Directly and Indirectly Anaphoric Demonstrative and Personal Pronouns in Newspaper Articles

Nancy Hedberg

Simon Fraser University
Vancouver, BC Canada
hedberg@sfu.ca

Jeanette K. Gundel

University of Minnesota
Minneapolis, Minnesota USA
gunde003@tc.umn.edu

Ron Zacharski

New Mexico State University
Las Cruces, New Mexico USA
ron.zacharski@gmail.com

Abstract

This paper reports on a study of pronouns *this*, *that*, and *it* in articles in the New York Times, testing the following hypotheses: (1) the pronoun *it* requires its referent to be in the addressee's focus of attention; demonstrative pronouns only require activation; (2) the anaphoric relation between *it* and its antecedent tends to be direct (co-referential); the relation between a demonstrative pronoun and its antecedent tends to be indirect (non-coreferential). A pilot study of 33 *it* found that 81.8% agreed-upon tokens were 'in-focus direct'. A study of 181 *that* and 140 *this* pronominals resulted in annotator agreement of 79.8% for in focus vs. activated and direct vs. indirect anaphora coding. Within the agreed-upon data, 69.5% of demonstrative pronouns were 'activated indirect'. These results support the primary hypotheses as well as the proposed explanation for why demonstrative pronouns tend to have non-nominal antecedents. Non-nominal constituents describe/denote eventualities and activities. However, pronouns with non-nominal antecedents most often refer to facts, situations, etc. that are only indirectly related to such eventualities. The referents of such pronouns are thus activated, but not in focus, and they therefore require a demonstrative, since the pronoun *it* can only access an in focus referent.

I. Introduction

According to Gundel, Hedberg and Zacharski (1993), unstressed personal pronouns in English can only be used appropriately when the referent is in the addressee's focus of attention, while demonstrative pronoun referents need only be activated. Referents can be brought into focus when a sufficiently salient contextual object or NP/DP introduces them. Gundel, Hegarty and Borthen (2003) propose that events, states or activities directly introduced by clauses and verb phrases can also be brought into focus and thus are able to be referred to by either personal pronouns or demonstratives; however, propositions, speech acts, situations and facts introduced by non-nominal expressions are typically accessible only to reference by demonstratives, since their referents are activated, but not in focus.

Two examples supporting such an analysis are shown in (1)-(2):

(1) a.   'We believe her, the court does not, and **that** resolves the matter,' Mr. Montanarelli said today of Ms. Lewinsky's testimony that… (NY Times, 5/24/00)
     b.   'We believe her, the court does not, and **it** resolves the matter,' Mr. Montanarelli said today of Ms. Lewinsky's testimony that…

(2) A: I just ate three pieces of cake.
     B: Can your repeat **that**?
     B': Can you repeat **it**?

In (1a), *that* is interpreted as referring to the complex fact that a group including Mr. Montanarelli believes her and the court does not, which can be inferred from the state of affairs described by the first two clauses; *it*, on the other hand, would be taken as referring to the individual entity, the court, directly introduced in subject position in the previous clause. In (2B), *that* can refer either to A's speech act of stating that s/he just ate three pieces of cake or to the activity directly denoted by *ate three pieces of cake*; but *it* in (2B') can only have the latter interpretation.

It is proposed that the difference in cognitive status signaled by demonstrative and personal pronouns can account for the findings of Webber (1991), Hegarty, Gundel and Borthen (2001), Byron (2002), among others, that most entities introduced by non-nominal constituents are referred to subsequently by demonstrative pronouns and only a relatively small percentage are referred to by personal pronouns. Eckert and Strube (2000) further discuss the use of demonstrative pronouns as opposed to personal pronouns for 'discourse deictic' reference, whose resolution, in their view requires 'referent coercion', i.e. an inferential process, upon encountering the anaphor.

