the transaction is observed as a slight decrease in the frequency (or increase in the wavelength) of the radiation. Inverse Compton scattering is observed when a photon encounters a high-energy electron. Then it is the electron that loses energy to the photon. [*Scientific American*, 8/77: 38, Borkin 1984]

Ball and Prince point out that such it-clefts constitute exceptions to the requirement that the cleft clause be a presupposition. Instead these seem to be more like double focus clefts, where the presupposition is, respectively, 'someone called someone' and 'something loses energy to something'.

Ball and Prince point out that the presupposition of a wh-cleft cannot be suspended in this way: See for example, (43):

---

<sup>9</sup> Topic-comment it-clefted constituents and reverse wh-clefted constituents can also be modified by *even* with primary stress on the cleft clause (and an appropriate context). Here, the contrast with wh-clefts is even more pronounced:

- (i) It was even location work that gave Sheen his first **ACTING** break.  
 (ii) Location work was even what gave Sheen his first **ACTING** break.  
 (iii) \*What gave Sheen his first **ACTING** break was even location work.

- (43) Anna: So, what's the case you're working on?  
 Robert: Nothing I need bother you with now. #The one who called me is you.

No matter what stress pattern is placed on the wh-cleft, a vice-versa reading is impossible.

However, vice-versa reverse wh-clefts do exist, as shown in (44) and (45):

- (44) "She was such a wonderful person," the woman continued. "I know I wouldn't be here today if it weren't for your mother, and everything she did for me. I just can't believe she's gone...."

A great lady? A wonderful person? Who on earth was this woman talking about? Bonnie looked toward Rod, who was staring at the woman with bemused detachment.

Lauren stood up, drew the woman into a close embrace.

"I'm the one who should be comforting you," the woman told her, pulling back, wiping stubborn tears from her eyes.

"I'll be all right," Lauren assured her.

[Joy Fielding, *Don't Cry Now*, p. 84, Random House Canada, 1995]

- (45) "...And face it, it doesn't look as if my life will be normal any time soon."

"Doesn't look like it," he agreed. "What's your point?"

"My point is that unless you have a taste for the bizarre situation I'm not sure you want to continue this walk with me."

"Well," he drawled "I guess we're even now."

"Oh, how so?"

"This time you're the one who's ambushed me."

The phrase vividly called up a picture of him in the doorway of his bedroom; Laura could practically hear her own moans.

"I didn't intend to ambush you," she said, pressing her lips together. "I just want you to know that there's no obligation to get to know me well enough to tell me dirty jokes.

[Antoinette Stockenberg, *A Month At the Shore*, p. 222, St. Martin's Press, 2003.]

That primary stress falls on the cleft clause in such examples can be seen by the fact that the author italicized the focus in the cleft clause in (45). We suggest that vice versa it-clefts and reverse wh-clefts are actually topic-comment clefts with a contrastive topic expressed in the clefted constituent, which is therefore highly stressed and a contrastive focus expressed in the cleft clause, which therefore receives the primary sentence stress. Vice-versa wh-clefts are not possible because the cleft clause is necessarily the topic in that construction.

#### 4.6. Anti-reconstruction effects



A final argument distinguishing it-clefts and reverse wh-clefts from wh-clefts derives from Heycock and Kroch (2002). They applied the argument to wh-clefts and reverse wh-clefts; we extend the argument to it-clefts. First of all, note that in simple sentences such as those in (46), a full noun phrase (R-expression) cannot be co-referential with a c-commanding pronoun. This is an illustration of Condition C of the Binding Theory of Chomsky (1981).

- (46) a. He<sub>\*i/j</sub> was proud of John<sub>i</sub>.  
 b. He<sub>\*i/j</sub> will scold John<sub>i</sub>'s children.  
 c. He<sub>\*i/j</sub> really missed John<sub>i</sub>'s dog.  
 d. He<sub>\*i/j</sub> had always claimed that John<sub>i</sub> was innocent..

Heycock and Kroch note that wh-clefts exhibit 'reconstruction' effects with respect to Condition C. That is, while the full noun phrase is not c-commanded by the pronoun in the wh-clefts on the surface in (47), it still cannot be interpreted as coreferential with it. The pronoun thus behaves as though it were 'reconstructed' into the position it would inhabit in the unclefted sentence.

- (47) a. What he<sub>\*i/j</sub> was was proud of John<sub>i</sub>.  
 b. What he<sub>\*i/j</sub> will never do is scold John<sub>i</sub>'s children.  
 c. What he<sub>\*i/j</sub> really missed was John<sub>i</sub>'s dog.  
 d. What he<sub>\*i/j</sub> had always claimed was that John<sub>i</sub> was innocent.

This phenomenon is an instance of 'connectivity' effects that have been observed for wh-clefts sentences as well as other kinds of clefts and identificational sentences ever since Higgins (1979) (for discussion and further examples, see Heycock and Kroch 2002).

