CHAPTER 12

A grammar in every register?
The case of definite descriptions

Mira Ariel
Tel Aviv University

1. Introduction: Different grammars for different registers?

Research in grammaticization has benefited tremendously from statistical data regarding frequency of use. Most functionalists are in agreement that very often the frequent discourse pattern of yesterday is the grammar of today (see Hopper and Thompson 1980, 1984; Traugott and König 1991), because "grammars code best what speakers do most" (Du Bois 1985). Recent research has, in addition, emphasized the importance of less than perfectly general rules (e.g., Thompson 2001). Thus, many have noticed that linguistic expressions have different frequency rates and different distributional patterns in different modalities, registers and genres.1 Both claims rely heavily on statistical counts of various expressions as they occur within specific discourse profiles. Thus, linguistic expressions are seen as associated with specific discourse profiles, those conditions which prototypically obtain when the expression is used. Such profiles may make reference to pragmatic features of the concept involved (e.g., Agents tend to be Given entities—Du Bois 1987), or they may make reference to contextual circumstances (e.g., ’know tends to occur in informal conversations).

Genres seem to be prime candidates for the creation of discourse profiles for linguistic forms. According to Miller (1984; see also Bazerman 1988; Swales 1990), genres constitute recurrent patterns of language use, defined according to social acts/motives. These recurrent patterns are only natural, given that similar social occasions tend to recur, with typical rhetorical problems, which trigger similar responses. Crucially, a genre relates both to substance and to form. Following Biber (1988, 1995) and Biber et al. (1999), I will call "any variety associated with particular situational contexts or purposes" (Biber 1995:1) a register, so as to generalize over all specifically situated speech (genres included). I propose to examine the nature of the association between linguistic forms and register.

---

1. For example, a third of the articles in nine issues of Journal of Pragmatics 33 (20/61) included some specific speech event in the title, thus restricting the applicability of the analysis to a specific discourse variety. For example, “The interactional organization of pharmacist consultations in a hospital setting” (JoP 33:12), “Meta discourse in slogans and headlines” (JoP 33:8).
Formal differences between registers are not hard to find. For example, West (1980) found that *that*-nominal clauses mostly occurred in the introduction and in the discussion sections of research papers in biology. Heslot (1982) found that simple past tense was extremely frequent in the methodology and the results sections of research papers. Simple present tense, on the other hand is much more frequent in the introduction and the discussion sections. Passive voice is quite prevalent in all sections except for the methodology one. Chafe (1982) found (among many other things) that first-person references are much more frequent in conversations than in written English (13.4 times more so).

But is every form that occurs in some register but not others with a statistically significant frequency directly register-dependent? Biber et al. (1995) testify that systematic patterns of use resulting from recurrent communication goals, interpersonal relations between addressees and addressees, and other extralinguistic factors have traditionally not been deemed relevant to the grammar. They too refrain from calling their statistical work grammar. Rather, what the analysis they provide describes is “the actual use of grammatical features in different varieties of English” (p. 4, emphasis added). Indeed, not all patterns of use amount to grammar. I would like to argue that not every statistically significant difference is grammatically significant.

For example, as West (1980) explains, *that*-nominal clauses are perfectly suitable for the introduction and the discussion sections of an article, because they provide a sentential construction for commenting (in the main clause) on statements (encoded in the nominalization). In the introduction, statements are offered on previous claims in the literature. In the discussion, conclusions are drawn from factual results. Given the natural explanation for the high frequency of nominalized clauses in certain sections of the research article, there is actually no need to postulate a direct conventional association between nominalized *that*-clauses and the specific article sections it tends to occur in. Instead, we can assume that these constructions are used according to their grammatical and discourse functional specifications throughout the research article. It just so happens that for certain sections of the article, writers find them more useful, and hence, tend to use them with greater frequency (see also the motivated frequent use of nominalizations and passives discussed by Halmari and Östman (2001), and the changes in the frequency of relative clauses, nominalizations, adverbial subordinate clauses and abstract subjects in the last 100 years of research articles, as discussed by Bazerman (1988)).