Gundel, Hedberg and Zacharski (2002/2005) further test the hypothesis explaining the differential distribution of personal and demonstrative pronouns, based on a study of the Santa Barbara Corpus of American English (Dubois et al. 2000), which contains transcripts of informal, casual speech. They found that 1,676 of 2,006 (83.6%) personal pronouns had NP/DP antecedents, while only 168 of 601 (28%) of demonstrative pronouns had NP/DP antecedents, thus corroborating results of previous studies and indirectly supporting the hypothesis that

personal pronouns often cannot access referents associated with non-nominal constituents because they are merely activated, but not in focus.

Gundel, Hedberg and Zacharski (2004) focused on demonstrative pronouns in the same corpus, explicitly testing the hypothesis that demonstrative pronouns can be used in referring to entities that do not have direct (i.e. co-referential) NP/DP antecedents, including those entities which are only indirectly introduced by non-nominal constituents, and thus require additional processing (e.g. situations, facts, speech acts and propositions), whereas such entities are inaccessible to reference with a personal pronoun because they are not in focus. In order to verify that the coding was reliable, each of the three authors separately coded 99 demonstrative pronouns in the corpus. The coding scheme distinguished between NP/DP antecedent versus non-NP/DP antecedent material and between direct versus indirect relationship of the anaphor to the antecedent or antecedent trigger (Cornish 1999). It was found that all three coders agreed on the coding in only 56 cases (56.5%) However, within the 56 agreed-upon tokens, 24 (42.9%) had indirect non-NP/DP antecedents, thus weakly supporting the basic hypothesis.

Upon analyzing the tokens that were disagreed upon more closely, it was determined that some examples were ambiguous (cf. Arstein and Poesio 2005), leading to coders disagreeing on the referent of the pronoun, and thereby also on whether the antecedent (trigger)-anaphor relation was direct or indirect. It was noted that such problems in reliable understanding were perhaps partly due to the difficulty of analyzing casual unplanned speech, which is subject to a greater amount of vagueness and ambiguity of reference than other genres.

For example, (3) is ambiguous as to whether the *these* refers to the beans in the extralinguistic context, and thus has no NP/DP antecedent, or anaphorically to the beans that were previously introduced linguistically, and thus has an NP/DP antecedent:

(3)   MARILYN: Would [2you like to2] … string the beans?
        PETE: [2What can I do2]. Sure.
        [Intervening Material]
        PETE: **These** aren't particularly stringy.

In example (4), it isn't clear whether the referent of *that* is a basil plant or basil leaves. If the former, the reference is direct, as it is coreferential with *it* in Sam's utterance and *basil* in Angela's utterance, but if the latter, it is indirect.

(4)   ANGELA: … (H) Can I grow some basil? … From seed?
        SAM: … Yes, that's how I've- -- I don't have any this year, but I've grown it other years.

[Intervening Material]
DORIS: … Isn't **that** what you gave the neighbor one time? … You gave him some kind of herb.

In light of this difficulty with unplanned speech, the current project investigates planned discourse in order to try to improve coding reliability while at the same time testing our primary hypothesis that entities directly introduced (whether by NP/DP, VP or S/CP) may be brought into focus and are thus accessible with a personal pronoun, while entities not directly introduced (including propositions, facts, situations and speech acts associated with eventualities directly introduced by non-nominal constituents) are only activated and thus can be accessed with a demonstrative pronoun, but not with a personal pronoun. In this paper, we report on a pilot study and a full study of demonstrative pronoun use in a corpus of newspaper articles from the New York Times, and a pilot study of the use of *it* in the same corpus.

2. Pilot Study on *That* and *This*

As a pilot study, we coded a single issue of the New York Times, from January 1, 2000, so that we could compare our three-way initial coding. The study concentrated on singular neuter pronouns because they can in general be substituted for each other, thus making it possible to clearly examine the factors that go into choosing among this set of pronouns. We tagged the first 50 *that* pronominals and then the *this* and *it* pronominals that occurred in that span of text. In this way, we found 50 occurrences of *that*, 53 occurrences of *this,* and 234 occurrences of *it*. Each occurrence of one of these pronouns was coded in an XML-compatible format as in (5), so that it could be extracted from the text-file automatically and its properties could be coded.

