Abstract

This paper is concerned with such concepts as `topic`, `focus` and `cognitive status of discourse referents', which have been included under the label 'information structure', as they relate in some sense to the distribution of given and new information. It addresses the question of which information structural properties are best accounted for by grammatical constraints and which can be attributed to non-linguistic constraints on the way information is processed and communicated. Two logically independent senses of given-new information are distinguished, one referential and the other relational. I argue that some phenomena pertaining to each of these senses must be accounted for in the grammar, while others are pragmatic effects that do not have to be represented in the grammar, since they result from interaction of the language system with general pragmatic principles that constrain inferential processes involved in language production and understanding.

1. Introduction

I will be concerned in this paper with such concepts as `topic`, `focus` and `cognitive status of discourse referents', which have been included under the label 'information structure' (alternatively 'information status'), as they relate in some sense to the distribution of given and new information. As an invited speaker at this conference, I was asked to address the question: What do we know about information structure that would bear on what a grammatical theory like HPSG needs to take into account? With this in mind, I will focus on the question of which aspects of information structural concepts and their properties are grammatically constrained and which are constrained by general cognitive and communicative principles that are independent of grammar. These are broad questions, and I obviously cannot hope to answer them fully and completely here. Instead, I will outline the kind of framework that I think needs to serve as the background for asking these questions and will make some tentative proposals for selected informational structural facts and properties within that framework.

The approach to pragmatics I will assume here is that of Relevance Theory (henceforth RT). Within this framework, pragmatics is construed as an account of the inferential processes involved in understanding utterances, processes which take as their input the result of linguistic decoding and 'enrich' that input by way of pragmatic inferences for those aspects of a speaker's intended meaning that are left underspecified by linguistic form, e.g. reference and ambiguity resolution and conversational implicature. Language generation and interpretation is thus seen as

constrained by the interaction of two independent systems, one grammatical the other pragmatic, where constraints imposed by the latter follow from the Principle of Relevance (Sperber and Wilson 1986, 1996). The fundamental goal of relevance theoretic pragmatics is to explain how the hearer is able to access the appropriate cognitive context for interpreting an utterance, i.e. which of the grammatically constrained, but still grossly underdetermined set of assumptions available to her is the one she is intended to use in processing the utterance. The distinction within HPSG between CONTENT and CONTEXT (Pollard and Sag 1994), where the value of the latter is the locus of pragmatic information, might at first seem anomalous on such an approach since, within RT, all linguistic input is viewed as constraining the context in which an utterance will be relevant. But the anomaly is only apparent, as it results from equivocation in the use of the terms 'context' and 'pragmatics', specifically whether these are construed as fully cognitive or not. For the purpose of this paper, I will take the formal construct CONTEXT within HPSG in a narrow sense to include those aspects of linguistic form represented by attributes whose values make direct reference to the utterance act and its participants. I take no position here on the question of whether the CONTEXT-CONTENT distinction is still necessary or even feasible under the relevance theoretic view of pragmatics outlined above, but this should have no bearing on the arguments presented. The main question will be what needs to be represented in the grammar and what doesn't, independent of where and how it is represented.

2. What is Information Structure? Referential vs. Relational Givenness.

'Information structure' is a cover label for a number of distinct , though partly overlapping, concepts that have often been conflated in the literature. While many researchers have recognized that there are distinct notions involved here (cf. Birner and Ward 1998, Chafe 1976, Gundel 1988, Halliday 1967, Lambrecht 1994, Prince 1992, inter alia), there is as yet no general agreement on what the linguistically relevant constructs are, how many of them there are, and how and if they are related (see Gundel 1999a and Gundel and Fretheim 2003.) The situation is confounded by the fact that the different concepts all relate in one way or another to the distinction between given and new information, but in different ways; and even those who recognize the distinction between different informational structural concepts, treat the given-new distinction (at least implicitly) as if it were a unitary phenomenon. As Birner and Ward note (1998, p. 9) this work shares "a general approach based on the degree to which information is assumed to be available to the hearer prior to its evocation." Their own work, following Prince (1992), recognizes a three-way distinction between what is old/new to the hearer, what is old/new to the discourse, and an

'open proposition that is shared knowledge' and 'represents what is assumed by the speaker to be salient (or inferable) in the discourse' (p. 12). But these three senses of givenness-newness are not logically independent. An 'open proposition that is shared knowledge' as well as anything that is Discourse Old is, by definition, also Hearer Old; these concepts differ only in the source of the givenness/newness (the discourse or general knowledge) and the nature of the object that has the givenness/newness property (a discourse entity or an open proposition). Since it is the link to given and new information that has been assumed to tie the various information structural concepts to 'contextual/pragmatic' information, a clear distinction between different senses of givenness/newness is crucial for understanding how and if various information structural properties are constrained by the grammar.