Most early studies of register differences (most notably those distinguishing spoken and written English) have adopted precisely this approach. In other words, the frequent linguistic forms are motivated by addressees' goals, rather than being directly coded for specific registers (e.g. Ochs 1979; Chafe 1982; Chafe and Danielewicz 1987; Biber 1986 and onwards—see also Cumming 1994). The reason for assuming only an indirect link between registers and linguistic forms is quite obvious. As Biber et al. (1999) argue, choices among alternative grammatical forms (e.g., between various referring expres-
sions) are pragmatically motivated. These pragmatic motivations are crucially dependent on addressee goals in the various discourse varieties. It is then quite plausible that the application of one and the same grammar would result in systematically different choices in different registers if the different registers are used for different goals, and are uttered under different circumstances. Thus, Chafe (1982) explains all the differences he found between spoken and written English by reference to a few general characteristics of these two registers: Speaking is faster than writing, and affords a direct interaction between the speaker and her addressee. The very large gap in self-references between spoken and written English is explained by noting that in general, informal conversations are characterized by a high degree of involvement. A similar explanation is provided by Argamon et al. (2003) for the cross-genre (rather modest) differences they found between female and male addressees. Men’s discourse is characterized by a lower involvement style, which accounts for why, for example, they have more definite descriptions than women, and for why women have more second-person pronouns.

1.1 Register-related patterns of referring expressions

Both Accessibility theory (Ariel 1985, 1990, 2001) and the Givenness Hierarchy (Gundel et al. 1993) have proposed a general (albeit different) account for the use of referring expressions in general, and of definite descriptions specifically. Both theories hardly address themselves to register differences, although differences between languages are recognized, and there is nothing inherent blocking the assumption of such differences within these theories (in fact, see Ariel 1990:6.1 and section 4 below). Should these theories be modified in order to accommodate the fact that statistically valid register differences have been found with respect to definite descriptions? I would argue that this is not necessary.

A good example of the type of analysis I advocate is Kirsner (1996). Kirsner’s findings (Kirsner and van Heuven 1988) showed that the ratio of the Dutch proximate demonstrative (deze) versus the distal demonstrative (die) was very different in different genres. For example, while the two forms were equally frequent in family magazines, the proximate demonstrative was much more frequent than the distal one in government publications. In general, there was hardly any difference in the frequency of die across genres, but deze was rare in novels but quite frequent in government publications. These distributional patterns undoubtedly constitute different discourse profiles for deze. However, rather than conclude that there are genre-specific use conventions for the proximate demonstrative, Kirsner argued that the different frequency ratios of the two demonstratives stem from the different discourse functions of deze and die. The discourse function of the former (what he terms “high deixis”) is simply more compatible with complex sentences and complex texts. And since some genres (e.g., government publications) contain more complex messages than others (e.g.,
novels and family magazines), the statistical skewing of the distribution of the two demonstrative forms is accounted for without having to attribute to the *deze* discourse functions per se a specification about genres as well. Hence, while the discourse profiles of *deze* remain distinct in different genres, its discourse function remains constant across genres. I propose we offer similar accounts for the register-specific patterns below.

More recent findings which I interpret in support of this approach can be found in Swanson (2003), who compares modes of coreference in academic journals, news magazines and fiction narratives, all concerning the Middle East. Swanson carefully analyzes the differences between the three genres (also noting quite a few individual intra-genre variations), and relates these differences to the different goals of the genres. For example, places are hardly ever coreferred to in the narratives, but are often referred to in the academic writing. As Swanson notes, the role of places is quite different in the different genres. In the narratives, places are merely scene setters, so they are transient referents, not often mentioned a second time. Not so for the journals, where territories may constitute a major topic. Similarly, while humans figure prominently as anaphoric entities in all three genres, in the narratives they constitute over 92% of the anaphors. They take up over 43% of the anaphoric references in the magazines, and just under 20% in the academic journals, where they also seem not to be individuated. Indeed, a quick look at the tabulated data shows an ascending use of personal pronouns starting with the academic journals, where they are rather rare, and ending with the narratives, where they are quite frequent. On the other hand, probably since concepts are anaphorically referred to quite often in both the journals and the magazines (14% and 13% respectively) (but not in the fiction), these two text types contain many more elaborate anaphoric definite descriptions. Different goals, then, call for different types of entities. These, in turn, call for different referring expressions.

Recently, however, other conclusions have been drawn from register-related distributional patterns. Specifically with respect to referring expressions, some researchers have argued for “different grammars for different genres”. Thus, Fox (1987:152) concludes that

> there is no single rule for anaphora that can be specified for all of English . . . instead, we have a variety of specific patterns which obviously share a number of general characteristics, but which nevertheless differ enough to require separate formulations [emphasis added].

In fact, Fox also proposes that there are different conventions for different written genres (e.g., narrative versus informative texts, p. 143). Lord and Dahlgren (1997:339) seem to agree with her:

> the choice of anaphor is a function of segmentation. Since different genres may have different discourse structures and segmentation options . . . this finding means that choice of anaphor form is a function of genre [emphasis added].
And about the distinction between proximal and distal demonstrative distribution, they say (p. 346) that “These patterns . . . have become part of the tacit knowledge of the genre, shared by the writer and reader . . .” [emphasis added]. In other words, these authors are not content to note that speakers’ goals mediate between the register and the distributional patterns. They see a direct link between the linguistic patterns and specific registers.