(5)   **<it num="3">It</it>** 's the basics: shoot a high percentage, rebound well and …..

As the focus of the pilot study was to develop the coding scheme, we analyzed the referential details only of demonstrative pronouns. Pronouns within quotations were excluded from the study because it was unclear whether they would relate back to previous discourse in the same way as non-quoted pronouns. Pronouns that were 'nonreferential' or 'cataphoric' were also excluded. In the end, there were 27 *that* pronominals which all three of the coders agreed on as fitting the criteria for analysis, and 16 *this* pronominals, 43 tokens in total. The referent of the pronoun was indicated as a check on whether coders interpreted the pronoun in the same way. The linguistic material that constituted the antecedent was also annotated. Each demonstrative pronoun was coded for

highest cognitive status of the referent (e.g. 'in focus' or 'activated'), following the coding guidelines of Gundel (2004). Although the guidelines only include sufficient criteria for a given status, not necessary ones, coding was limited to these guidelines in order to ensure greater reliability. The anaphoric relation was coded as 'direct' or 'indirect', depending on whether or not the pronoun was co-referential with its 'antecedent'. The relation was counted as direct, for example, if a pronoun referred to an individual referent of a previous NP/DP or to an event described by a previous clause. The relation was counted as indirect in the case of 'inferrables' such as *the bus… .the driver*, or if it referred to a fact inferred from an event or state of affairs described by a previous clause.

For example, in (6), the referent of *that* is the number 1,350. The highest cognitive status of the referent is 'activated', as the number was introduced in the previous sentence, but not in a syntactically prominent position. It is part of the subject of the previous sentence but the subject itself is *1,350 commercial jets*. The anaphoric relation is coded as 'direct', because *that* is coreferential with *1,350.*

(6)   <P num="52"> Worldwide, about <1>1,350</1> commercial jets were in the air at the time, according to the Air Transport Association, the airline industry's trade group. **<that ACTIVATED DIRECT "the number 1,350" 52.1 num="1">That</that>** was less than half the number of planes that flew last New Year's Eve.</P>

In (7), the referent of *this* is the fact that Giuliani was sleepy. Its highest cognitive status is 'activated' as it was not directly referred to in the previous sentence. The anaphoric relation is coded as indirect because the clause *I was sleepy* directly introduces the state of Giuliani being sleepy, but *this* refers to the fact that must be inferred from his assertion of being in that state.

(7)   <P num="405"> NEW YORK _ Mayor Rudolph Giuliani, who gave himself the job of ubiquitous master of ceremonies of the city's New Year celebration, said he began his last day of the 1900s at 5:30 a.m. having trouble getting his lights on. ``I was convinced that it was Y2K,'' the mayor said, but <1>``actually I was sleepy."</1></P>
        <P num="406"> **<this ACTIVATED INDIRECT "the fact that Giuliani was sleepy" 405.1 num="12">This</this>** perhaps explains an interesting mishap...

In our pilot study, 63% of *that* pronouns and 69% of *this* pronouns were agreed upon by all three coders with

regard to both cognitive status and direct/indirect relation of anaphor to antecedent (trigger), a total of 65% overall agreement. This agreement is an improvement over the result obtained in the casual speech corpus (56.5%)

Of the 28 demonstratives that all coders agreed on, 23 (82.1%) were 'activated indirect', while 5 (17.9%) were 'activated direct', thus supporting our hypothesis. None of the agreed-upon tokens was coded as 'in focus'.

In the next sections, we report on a fuller study of demonstrative pronouns and a pilot study of personal pronoun *it*. Our goal was to test the primary hypotheses that *it* requires the referent to be in focus, whereas *this* and *that* only require activation and that the anaphoric relation between a pronoun and its antecedent tends to be direct for the personal pronoun *it*, but indirect for demonstratives *this* and *that*.