What Heycock and Kroch (2002) point out, however, is that the connectivity effects with respect to condition C in reverse wh-clefts differ from those in wh-clefts. In (48c) and (48d), the pronoun can be coreferential with the full noun phrase.<sup>10</sup>

- (48) a. Proud of John<sub>i</sub> was what he<sub>\*i/j</sub> was.  
 b. Scold John<sub>i</sub>'s children is what he<sub>\*i/j</sub> will never do.  
 c. John<sub>i</sub>'s dog was what he<sub>i/j</sub> really missed.  
 d. That John<sub>i</sub> was innocent was what he<sub>i/j</sub> had always claimed.

In (48c,d), the full noun phrase is syntactically an accessible antecedent for the pronoun in the surface structure, and reconstruction does not appear to have taken place.

Heycock and Kroch point out that the pattern of results in (48) is duplicated by the pattern of results in the verb-complement preposed constructions shown in (49). It is possible to construe the pronouns in (49c,d) and not (49a,b) as coreferential with the preceding full noun phrases.

---

<sup>10</sup> Heycock and Kroch describe the antireconstruction effects as follows: "R-expressions buried inside other referential expressions or inside CPs can be coreferential with a pronoun c-commanding the 'gap'." (p. 159). (48a) and (48b) don't meet this description, as the R expression is inside an AP or a VP in these examples.

- (49) a. Proud of John<sub>i</sub>, he<sub>\*i/j</sub> was.  
 b. Scold John<sub>i</sub>'s children he<sub>\*i/j</sub> never will.  
 c. John<sub>i</sub>'s dog, he<sub>i/j</sub> really missed.  
 d. That John<sub>i</sub> was innocent, he<sub>i/j</sub> had always claimed.

Heycock and Kroch claim that information structure can explain these results. In the preposing and reverse wh-cleft examples, the initial constituent can function as the topic of the utterance, but in the wh-cleft the clefted constituent is necessarily the focus.

Heycock and Kroch point out that preposed arguments can function as preposed topics (c.f. the 'topicalization' of Prince 1986 and the 'topic topicalization' of Gundel 1974) or as preposed foci (c.f. the 'focus preposing' of Prince 1986 and the 'focus topicalization' of Gundel 1974). They claim that the anti-reconstruction effects arise only when the preposed constituent can be construed as the topic.

This argument is persuasive and has a functional basis in that it is reasonable to assume that a pronoun in a focus expression can have as its antecedent a full noun phrase in the associated topic expressions. Topics by definition constitute the element about which the focus serves as the pragmatic predication, and thus are 'prior' to it in relational information structure and should be able to serve as antecedents. Conversely, we wouldn't expect to find pronouns in the topic depending for their reference on full noun phrases in the focus.

Since we are exploring the information structure of it-clefts as well as wh-clefts and reverse wh-clefts in this paper, it is important for us to point out that the pattern of results found by Heycock and Kroch in reverse wh-clefts is also found in it-clefts. Here, however, the inability of the (a, b) examples to occur is independently explained by the inability of adjective phrases and verb phrases to function as it-clefted constituents in most dialects of English (see Delin 1989, Hedberg 1990 for discussion). But if we examine the counterparts of (48c, d) and (49c, d) we see that once again they allow coreference, i.e. exhibit anti-reconstruction effects.

- (50) a. \*It was proud of John<sub>i</sub> that he<sub>i/j</sub> was.  
 b. \*It was scold John<sub>i</sub>'s children that he<sub>i/j</sub> never will.  
 c. It was John<sub>i</sub>'s dog that he<sub>i/j</sub> really missed.  
 d. It was (the fact) that John<sub>i</sub> was innocent that he<sub>i/j</sub> had always claimed.

We can conclude, then, that based on anti-reconstruction effects the clefted constituent in reverse wh-clefts and it-clefts but not wh-clefts can function as the topic of the utterance. As mentioned above, Heycock and Kroch (2002) also conclude that the clefted constituent in most reverse wh-clefts functions as the topic. We support this conclusion in our corpus study.

## 5. Conclusion

In this paper we have argued that wh-clefts are associated with a single relational givenness structure—topic-comment, while reverse wh-clefts and it-clefts are associated

with two relational givenness structures—topic-comment or comment-topic (or sometimes, all-comment). We first examined a corpus of naturally occurring examples of all three types of clefts and analysed the referential and relational givenness status of the two major parts, finding that these examples support our hypothesis. We use prosody as the primary cue to relational givenness structure, claiming that primary stress falls on the relationally new comment, even though the comment may present activated (i.e. referentially given) material.

We then presented six arguments in support of our proposal, drawing on data from the corpus, an earlier corpus, as well as other naturally occurring examples and constructed examples. We looked at reverse wh-clefts with personal pronominal clefted constituents, relative clause reverse-wh clefts, right-dislocated reverse wh-clefted constituents, reverse wh-clefts and it-clefted constituents modified by *also* or *even*, vice-versa reverse-wh clefts and it-clefts, and at anti-reconstruction effects in reverse-wh-clefts and it-clefts. In all cases, we found that wh-cleft cleft clauses behave like topics but reverse wh-cleft or it-cleft cleft clauses behave either as topics or comments.