In my own work (e.g. Gundel 1988, 1999) I have argued that there are two distinct and logically independent senses of givenness-newness, one referential and the other relational. Referential givenness describes a relation between a linguistic expression and a corresponding non-linguistic (conceptual) entity in (a model of) the speaker/hearer's mind, the discourse, or some real or possible world, depending on where the referents or corresponding meanings of these linguistic expressions are assumed to reside. The relevant parameters are whether or not it already exists in the model, its degree of salience and, for some authors (e.g. the distinction made by Prince 1992 and Birner and Ward 1998), how it got there and what kind of entity it is. Some representative examples include existential presupposition (e.g. Strawson 1964), various senses of referentiality and specificity (e.g. Fodor and Sag 1982, Enç 1991), the familiarity condition on definite descriptions (e.g. Heim 1982), the accessibility levels of Ariel (1988), the activation and identifiability statuses of Chafe (1994) and Lambrecht (1994), the familiarity scale of Prince (1981), and the cognitive statuses of Gundel, Hedberg and Zacharski (1993).

Relational givenness-newness, by contrast, involves a partition of the semantic/conceptual representation of a sentence into two complementary parts, X and Y, where X is what the sentence is about (the topic, theme, ground, logical/psychological subject) and Y is what is predicated about X (the comment, rheme, focus, logical/psychological predicate). X is given in relation to Y in the sense that it is independent, and outside the scope of, what is predicated in Y. Y is new in relation to X in the sense that it is information that is predicated (asserted, questioned, etc.) about X. Unlike referential givenness, this sense is a relation between two elements on the same level of representation, and can be defined independent of a speaker's assumptions about the hearer's knowledge or attention state. The relation may be construed as logico-semantic, a subject – predicate relation, or as conceptual/psychological/cognitive, the relation between an

entity represented in the hearer's memory (a file card to use a common metaphor) and what is added in relation to that entity. In either case, the distinction can be taken to reflect how the informational content of a particular event or state of affairs expressed by a sentence is represented and how its truth value is to be assessed. Examples of relational givenness-newness pairs include traditional notions of logical/psychological subject and predicate (e.g. van der Gabelenz 1868) presupposition-focus (e.g. Chomsky 1971, Jackendoff 1972, 2000), topic-comment (e.g. Gundel 1974/89, Reinhart 1981), theme-rheme (e.g. Mathesius 1928, Kuno 1972, Sgall et al 1973, 1986, Vallduví 1992), and topic-predicate (Erteschik-Shir 1997).

Referential and Relational Givenness-Newness are logically and empirically independent of one another. An entity can be referentially given, but part of what is relationally new, as in (1).

(1)!!! A. Who called?!!!!!!!!!!

    B. Pat said SHE called. (Gundel 1980)

If *SHE*!refers to Pat, its referent is referentially given in virtually every possible sense. It is presupposed, specific, familiar, activated, in focus, hearer old, discourse old, and so on. But Pat is relationally new, (part of) the focus/comment/main predication, and so receives a focal accent here. Similarly, the referent of HER in (2), Mrs. Clinton, is referentially given, but relationally new, i.e. (part of) the focus of the sentence.

2) A. Good morning. I'm here to see Mrs. Clinton again.

    B: Sure. Mr.Smith. Let's see…One of her assistants will be with you in a second.

    C. I'd like to see [HER$_F$] today. I'm always talking to her assistants. (Vallduví and Engdahl 1996)

So-called `informative presupposition` clefts (Prince 1978) provide another example.

(3)The federal government is dealing with AIDS as if the virus was a problem that didn't travel along interstate highways and was none of its business**.** It's this lethal national inertia in the face of the most devastating epidemic of the late 20th century that finally prompted one congressman to strike out on his own. [Ellen Goodman, op-ed column, 5/35/87, cited in Hedberg (1990)]

The underlined cleft clause in (3) is part of the relationally new information predicated about the topic of this sentence (the national inertia regarding AIDS), as indicated by the fact that it is the locus of focal stress

(on 'own'). However, like all cleft clauses, it also has some degree of referential givenness. As Prince (1978) notes, it is treated by the speaker as if it were generally known, even though it may not be known to the hearer. Hedberg (2000) proposes an account that treats the content of the cleft clause as having some degree of referential givenness (albeit the lowest possible one) even for the hearer, since the hearer is expected to be able to construct a unique representation, the x that prompted one congressman to strike out on his own (against the AIDS epidemic), even if she has no previous knowledge that something fits this description. Hedberg argues that this property follows from the fact that the cleft pronoun and cleft clause form a discontinuous definite description and thus have the same referential givenness property as other definite descriptions, i.e. it must be uniquely identifiable (see below).

3. Grammatical Constraints or Pragmatic Constraints?

Having distinguished the two different senses of givenness-newness that pertain to various information structural concepts, we are now ready to ask the main question: how much belongs in the grammar ? Due to space and time limitations. I will restrict the discussion here to referential givenness. See Gundel (forthcoming) for more complete discussion of both types of givenness-newness.

3.1 Referential Givenness-Newness. What's in the grammar?

The referential givenness-newness concepts I will assume here are the cognitive statuses proposed in Gundel, Hedberg and Zacharksi (1988, 1993). While these were originally proposed to account for the distribution and interpretation of referring expressions, they could in principle play a role in other aspects of language as well. It also remains to be demonstrated whether they are the only referential givenness notions that are linguistically relevant and whether related concepts, such as those noted in section 2, can be reduced to these. I think they can, but I will not be concerned with this question here.