3.   Full Study on *That* and *This*

3.1. Method and Results

For the full study, we analyzed 321 demonstratives in two issues of the New York Times (January 2 and 3, 2000), 65 instances of *that* and 35 instances of *this* in the Jan. 2 issue (henceforth 'Transcript 2') and 116 *that* and 105 *this* in the Jan. 3 issue (henceforth 'Transcript 3'). As in the pilot study, we excluded pronouns in quotations, as well as nonreferential and cataphoric pronouns.

The coding scheme is shown in (8).

(8)   <FORM, COGSTATUS, TYPE, REF, REF_INDEX, CRITERION, NUM>

FORM was *that* or *this*. COGSTATUS was INFOCUS or ACTIVATED (none of the tokens was coded with a lower status, though that was not excluded as a possibility), TYPE was DIRECT or INDIRECT, REF was a description of the referent of the pronoun, REF_INDEX was a pointer to the marked place in the text where the antecedent material was located, CRITERION was an index to the criterion for 'in focus' or 'activated' status in the coding protocol for statuses on the Givenness Hierarchy (Gundel 2004), and NUM was the numerical index of the form in the particular transcript.

Two coders annotated the transcripts. The results were automatically extracted and placed on a spreadsheet, where they could be compared and tabulated. We calculated agreement statistics on COGSTATUS and TYPE. Agreement was 80% on Transcript 1 and 79.6% on Transcript 2, where both factors have to match in order for the token to be counted as agreed upon.

Next, we combined the results on the two transcripts and calculated a kappa statistic for COGSTAT and TYPE separately. For cognitive status, expected agreement was .80 and observed agreement was .89, which yielded a

kappa score of .46, a level often taken to indicate 'moderate agreement.' For direct/indirect type, expected agreement was .52 and observed agreement was .85, which yielded a kappa score of .70, a level often taken to indicate 'substantial agreement'.

The reason the kappa score for COGSTAT is relatively low is that 83% of the time both coders said 'activated', so chance agreement on 'activated' is high. Only 6% of the time did both coders say 'in focus', and 12% of the time the two coders disagreed. Thus there was substantial disagreement on coding tokens as 'in focus'.

Table 1 shows the combined results for both transcripts. 70% of agreed-upon demonstrative pronouns were coded indirect in Transcript 2, and 69.3% were found to be indirect in Transcript 3. The results for percent indirect were remarkably consistent across transcripts. Thus, *that* was 78% and 77.3% indirect in

	That	This	Total
Activated			
Direct	32	35	67
Indirect	114	64	178
In Focus			
Direct	1	10	11
Indirect	0	0	0
Total	147	109	256
% Indirect	77.5%	58.7%	69.5%

Table 1: Cognitive Status and Type in Agreed-upon Tokens in Transcripts 2 and 3

Transcripts 2 and 3, respectively; while *this* was 56.7% and 59.4% indirect, respectively. This uniformity perhaps tells us that we have analyzed enough tokens to have uncovered a regularity in the distribution of demonstrative pronouns in this genre.

To summarize, the results support both hypotheses concerning demonstrative pronouns: *this* and *that* only require activation, and the anaphoric relation between a demonstrative pronoun and its antecedent tends to be indirect. We turn to a discussion of anaphora type and cognitive status in the data in the next two subsections.