We leave it up to syntacticians and semanticists to explain why wh-cleft clauses are necessarily topics while reverse wh-cleft and it-cleft clauses can be either topics or comments. Perhaps it is because the wh-cleft subject as a sentential subject is preposed into a topic position; whereas the subjects of the other two types of clefts are in a purely subject position in spec-IP, whether base-generated there or moved from spec-VP (to speak in terms of Principles and Parameters theory). Whatever the derivation, we simply claim that the two classes of clefts are different in terms of their pragmatic interpretation.

## 6. References

- Ball, Catherine N. and Ellen Prince. 1978. A note on stress and presupposition. *Linguistic Inquiry* 8. 585.
- Beckman, Mary and Gayle Ayers Elam. Guidelines for ToBI Labeling. Version 3. Ohio State University.
- Borkin, Ann. 1984. *Problems in Form and Function*. Norwood, NJ: Ablex Publishing Corporation.
- Bresnan, Joan and Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chichewa. *Language* 63: 741-782.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Foris: Dordrecht.
- Delin, J. L. 1989. *Cleft Constructions in Discourse*. Ph.D. dissertation, University of Edinburgh.
- É Kiss, Katalin. 1998. Identificational focus versus informational focus. *Language* 74. 245-273.
- Gundel, Jeanette K. 1974. *The role of topic and comment in linguistic theory*. Ph.D. dissertation, University of Texas at Austin. Re-issued 1988, in Jorge Hankamer, ed. *Outstanding Dissertations in Linguistics*, Garland Publishing Co.
- Gundel, Jeanette K. 1985. Shared knowledge and topicality. *Journal of Pragmatics*, 9: 83-107.
- Gundel, Jeanette K. 1988. Universals of topic-comment structure In M. Hammond,

- E. Moravcsik and J. Wirth (eds.), *Studies in syntactic typology*, 209-239, Amsterdam:John Benjamins, 209-239.
- Gundel, Jeanette K. 1999b. Topic, focus and the grammar pragmatics interface. In J. Alexander, N. Han and M. Minnick (eds.), *Proceedings of the 23rd Annual Penn Linguistics Colloquium*. Penn Working Papers in Linguistics, vol. 6.1, pp. 185-200.
- Gundel, Jeanette K. and Thorstein Fretheim. 2004. Topic and focus. In Laurence R. Horn and Gregory Ward, eds., *The Handbook of Pragmatics*. London:Blackwell.175-196.
- Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski. 2001. Definite descriptions and cognitive status in English: Why accommodation is unnecessary. *English Language and Linguistics* 5: 273-295.
- Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69:274-307.
- Halliday, M. A. K. 1970. Language structure and language function. *New Horizons in Linguistics*, John Lyons (ed.), 140-164. Harmondsworth, England: Penguin.
- Hedberg, Nancy. 2000. The referential status of clefts. *Language* 76: 891-920.
- Hedberg, Nancy 1990. *Discourse pragmatics and cleft sentences in English*. Minneapolis, MN: University of Minnesota dissertation
- Hedberg, Nancy. 1988. The Discourse Function of Cleft Sentences in Spoken English. Linguistic Society of America Meeting, Dec. 1988, New Orleans, Louisiana.
- Heycock, Caroline and Anthony Kroch. 2002. Topic, Focus, and Syntactic Representations. *WCCFL 21 Proceedings*, ed. L. Mikkelsen and C. Potts, pp. 141-165. Somerville, MA: Cascadilla Press.
- Higgins, F. Roger. 1979. *The Pseudo-Cleft Construction in English*. New York: Garland.
- Kuno, Susumo. 1972. *The Structure of the Japanese Language*. The MIT Press.
- Horn, Laurence R. 1969. A presuppositional analysis of *only* and *even*. *Proceedings of the Chicago Linguistics Society* 5. 98-108.
- Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, focus, and the mental representations of discourse referents*. Cambridge University Press.
- Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39. 463-516.
- Prince, Ellen F. 1978. A comparison of wh-clefts and it-clefts in discourse. *Language* 54. 883-906.
- Prince, Ellen F. 1985. Fancy syntax and shared knowledge. *Journal of Pragmatics* 9.1:65-81.
- Prince, Ellen F. 1986. On the syntactic marking of presupposed open propositions. *Papers from the Parasession on Pragmatics and Grammatical Theory*, ed. by A. M. Farley, P. T. Farley, and K.-E McCullough. Chicago: Chicago Linguistic Society. 208-222.
- Prince, Ellen F. 1992. The ZPG letter: Subjects, definiteness and information status. In Thompson, S and Mann, W. *Discourse descriptions: Diverse analyses of a fund-raising text*. Philadelphia/Amsterdam: John Benjamins. 295-325.
- Reinhart, Tanya 1981. Pragmatics and linguistics. An analysis of sentence topics. *Philosophica* 27:53-94.

- Ross, John R. 1967. *Constraints on Variables in Syntax*. Ph.D. Dissertation, Massachusetts Institute of Technology.
- Vallduví, Enric and Maria Vilkuna. 1998. On rheme and kontrast. In Culicover, Peter and Louise McNally (eds.), *The Limits of Syntax*. Syntax and Semantics 29. New York: Academic Press, 79-108.