Gundel, Hedberg and Zacharski start from the (uncontroversial) premise that the descriptive content of a nominal expression grossly underdetermines its interpretation. For example, the conceptual content encoded in the phrase *these primitive reptiles* in (4) constrains possible interpretations to primitive reptiles (assuming it is not used metaphorically), [1] but it provides no information about which primitive

---

[1] As Green (1997:9) notes, the condition on the descriptive content is pragmatic rather than semantic, namely "that the speaker believe that the addressee will recognize the speaker's intention in USING the expression that its index be

reptiles are intended. In (5), the pronoun *they* provides even less descriptive information, as it only encodes the conceptual content that the intended referent is third person plural.

(4) A restudy of pareiasaurs reveals that these primitive reptiles are the nearest relatives of turtles. [M.S.Y. Lee, The origin of the Turtle Body Plan.Science, 1993, p. 1649].

(5") A restudy of pareiasaurs reveals that **t**hey are the nearest relatives of turtles.

Yet English speakers have little trouble identifying the intended referents of both phrases as pareiasaurs, even if they don't know what pareiasaurs are.

The referent of (6), on the other hand, is not so easily resolved, and the most accessible interpretation here is one that is not coreferential with pareiasaurs (though it may be a set that includes pareiasaurs), despite the fact that the descriptive content is the same as for *these primitive reptiles* in (4).

(6) A restudy of **pareiasaurs** reveals that **the primitive reptiles** are the nearest relatives of turtles.

Gundel et al propose to account for such facts with a theory whose basic premise is that some determiners and pronouns constrain possible interpretations by conventionally signaling different cognitive statuses (memory and attention states) that the intended referent is assumed to have in the mind of the addressee. They propose six cognitive statuses, which are implicationally related in the Givenness Hierarchy in (7)

(7)   The Givenness Hierarchy   (GH) (Gundel, Hedberg and Zacharski 1993)

in focus	> activated	> familiar >	uniquely identifiable>	referential >	type identifiable
it	*this/that/ this* N	*that* N	*the* N	indefinite *this* N	*a* N

anchored to" the intended referent. In most uses, this involves an assumption that the expression is normally used to refer to objects that have the properties encoded by the descriptive content of the phrase.

Statuses on the GH are conventional meanings of the form or forms listed under them.[2] Since each status entails all statuses to the right on the hierarchy (anything in focus is by definition also activated, anything activated is also familiar, and so on), a form that has a particular status as its conventional meaning is unspecified for higher statuses (statuses to the left) on the hierarchy, but does not exclude them. The forms thus restrict possible referents to those that are assumed to have (at least) the designated memory and attention status for the addressee. They can be thought of procedurally as processing instructions, as follows:

Type identifiable - identify what kind of thing this is.
Referential- associate a unique representation by the time the sentence is
    processed
Uniquely identifiable - associate a unique representation by the time the
    nominal is processed
Familiar - associate a representation already in memory
Activated - associate a representation from working memory
In focus - associate a representation that your attention is currently
    focused on.

Consider, for example, the sentences in (8a-f).

(8) I couldn't sleep last night.
a. A train kept me awake.
b.This train kept me awake
c. The train kept me awake.
d. That train kept me awake.
e. This train/this/that kept me awake.
f. It kept me awake.

The statuses range from least restrictive, 'type identifiable', to most restrictive, 'in focus'. In (8a) the addressee is only expected to identify what kind of thing a train is. In (8b) (on the 'indefinite *this*' interpretation), he is expected to associate a unique representation with the phrase *this train* by the time the sentence is processed. (8c) tells the addressee that he is expected to associate a unique representation by the time the noun phrase is processed. He can do this either by retrieving an existing representation from memory or by constructing a new unique representation. In (8d), he is told that he already has a representation of the train in memory; in (8e) he is instructed to associate a representation from

---

[2] Gundel, Hedberg and Zacharski investigated 5 languages in their 1993 work (English, Russian, Japanese, Mandarin, and Spanish,) The theory will be illustrated here using only English examples.

working memory; and in (8f) he is told to associate a representation that is currently in focus.

The theory makes a wide range of predictions, both categorical and probabilistic, about the distribution and interpretation of referring expressions. I discuss only a sample of these here. (See Gundel, Hedberg and Zacharski 1993, 2001 for more detailed discussion.) The interpretive facts in (4)-(6) above are explained as follows. The demonstrative determiner *this/these* explicitly signals that its referent is at least activated. Since there is only one plural entity in working memory at the particular point when the phrase *these primitive reptiles* is encountered, the referent is automatically resolved as pareiasaurs, even if the reader doesn't know what pareiasaurs are. The explanation in (5) is similar. The pronoun *they* requires its referent to be at least activated, and if unstressed (as is probably the case here) in the current focus of attention. Only one entity meets this condition here, pareiasaurs. So again the reference is automatically resolved, even without knowledge of what pareiasaurs are. The definite article in the phrase *the primitive reptiles* in (6), on the other hand, only requires the referent to be uniquely identifiable. The activated/in focus pareiasaurs meets this condition as it is already represented in working memory due to its mention in the previous sentence and anything activated is, by definition, also uniquely identifiable. But successful resolution here would depend on the interpreter's knowledge that pareiasaurs are primitive reptiles. Moreover, other primitive reptiles that might be represented in memory would meet the condition of being uniquely identifiable as well. And it would also be possible to construct a new unique representation of the whole class of primitive reptiles, if one doesn't already exist in memory. This is why the phrase in (6) has a different interpretation, and is also more difficult to resolve, than the corresponding phrases in (4) and (5).