3.2. Discussion: Direct

Sentence-level referents that were coded INFOCUS conformed to criterion IF6: "It is the event denoted by the immediately preceding sentence." We took states and activities denoted by verb phrases and states (of affairs) denoted by sentences to also fall under this criterion. Such tokens were also coded DIRECT. Only a small number of demonstrative pronouns were in this category (6 total), but a clear example is shown in (9):

(9)   The winner was Internet Capital Group, a company that invests in other Internet companies. <1> It more

than doubled its first day of trading, Aug. 5. </1> and **<that INFOCUS DIRECT "the event of the stock doubling on its first day of trading" 929.1 IF6 num="14">that</that>** was just the beginning. </P>

Events were not always coded INFOCUS, however. There were seven examples like example (10), which was coded as ACTIVATED because the event reference is not made by the main clause:

(10)   <P num="8879"> But Holmes pummeled Cooney, prompting <1>Valle to jump into the ring in Round 13 to get the fight stopped.</1> **<that ACTIVATED DIRECT "the event of Valle jumping into the ring in Round 13 to get the fight stopped" 8879.1 A1 num="126">That</that>** was in character with Valle's credo that a trainer must demand respect but also ``show the fighter that he cares _ that's not a machine out there." </P>

When the antecedent of a pronoun referring to a fact or proposition directly introduces a fact or proposition, the referent of the pronoun is coded as DIRECT rather than INDIRECT since the pronoun and its antecedent are coreferential, as in example (11).

(11)   <P num="1128"> The confusion, say Encarta specialists (perhaps assuming that a classical education is more ubiquitous than it is), relates to <1>the fact that although millennium and millennial have two n's, millenarian and millenary have only one.</1> **<this ACTIVATED DIRECT "the fact that millennium and millennial have two n's, while millenarian and millenary have only one" 1120.1 A1 num="09">This</this>** is due to the words' different Latin origins. </P>

This was coded as ACTIVATED because it didn't fall under any of the criteria for 'in focus' status, for example criterion IF1, "It is mentioned in main clause subject position in the immediately preceding sentence/clause", or IF2, "It is mentioned earlier in the same sentence". Instead it fell under criterion A1, "It is mentioned in one of immediately preceding two sentences".

3.3. Discussion: Indirect

ACTIVATED INDIRECT was the largest class of demonstrative tokens, and the cognitive status criterion was A3, which reads as follows: "It is a proposition, fact, or speech act associated with the eventuality (event or state) denoted by the immediately preceding sentence(s)." We also included 'situations' under this category.

An agreed-upon example of a fact was shown in (7), and an example of a proposition is shown in (12).

(12) <P num="1500"> But the same aid official, who spends much of his time on emergency work in Pristina, admitted to me that <1>``every single Serbian here has been intimidated _ verbally in the street, on the telephone, physically."</1> </P>

<P num="1501"> As if to prove **<this ACTIVATED INDIRECT "the proposition that every Serbian has been intimidated" 1500.1 A3 num="16">this</this>**, I was confronted a few hours later by a 64-year-old Serb woman, Milunka Josic, who had just spent the night trying to keep Albanian youths from breaking down her front door. Her right hand was covered in bruises. </P>

By 'situation' we intend what Hegarty (in press) calls a 'situation with manifestations'; that is, "a complex of events or states consisting of a causally basic event or state, together with the ramifications of the basic event or state in the prevailing circumstances" (p. 171). An example is shown in (13):

(13) <P num="405"> With the exception of Japanese equities, <1>Americans have been selling more foreign stocks than they have been buying in recent months.</1> </P>

<P num="406"> But **<that ACTIVATED INDIRECT "the situation that Americans have been selling foreign stocks more than buying them" 405.1 A3 num="06">that</that>** could change.

It proved to be quite difficult to agree on how to classify sentence-level referents. Table 2 shows the agreement data on a combination of COGSTAT and TYPE. In 45 out of 74 instances of disagreement (or 61%), the coders disagreed on directness. Agreements are shown in boldface.

C1/C2	ACTIND	ACTDIR	INFIND	INFDIR
ACTIND	**178**	17		13
ACTDIR	11	**67**		16
INFIND				
INFDIR	4	2		**11**

Table 2: Agreements and Disagreements between Two Coders on Cognitive Status and Anaphora Type

In a number of cases, one coder coded a given token as a situation, while the other analyzed the same token as an event or an activity, as in the example in (14).