The goal of this article is to examine cases presented in the literature, all pertaining to the use of definite descriptions, all demonstrating statistically significant register differences. I will argue, however, that significant statistical findings do not necessarily reflect significant grammatical generalizations. More controversially, I will tentatively propose that the discourse patterns revealed tend not to grammaticize. The claim is, then, that whereas some statistical findings reflect potential grammar in the making, for others, grammar is quite irrelevant. Some significant statistical differences are to be attributed directly to speakers’ goals in various contexts, rather than to their “specific-register grammar”. Thus, even though speakers employ the very same grammatical rule across different registers, the discourse profiles of the linguistic expression come out differently in different registers, because different goals/functions are better served by different means. The different discourse patterns are then accounted for by factors external to the linguistic forms at hand, and there is no need to assume different grammatical generalizations for different discourse varieties.

In section 3, I will present various findings attesting to register-related differences in the distribution of definite descriptions. However, rather than conclude that there are register-specific use conventions for definite descriptions, I will propose that different registers prototypically call for different types of discourse entities. While the discourse function of definite descriptions remains constant across all registers (indicating a low degree of accessibility for the mental representation of the entity referred to), its implementation naturally varies according to the discourse entities involved. In other words, the same grammatical rule, used to encode different entities, will only naturally yield different surface realizations. To give just one example, assuming that definite descriptions code various degrees of low accessibility, it is only to be expected that definite descriptions should be more frequent in registers which call for the use of many low accessibility discourse referents.

But before we turn our attention to real statistical differences in the use of definite descriptions, we should distinguish between real register findings, as will be discussed in section 3, and ones which are only apparent, as will be discussed in section 2. The data discussed below is taken from various published sources (Fox 1987; Ariel 1990, 1996; Biber et al. 1999; Wolf and Walters 2001a,b), as well as from a slightly edited narrative examined especially for this chapter (Morris 1994: 129–39), containing 204 referential definite descriptions. Unless otherwise indicated, all the cited examples in this chapter are drawn from Morris (1994), and each example is identified by the page number followed by the number I assigned it.
2. Language-dependent differences, or what's discourse anaphoric?

The appropriateness condition on the use of definite descriptions is that the entities they refer to are identifiable. Linguists had to then characterize those sources which can provide a basis for this identifiability. It has long been established that our general knowledge, the speech situation and a previous discoursal mention may all enable identification of entities, and thus allow for the use of definite descriptions (see the discussion in Ariel 1988, 1998). It would then seem reasonable to characterize the prototypical discourse profile of definite descriptions by reference to these contexts (see Clark and Marshall 1981, for example). In fact, the three-way distinction among contexts has often been reduced to a two-way distinction: First-mention (based on general or situational knowledge) versus discourse anaphoric (based on prior mention). Indeed, some version of this distinction seems relevant for at least some languages, which seem to distinguish between anaphoric and general knowledge definite articles. According to Lyons (1999), some languages only have anaphoric definiteness articles, while others have distinct forms for anaphoric and general knowledge uses. The anaphoric (including the situational) usage and the general knowledge usage are potentially grammatically distinct, then.² No wonder many linguists have attempted to characterize the discourse function of definite descriptions by reference to a first-mention/discourse anaphoric distinction. I have argued, however, that the distinction is not all that easy to make, nor all that important in explaining the use and interpretation of referring expressions.

There is no agreement among researchers about what the prototypical discourse profile of definite descriptions is re first-mention vs. discourse anaphoric. Are definite descriptions mainly anaphoric (referring to a previously mentioned entity) or are they mainly first-mention referring expressions? The philosophers, who initiated the discussion of definite descriptions, devoted their analyses to isolated first-mentions (e.g., The king of France). Computational linguists (e.g., Sidner 1979) and semanticists, such as Heim (1982), consider the anaphoric (subsequent mention) use as the basic use of definite NPs. I have found that on the average, about a third of the definite descriptions in my Hebrew written data were first-mentions, and two thirds were discourse anaphoric (see Ariel 1990: 35), but Fraurud (1992) is virtually dedicated to proving the opposite, that overwhelmingly, definite descriptions introduce new entities into the discourse. Maes and Noordman's (1995) results are closer to Fraurud's, but are still different: Nearly a quarter of the definite NPs in their written Dutch data are classified as anaphoric. Gunde et al. (2001) find that 44% of the definite descriptions in their data are first mention.