Since cognitive statuses are properties of mental representations, not linguistic entities, it should be irrelevant how something acquires a particular status, e.g. whether by being linguistically introduced, by being present in the spatiotemporal context, or by being part of general background knowledge. The theory thus predicts correctly that linguistically introduced and non-linguistically introduced entities will be encoded in the same way. It also doesn't make a difference what type of thing is being referred to, e.g. whether it is a concrete object or an abstract entity such as a proposition or a fact, except in cases where the way such entities are introduced has bearing on cognitive status. This is illustrated in the following examples.

(9) Dentist to patient:

Did that hurt? [from Jackendoff 2002]

(10)"We believe her, the court does not, and that resolves the matter,"
     [NY Times, 5/24/ 00]

(11) I tried the shirt on, but that was too big.

In (9) the pronoun *that* is used to refer to something the dentist just did, a representation of which can be assumed to be activated for the addressee, and thus meets the necessary condition for using this form. In (10), the same form is used to refer to a fact that can also be assumed to be activated, in this case because it has just been introduced linguistically by uttering the preceding clause. And in (11) *that* is being used to refer to an object, the shirt, that was activated by its mention in the preceding sentence.

(12) At one point, the hijacker fired a shot inside *the cockpit*, perhaps accidentally, one of the three pilots aboard said....[14 sentences later]Those aboard the plane did not get a good look at the hijacker because when he stood up, he told everyone to hide their faces in their laps and not look at him, then he walked to *the cockpit*, passengers said in radio reports."   [Associated Press, Hijacker Leaps to Safety after Robbing Passengers. 5.25.2000.]

(13) (Passenger on a plane) Do you know if the cockpit door is locked?

In (12), the definite article is used in referring to a cockpit that the hearer can be expected to uniquely identify, either by associating it with an existing representation in memory or by constructing a new representation that links it, by way of a bridging inference (Clark and Havilland 1977), to the recently mentioned plane. [3] In (13), the phrase *the cockpit* is also used to refer to an entity that can be uniquely identified/represented by way of a bridging inference to an already activated entity, in this case the plane that the speaker and addressee are in.

(14) (Dentist to patient, who just winced) Did it hurt?

---

[3] Note that the constraint on cognitive status itself cannot explain why the interpreter chooses the cockpit of the currently active plane over other cockpits that may be represented in memory and would thus be uniquely identifiable. An explanation of this requires an appeal to pragmatic (i.e. non-grammatical) constraints, specifically Relevance (see Gundel 1996 for further discussion of this point.)

(15)   A. I finally had my wisdom tooth pulled.

       B. Did it hurt?

In (14), the patient makes it clear that whatever the dentist just did is in his focus of attention, thus licensing the use of *it*. In (15*), it* is ambiguous between an interpretation where it refers to the process of A having his tooth pulled and one where it refers to the tooth itself. Each of these interpretations can be assumed to be in A`s focus of attention because he just mentioned it.

    Facts like those discussed above and many more like them can be accounted for straightforwardly in the grammar by constraining the relevant pronouns and determiners so that their CONTEXT attributes, and those of the phrases they are a part of, have the required cognitive status values associated with them.[4] The constraints have access to pragmatic/contextual information only in the narrow sense that they make reference to the addressee's memory and attention state (more specifically to the speaker's mental model of that state). But in other respects, they are no different than other aspects of the conventional meaning of lexical items and thus clearly belong in the grammar. The cognitive status constraints could be viewed as an extension of the general framework for representing reference outlined in Green (1997), (or some version thereof) where contextual information is necessarily a part of the representation of all reference.

3.2. Referential givenness-newness. What's not in the grammar?

3.2.1. Salience-promoting factors

    While statuses themselves are independent of how and if a particular entity was linguistically introduced, linguistic factors can influence the hearer's attention state with respect to some entity, specifically whether it is merely activated or brought into focus of attention. In English, this is most evident in the distribution of the personal pronoun *it* compared with the demonstrative pronouns *this* and *that*. As seen in (7), Gundel et al (1993) hypothesize that unstressed personal pronouns, including *it,* require their referents to be in focus. The demonstrative pronouns *this* and *that*, on the other hand, only require their

---

[4] Since cognitive status is associated with the referent of the whole phrase and not just the determiner, I am assuming some mechanism for projecting the cognitive status value of individual lexical items to the noun phrase (or determiner phrase) as a whole.

referents to be activated, i.e. in working memory. Since anything in focus is by definition also activated, referents of demonstrative pronouns could be in focus, but they don't have to be, while the referent of *it* must be in focus, as illustrated in (16) and (17)

(16)   The package was on the table.   That looked new.

(17)   The package was on the table.   It looked new.

The demonstrative *that* in (16), could refer either to the package or to the table, as both meet the condition of being at least activated. In (17), on the other hand, an interpretation where *it* refers to the table is much less accessible, if it is possible at all. The package has been introduced in subject position, which always brings an entity into focus, while the table is less likely to be in focus since it has been introduced in a syntactically less prominent position.