(14) Coder 1 <P num="1778"> <1>Some companies contracted work to firms in India, where wages are lower.</1> But some industry officials say there was also less of **<that ACTIVATED INDIRECT "the situation of companies contracting work to India" 1778.1 A3 num="35">that</that>** than had been predicted.</P>

Coder 2: **<that INFOCUS DIRECT "the activity of contracting work to firms in India" 1778.1 IF6 num="35">that</that>**

Even within the agreed-upon tokens classified as ACTIVATED INDIRECT, the precise referent type was only agreed on 59% of the time, as Table 3 shows.

C1/C2	PROP	FACT	SIT	SP.ACT	OTHER
PROP	**6**			1	
FACT	11	**59**			
SIT	3	35	**29**	2	11
SP.ACT	2			**7**	
OTHER					**4**

Table 3: Agreements and Disagreements between Two Coders on Referent Type

As can be seen from the table, the greatest difficulty was in distinguishing facts from situations. (15) shows an example of this type of coding disagreement:

(15) Coder 1. …<1>The result, though, is that nearly 75 percent of the delegates will be chosen by the middle of March, even as the primaries continue through June 6.</1> </P>

<P num="2945"> But is **<that ACTIVATED INDIRECT "the situation that 75% of the delegates are chosen early" 2944.1 A3 num="31">that</that>** progress?

Coder 2.: **<that ACTIVATED INDIRECT "the fact that nearly 75% of the delegates will be chosen by the middle of March" 2944.1 A3 num="31">that</that>**

We are analyzing the reliability data with a view to improving and making more precise our coding criteria, but it is also the case that sometimes a given example was genuinely ambiguous, like example (16). Interpreting the pronoun as referring to a particular rate or to the fact that the rate is what it is both seem to be valid interpretations.

(16) Coder 1 <P num="1656"> <1>The TB rate at the INS center in 1997, the latest year for which

statistics are available, was equivalent to 190 per 100,000 people.</1> To put **<that ACTIVATED INDIRECT "the fact of the INS TB rate to be 190/100k " 1656.1 A3 num="33">that</that>** into perspective, the country as a whole that year had 7.2 cases per 100,000.</P>

Coder 2: **<that ACTIVATED DIRECT "the rate of 190 per 100,000 people" 1656.1 A1 num="33">that</that>**

5. Pilot Study on *It*

Finally, we conducted a small pilot study of *it*, coding 43 tokens in Transcript 3. The results were that agreement on both COGSTAT and TYPE was 76.7%. Expected agreement for cognitive status was .67 and observed agreement was .86, yielding a kappa score of .58. For directness type, expected agreement was .81 and observed agreement was .84, yielding a kappa score of only.16. The kappa score is influenced by the fact that so few tokens were coded INDIRECT, but we plan to investigate the results, improve our coding guidelines, and perform a larger study.

The main results were that 97% of agreed-upon *it* tokens were coded as DIRECT, which supports our hypothesis that the anaphoric relationship tends to be direct for the personal pronoun *it*. Most of the direct cases had NP/DP antecedents; however some had non-nominal antecedents. Instances of *it* with non-nominal antecedents were classified as INFOCUS DIRECT if they expressed the event or activity denoted by a main clause antecedent. An example is shown in (18):

(18)   <P num="100"> <1>The likes of Hochstein, 300-pound Dave Volk, 295-pound Dominic Raiola, 300-pound James Sherman, 320-pound Adam Julch and 300-pound Jon Rutherford sent the normally stingy Tennessee defense sliding backward all night.</1> **<it INFOCUS DIRECT "the event of the Tennessee defense sliding backward all night" 100.1 IF6 num="14">It</it>** happened only one game after the Nebraska line did the same to Texas in a 22-6 Big 12 title game rematch of the October loss that proved to be no match. </P>

As 81.8% of *it* tokens were coded as INFOCUS. the pilot study only weakly supports our hypothesis that *it* requires its referent to be in focus; however, this percentage represents a conservative estimate and is most likely lower than it would be if coders had not been restricted to coding referents as being in focus only if they met one of the INFOCUS criteria in the coding protocol. This did not skew the results for demonstratives, as either activated or in focus status would be consistent with the

hypothesis that demonstratives only require activation (and anything in focus is also activated); however, it did cause the results to provide weaker support for the hypothesis concerning *it* and in focus status, than if coders had not been restricted to the guidelines.

Thus, if a referent was introduced in direct object position in the preceding sentence, it was coded as only activated, even though direct objects, and sometimes even objects of prepositions, can bring a referent into focus of attention under appropriate semantic and pragmatic conditions (see Gundel et al. (1993 inter alia)). That is, syntactic position is not the only factor that determines salience of the referent, but the protocol is not yet sufficiently refined to include semantic and pragmatic factors. An example of a referent that is likely in focus, but would only be coded as activated given the current coding protocol alone is shown in (17):

(17) <P num="815"> Somebody showed Janikowski <1>a football</1>. <2> He kicked **<it ACTIVATED DIRECT "the football" 815.1 A1 num="41">it</it>.**

6. Conclusion

In conclusion, we found support for both our hypotheses. First, *this* and *that* require only activation, while *it* requires in-focus status. Second, the anaphoric relation between a pronoun and its antecedent tends to be direct for *it* but indirect for *this* and *that*, which supports the proposed explanation for the strong association of demonstratives with non-nominal antecedents. We carefully calculated reliability statistics for the coding of both cognitive status and directness, and will compare the resulting data in order to improve our coding guidelines.

7. References

Arstein, R. and Poesio, M. (2005), Annotating (anaphoric) ambiguity. *Proceedings of the conference on Corpus Linguistics.* University of Birmingham, England.

Byron, D. (2002). Resolving pronominal reference to abstract entities. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics.* 80-87.

Cornish, F. (1999). *Anaphora, Discourse, and Understanding. Evidence from English and French.* Oxford: Clarendon Press.

DuBois, J.W., Chafe, W. L., Meyer C., and Thompson, S.A. (2000). Santa Barbara Corpus of Spoken American English—Part 1. Linguistic Data Consortium.

Eckert, M. and Strube, M. (2000). Dialogue acts, synchronizing units and anaphora resolution. *Journal of Semantics* 17. 51-89.

Gundel, J.K. (2004). Coding protocol for statuses on the Givenness Hierarchy. Unpublished manuscript, University of Minnesota.

Gundel, J.K., Hedberg, N. and Zacharski, R. (2005). Pronouns without NP Antecedents: How do we know when a pronoun is referential? In Antonio Branco, Tony McEnery and Ruslan Mitkov (eds.). *Anaphora Processing: Linguistic, Cognitive and Computational Modelling*. John Benjamins. 351-364.

Gundel, J. K., Hedberg, N. and Zacharski, R. (2004). Demonstrative pronouns in natural discourse. *Proceedings of the Fifth Discourse Anaphora and Anaphora Resolution Colloquium*. São Miguel, Portugal. 81-86.

Gundel, J.K., Hedberg, N. and Zacharski, R. (1993). Cognitive status and the form of referring expression in discourse. *Language* 69:274-307.

Gundel, J.K., Hegarty, M. and Borthen, K. (2003). Cognitive status, information structure, and pronominal reference to clausally introduced entities. *Journal of Logic, Language and Information* 12. 281-299.

Hegarty, M. (in press). Context dependence and semantic types in the interpretation of clausal arguments. In Nancy Hedberg and Ron Zacharski (eds.) *The Grammar-Pragmatics Interface*. Amsterdam: John Benjamins, 171-188.

Hegarty, M., Gundel, J.K. and Borthen, K. (2001). Information structure and the accessibility of clausally introduced referents. *Theoretical Linguistics* 27:1-24.

Webber, B. (1991). Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes* 6: 107-135.