     The interpretive facts in (16) and (17) would follow straightforwardly from cognitive status constraints placed on the pronouns *it* and *that* in the grammar. The distinction in cognitive status encoded by these two different kinds of pronoun also provides a clue to the difference in their distribution in referring to entities such as propositions, facts, and situations, when these are evoked by non-nominal expressions. As seen in the examples in (9)-(11) and (14)-(15) above, both forms can be used to refer to such entities as well as to entities that represent concrete objects and ones that are not linguistically introduced at all. However, as shown by a number of studies, the personal pronoun *it* is much less frequently used than the demonstrative when the antecedent is not an NP (Webber 1988, 1991, Hegarty , Gundel and Borthen 2002, Byron and Allen 1998, inter alia). The use of one form rather than the other also sometimes results in a different interpretation. Compare (10) above (repeated here for convenience) with (18), for example.

(10) "We believe her, the court does not, and that resolves the matter," [NY Times, 5/24/ 00]

(18) We believe her, the court does not, and it resolves the matter,

Gundel et al (1993) attribute such facts to the independently motivated assumption that non-nominal constituents are less likely to bring an entity into focus of attention. The semantic type of the entity and other salience-promoting factors also play a role here (see Gundel, Hegarty and Borthen 2003). Thus, in (19), where the subject of the second sentence refers to the event directly introduced by the first sentence, reference with *it* is acceptable.

(19) Mary fell off her bike. It happened yesterday.

But since the act performed in uttering a sentence is activated, but never brought into focus (as focus of attention will be on some aspect of the content of the speech act, not the act itself) speech acts can only be referenced with a demonstrative, never with the pronoun *it*, as seen in (20) and (21).

(20) A. John snores.
B. That's rude.
B'. It's rude.

(21)   A. I just ate three pieces of cake.
B. Can you repeat that.
B'. ? Can you repeat it.

In (21), *that* is ambiguous between an interpretation where it refers to John's snoring and one where it refers to the addressee's act of saying that John snores. But *it* can only refer to the snoring itself. Similarly, in (21), *that* can refer either to the act of eating three pieces of cake or to the addressee's act of saying that she just ate three pieces of cake. But *it* can only refer to the act of eating the cake.

  The fact that entities introduced by non-nominal expressions are less likely to be accessible to reference with the personal pronoun *it* than with a demonstrative pronoun can thus be shown to follow from interaction of the grammatical constraint that *it*, unlik*e this/that*, requires its referent to be in focus with the non-grammatical fact that certain contexts are more salience promoting than others. For example, introduction in syntactically prominent positions promotes the salience of a referent, whereas performing a speech act directs the addressee's focus of attention to certain aspects of the content of the act, not to the act itself. It may, however, also be possible to account for the facts in question, at least partially, by representing the structural and semantic properties that correlate with the distribution and interpretation of *it* vs. *this/that* directly in the grammar. For example, *it* might be constrained so that it can only refer to entities introduced in certain NP positions (e.g. subject), to clausal complements of factive verbs (see Hegarty et al 2002), to certain semantic types (e.g. objects and events) and so on. Depending on one's goals, such an account might even be preferable to the one proposed here, as it would directly align the facts about referring forms and linguistic contexts without appealing to cognitive status, and specifically to attention states such as 'activated' and 'in focus', which cannot easily be determined by the grammar. But it would fail to explain why the correlations between

referring forms and linguistic contexts are as they are and not otherwise, and would provide little insight into how such forms are processed and interpreted. It would also preclude a principled distinction between facts that are due to (knowledge of) the language system and more general factors governing information processing, such as the role played by linguistic and other factors in promoting the salience of representations. Moreover, there is no single structural context that can be directly correlated with the use of *it* vs. *this/that*, and the relevant factors are sometimes not linguistic at all (Gundel, Borthen, and Fretheim 1999, Hegarty, Gundel and Borthen 2002, Gundel, Hegarty and Borthen 2003). Unless the goals are purely practical ones, then, grammatical constraints on referring forms that make direct reference to cognitive status values would be preferable to ones that attempt to constrain referring forms in terms of the linguistic contexts that contribute to different statuses.

3.2.2. Conversational implicatures

As noted in section 2, the statuses are in a unidirectional entailment relation (anything in focus is, by definition, also activated; anything activated is also familiar, and so on). The informal notion of `definiteness` thus simply falls out as an effect of the hierarchy, since forms that have been characterized as 'definite' are all constrained to refer to entities that are uniquely identifiable by the addressee, either directly, as in the case of the definite article, or by implication, as with forms that overtly signal statuses that entail 'uniquely identifiable' (demonstratives and personal pronouns like 'it', 'she', etc.). This much can be predicted by the grammar, assuming some statement about the unidirectional entailment relation that holds for statuses on the hierarchy, and there is no need for a separate 'definiteness' feature. The hierarchy also predicts correctly that there will be a one to many mapping between statuses and forms in language use, since forms are underspecified for higher statuses, rather than excluding them. Thus, for example, corpus studies have found that less than half of the phrases introduced by a definite article refer to entities that have been previously mentioned in the discourse and 30%-60% (depending partly on the genre examined) refer to entities that cannot be assumed to be familiar to the addressee in any sense, either from the discourse or from general experience (cf. Fraurud 1990, Gundel et al 1993, 2000, Poesio and Vieira 1998). This is perfectly consistent with the Givenness Hierarchy constraints imposed on the definite article by the grammar, since the definite article only restricts possible referents to ones that can be uniquely identified/represented, regardless of whether or not the addressee can be expected to already have an existing representation in memory beforehand. This restriction can be met by entities that are already familiar (regardless of how they became familiar), including ones

that are also activated and/or in focus, as in (22), since anything familiar, activated or in focus is by definition also uniquely identifiable.

(22) A. Oh. So you've only known the dog how long did you say?
       B. Well, about a year, I guess.
       A: Oh well. Is it, uh, how old is the dog?   (Switchboard corpus)

But the cognitive status restriction on appropriate use of the definite article can also be met by entities for which a new unique representation can be constructed, either by way of a bridging inference to a recently activated entity (as in (12) or (13) above), or on the basis of descriptive content encoded in the phrase alone, as is the case for the phrase *the maximum number of boxcars of oranges that I can get to Bath by 7 a.m. tomorrow morning in (23)*.

(23) I want t- I want to determine the maximum number of boxcars of oranges that I can get to Bath by 7 a.m. tomorrow morning …[Trains Corpus. Heeman & Allen 1995]

The various mappings between referring forms and cognitive statuses thus fall out automatically if cognitive status values for different determiners and pronouns are represented/constrained in the grammar, as suggested in the previous section. However, distribution of forms across statuses that meet necessary conditions for appropriate use is not random. Some forms are rarely used, even when necessary conditions for use are met. For example, since the indefinite article only requires type identifiability it should, in principle, be appropriate to use this form for all statuses. In fact, however, the indefinite article is rarely used for statuses higher than 'referential'. Traditional accounts of the difference between definite and indefinite determiner use have accounted for such facts by assuming that non-familiarity (and non-uniqueness) is part of the conventional meaning of the indefinite article. Gundel et al (1993) propose, instead, that the association of indefiniteness with non-familiarity follows from interaction of the conventional meaning of the indefinite article (i.e. type identifiability) with the first part of the Quantity Maxim (make your contribution as informative as appropriate).[5] Since, in most cases, it would be informative (and relevant) to the addressee to know whether or not there is an intended referent that she can uniquely identify, use of the indefinite article (which is unspecified for any status above 'type identifiable') would normally implicate that the addressee cannot

---

[5] An alternative formulation is proposed in Green (2000: 117) – "An agent will do as much as is required for the achievement of the current goal."

uniquely identify the referent.[6]

Similarly, Gundel et al argue, demonstrative pronouns, which require only activation, often implicate that the referent is not in focus, which accounts for their relatively infrequent use compared to the personal pronoun. Demonstrative pronouns are typically used only when conditions for using the more restrictive (hence more informative) pronoun , *it* , are not met. Compare (24) and (25), for example.

(24) Anyway , going back from the kitchen then is a little hallway leading to a window. Across from the kitchen is a big walk-through closet. And next to it**,**

(25) Anyway , going back from the kitchen then is a little hallway leading to a window. Across from the kitchen is a big walk-through closet. And next to that**…**

It in (24) is most naturally interpreted as referring to the kitchen, not the hallway or the closet. This is as predicted by the cognitive status constraint on unstressed personal pronouns, namely that their referent must be in focus. Since the kitchen, unlike the hallway and the closet, is the focal point for the description and has been mentioned twice, it is likely to be in focus at the point when the pronoun is encountered. In (25), on the other hand, the demonstrative *that* is interpreted as referring to the closet, which is activated, but not yet in focus. It is not interpreted as referring to the kitchen, even though the kitchen meets necessary conditions for using a demonstrative pronoun, since anything in focus is also activated. Thus, just as the indefinite article, which is unspecified for statuses above type identifiable, implicates that the referent is not uniquely identifiable, a demonstrative pronoun, which is unspecified for the status 'in focus', typically implicates that the referent is not already in focus, i.e. it implicates a focus shift.

Use of a weaker, less restrictive form doesn't always implicate that a stronger form would not have been licensed, however. For example, the definite article doesn't implicate non-familiarity. As noted above, it is typically used for familiar, and even activated and in focus entities. Gundel et al argue that this is because scalar implicatures arise only when the information that would be conveyed by the stronger form is relevant. For full definite NPs, signaling that the addressee can uniquely identify the

---

[6] Note that use of the indefinite article does not implicate non-referentiality. This is because there is no generally available form in English which explicitly signals referentiality, as indefinite 'this' is restricted to casual speech. The determiner *a* is therefore the most informative choice when the cognitive status is `referential`, but not `uniquely identifiable.'

referent is usually sufficient to allow her to interpret it (given the descriptive content of the NP); so the extra information about cognitive status provided by the demonstrative is typically necessary only in cases like the 'pareiasaurs' example in (4), where the descriptive content is insufficient to allow the addressee to identify the referent. This also explains the relative infrequency of demonstrative determiners as compared to the definite article (see Gundel, Hedberg and Zacharski 1993, 2000 and Gundel and Mulkern 1998 for more detailed discussion.).

Facts like the ones discussed above follow from interaction of the Givenness Hierarchy (specifically constraints on the cognitive statuses signaled by different forms) with general pragmatic principles. As such, they do not have to be directly represented in the grammar, e.g. by constraining the indefinite article so that it refers only to non-familiar entities or demonstrative pronouns so they do not refer to entities in focus. In fact, imposing such restrictions would make incorrect predictions in examples like (26), where *that* refers to the `in focus` kitchen or (27), where *a student of yours* clearly does not refer to someone the addressee is not already familiar with.

(26) John's kitchen is really cozy. That's my favorite room in the house.

(27) A student of yours came to see me today.

4. Conclusion

I have distinguished here two distinct senses of givenness/newness, one referential and the other relational, and have discussed facts relating to the referential givenness notion of cognitive status, demonstrating the relevance of this notion for the distribution and interpretation of different forms of referring expression. Some of these facts can be accounted for by directly incorporating cognitive status into the grammar, specifically as a constraint on specific lexical items (determiners and pronouns). These include, among other things, the fact that determiners and pronouns are not sensitive to whether or not a referent has been linguistically introduced; the infelicity of unstressed personal pronouns in referring to entities not in the addressee' focus of attention; and use of the definite article in referring to non-familiar, but still uniquely identifiable, entities, as well as entities that are not only familiar, but also in focus. Other facts, I have argued, can be attributed to interaction of the language system with non-linguistic principles that govern information processing and therefore do not need to be directly represented in the grammar. These include association of the indefinite article with non-familiarity; association of demonstrative pronouns with focus shift; and the fact that unstressed personal pronouns are more likely to refer to entities that have been

linguistically introduced in a syntactically prominent (e.g. subject) position. In a forthcomng article (Gundel , in preparation), I argue that the situation is similar for facts having to do with such relational givenness notions as topic and focus. These are linguistic concepts, which play a role in the syntax, morphology and phonology of natural languages. As such, they clearly belong in the grammar. But interpretive aspects of these concepts such as familiarity or salience conditions on topics and the 'new information' effect of focus follow from general pragmatic principles, and do not belong in the grammar.[7]

References

Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics* 24:67-87.

Birner, Betty J. and Gregory Ward. 1998. *Information Status and Noncanonical Word Order in English*. Amsterdam: John Benjamins.

Byron, Donna and James Allen. 1998. Resolving demonstrative pronouns in the TRAINS93 corpus. In the Proceedings of the Second Colloquium on Discourse Anaphora and Anaphor Resolution (DAARC-2*)*, 68-81.

Chafe, Wallace. L. 1976. Givenness, contrastiveness, subject, topic, and point of view. In C. Li, (ed.*), Subject and Topic*. New York:Academic Press.

Chafe, Wallace L.1994. *Discourse, Consciousness, and Time*. Chicago University Press.

Chomsky, Noam. 1971. Deep structure, surface structure and semantic interpretation. In D. Steinberg and L. Jakobovits (eds.), *Semantics, an Interdisciplinary Reader in Linguistics, Philosophy and Psychology*, 183-216, Cambridge:Cambridge University Press.

Clark, Herbert . and Susan D. Haviland. 1977. Comprehension and the given new contract. In R. Freedle (ed.) *Discourse Production and Comprehension*. Norwood, N.J.:Ablex, 1-40.

Erteschik-Shir, Nomi. 1997. *The Dynamics of Focus Structure*. Cambridge University Press.

Fodor, Janet D., and Ivan Sag. 1982, Referential and quantificational indefinites. *Linguistics and Philosophy* 5.355-398.

Fraurud, Kari. 1990. Definiteness and processing of NPs in natural discourse. *Journal of Semantics* 7:395-433.

Gabelenz, G. von der. 1868. Ideen zur einer vergleichenden Syntax: Wort und Satzstellung. *Zeitschrift für Völkerpsychologie und Sprachwissenschaft*.6:376-384.

---

[7] See also Gundel 1999b for some preliminary discussion of these points.

Green, Georgia. 2000. The nature of pragmatic information. In Ronnie Cann, Clair Grover and Phillip Miller (eds.) Grammatical Interfaces in HPSG. Stanford, CSLI.

Green, Georgia. 1997. The structure of CONTEXT: the representation of pragmatic restrictions in HPSG. In James Yoon (ed.), *Proceedings of the 5th Annual Meeting of the Formal Linguistics Society of the Midwest*. Studies in the Liiguistic Sciences, .

Gundel, Jeanette K. 1974. *The Role of Topic and Comment in Linguistic Theory*. Ph.D.Dissertation, University of Texas at Austin. Published by Garland, 1989.

Gundel, Jeanette K. 1980. Zero NP-anaphora in Russian: a case of topic-prominence. In *Proceedings from the 16th Meeting of the Chicago Linguistic Society. Parasession on Anaphora*, pp. 139-146.

Gundel, Jeanette K. 1988. Universals of topic-comment structure In M. Hammond, E. Moravczik and J. Wirth (eds.), *Studies in Syntactic Typology*, 209-239, Amsterdam:John Benjamins, 209-239..

Gundel, Jeanette K. 1996. Relevance theory meets the Givenness Hierarchy: an account of inferrables. In J. Gundel and T. Fretheim, (eds.), *Reference and Referent Accessibility*.Amsterdam: John Benjamins.

Gundel, Jeanette K. 1999a. On different kinds of focus. In P. Bosch and R. van der Sandt (eds.), *Focus in Natural Language Processing*, 293-305, Cambridge: Cambridge Univ. Press.

Gundel, J. K. 1999b. Topic, focus, and the grammar-pragmatics interface. In J. Alexander, N-R. Han, and M.M. Fox (eds.) Proceedings of the 23rd Annual Penn Linguistics Colloquium. University of Pennsylvania Working Papers in Linguistics 6.1:185-200.

Gundel, Jeanette K. In preparation. Information Structure: How much belongs in the grammar?

Gundel, Jeanette K., Nancy Hedberg, Ron Zacharski. 1988. The generation and interpretation of demonstrative expressions. *Proceedings of the XIIth International Conference on Computational Linguistics*. John Von Neumann Society for the Computing Sciences, Budapest, 216-221.

Gundel, Jeanette K., Nancy Hedberg, Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69:274-307.

Gundel, J. K., N, Hedberg and R, Zacharski. 2001. Definite descriptions and cognitive status in English:why accomodation is unnecessary. *Journal of English Language and Linguistics* 5.2:273-95.

Gundel, Jeanette K. and Ann Mulkern. 1998. Reference and scalar implicatures *Pragmatics and Cognition*. Special issue on the Concept of Reference in the Cognitive Sciences.

Gundel, Jeanette K., Kaja Borthen and Thorstein Fretheim. 1999. The role of context in pronominal reference to higher order entitites in English and Norwegian. In P. Bouquet et al, eds. *Modeling and Using Context.. Proceedings from the Second International and Interdisciplinary Conference, CONTEXT '99.* Lecture Notes in Artificial Intelligence 1688. Springer.

Gundel, Jeanette K., Michael Hegarty and Kaja Borthen. 2003. Cognitive Status, Information Structure, and Pronominal Reference to Clausally Introduced Entities. *Journal of Logic, Language and Information* 12:281-299.

Gundel, Jeanette K. and Thorstein Fretheim. I993. Topic and Focus. In G. Ward and L. Horn (eds.) Handbook of Pragmatics. London:Blackwell.

Halliday, M.A.K. 1967. Notes on transitivity and theme in English. Part II. *Journal of Linguistics* 3:199-244.

Hedberg, Nancy. 1990. *Discourse pragmatics and cleft sentences in English.* Minneapolis, MN: University of Minnesota dissertation

Hedberg, Nancy. 2000. The referential status of clefts. *Language* 76: 891-920.

Hegarty, Michael, Jeanette K. Gundel, and Kaja Borthen. 2002. Information structure and the accessibility of clausally introduced referents. *Theoretical Linguistics*, pp. 1-24.

Heim, Irena R. 1982. The semantics of definite and indefinite noun phrases. Amherst: University of Massachusetts dissertation.

Jackendoff, Ray.1972. Semantic Interpretation in Generative Grammar. Cambridge: MIT Press.

Jackendoff, Ray. 2002. Foundations of Language. Oxford University Press.

Kuno, Susumu. 1972. Functional sentence perspective. *Linguistic Inquiry* 3.3, 269-320.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form*: *Topic, Focus and the Mental Representation of Discourse Referents.* Cambridge University Press.

Mathesius, V. 1928. On linguistic characterology with illustrations from Modern English. *Actes du Premier Congrès International de Linguistes à La Haye*, pp. 56-63.

Poesio, Massimo and Renate Vieira. 1998. A corpus-based investigation of definite description use. Computational Linguistics 24:183-216.

Pollard, Carl and Ivan Sag. 1994. Head-Driven Phrase-Structure Grammar. University of Chicago Press.

Prince, Ellen F. 1978. A comparison of *wh*-clefts and *it*-clefts in discourse. *Language* 54:883-906.

Prince, Ellen F. 1981. Towards a taxonomy of given-new information. In P. Cole (ed.), *Radical Pragmatics*. New York:Academic Press.

Prince Ellen F. 1992. The ZPG letter: subjects, definiteness, and information status. I*n* S. Thompson and W. Mann (eds.), Discourse *Description: Diverse Analyses of a Fund Raising Text*. Amsterdam:John Benjamins , 295-325.

Reinhart, Tanya 1981. Pragmatics and linguistics. An analysis of sentence topics.*Philosophica* 27:53-94.

Sgall, Petr, E. Hajiìova and E. Beneîova. 1973. *Topic, Focus, and Generative Semantics*. Kronberg: Scriptor Verlag GmbH.

Sgall, Petr, E. Hajiìova , J. Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. Dordrecht: Reidel.

Sperber, Dan, Deirdre Wilson. 1986/95. *Relevance: Communication and Cognition* . London:Blackwell.

Strawson, P.F. 1964. Identifying reference and truth values. *Theoria* 3.96-118.

Vallduví, Enric. 1992. *The Informational component*. New York:Garland.

Vallduví, Enric and Elisabet Engdahl. 1996. The linguistic realization of information packaging. *Linguistics* 34:459-519.

Webber, Bonnie L. 1988. "Discourse deixis and discourse processing."Technical report, University of Pennsylvania.

Webber, Bonnie. L. 1991. "Structure and ostension in the interpretation of discourse deixis." Language and Cognitive Processes 6.2:107-135.