So what is the prototypical discourse profile of definite descriptions regarding first versus subsequent mention? Do definite descriptions in different languages simply have different prototypical discourse profiles? It would seem so. Since we all used writ-

---

². But note that situational references are grammatically classified with anaphoric, rather than with first-mention uses.
ten materials as sources, the different discourse profiles could not stem from register differences. However, I do not believe that the differences in the data presented in Ariel (1990), Fraurud (1992), Maes and Noordman (1995) and Gundel et al. (2001) necessarily stem from differences between the languages examined (Hebrew, Swedish, Dutch and English). The bulk of the differences stem from our different conceptions of what should be considered as discourse anaphoric. Fraurud takes a form to be anaphoric only if it is a case of referential dependency, where the addressee’s interpretation must absolutely rely on another expression, necessarily an NP antecedent. Furthermore, that antecedent must form an explicit reference to the very same discourse entity. That is certainly a legitimate semantic approach, and I estimate that in my data too, fewer instances of definite descriptions would count as discourse anaphoric under Fraurud’s restricted definition.

However, I believe that in order to account for the specific referring expression selected by the speaker, as well as for the processing involved in the interpretation by the addressee, a weaker notion of discourse anaphora is more fruitful. First, note that an interpretative dependence can be shown even if the material relied upon (as antecedent) is not an explicit mention of a referent. This occurs in (1a), where comments at the bottom in green ink do not constitute an NP, and do not refer to one discourse entity at that stage (see also Garnham et al. 1992 on conceptual anaphors and McKoon et al. 1993 on nonreferential antecedents), and in (1b) where a few propositions (rather than discourse entities) mentioned in the previous paragraph form the antecedent for the same thing:

(1) a. ... someone had written comments at the bottom in green ink. One had a naked woman with an older man, and the inked-in comments referred ... (131:65).

b. ... she told me that ... she implied ... Then she advised me ... And then she said ...

[New Paragraph] The dean pretty much told me the same thing (133:105).

Moreover, while we may independently access (i.e., as new) discourse entities such as the university, the state defamation suit and the dean, all referred to multiple times in Morris (1994), the narrative actually requires us to accumulate information regarding the same dean, suit and university. Thus, we need to connect the current reference with previous mentions of that entity, and that entity specifically (in the narrative examined, a few universities and suits are mentioned, and we need to track the “right” one each time). In other words, anaphoric definite descriptions need not have explicitly mentioned NP antecedents on the one hand, and on the other hand, they may be referentially dependent even if they are potentially accessible from our general knowledge store (see also Swanson 2003).

3. Maes and Noordman do not provide their working definition of an anaphoric usage.
Note that the second mention of 'the comments' above is shorter, omitting the location of the comments (at the bottom) and the color of the ink (green). Indeed, even when a referring expression is not dependent in Fraurud’s absolute way on an earlier linguistic antecedent for its referential interpretation, the degree of accessibility of the mental representation it retrieves tends to be higher, just because the entity/ies has/ve been previously mentioned in the discourse. Textual counts, as well as production experiments, have shown that speakers take these previous mentions into account when they choose the forms of words and referring expressions (see Fowler and Housum 1987; Gurman-Bard 1995). Thus, even if a definite description is used again (rather than a pronoun), it often shows sensitivity to the fact that it is a second mention. (2b), which occurs in the sentence following that of (2a), is another example of shortening:

(2)  a. Lurid headlines ... appeared in the newspapers every day ... (137:254).
    b. The papers didn’t begin covering the story ... (137:257).

The shorter form invariably follows, rather than precedes the longer form, since the first mention renders the entity a relatively high degree of accessibility. So, even though one could potentially access the papers from one’s general knowledge, the form selected for it shows that the speaker views it as a second mention. Note that this variation in the degree of informativity and expression length (shorter forms for higher accessibility) is applicable to all referring expressions, as (3) shows for a name ((3b) occurs later in the same paragraph as (3a):

(3)  a. They sent me to the academic Vice-president, a man named Leffler ...
    (134:127).
    b. ... the first thing she did was send a letter to Vice-president Leffler ...
    (134).

Note that the referring expression in (3b), although fully lexical and potentially independently referential, is typically shorter and less informative. In Ariel (1990, 1998), I argued that form variability is manifest even in first-person references, which surely do not depend on a previous linguistic mention (of the speaker) for their interpretation. Thus, in Hebrew, the speaker refers to herself with a full (even stressed) pronoun when deemed relatively less accessible in the discourse, but with a cliticized pronoun (or even zero) when highly accessible in the discourse. Her presence and accessibility in the speech situation remains constant, and cannot then be so crucial in and of itself in determining referential form. Hence, while I agree with Fraurud (1992) that not all previous mentions necessarily serve as the semantic sources for deriving the interpretation of the definite description (and definite NPs in general), I maintain that the form and distribution of all referring expressions often take into account previous mentions, because they are dependent on degree of accessibility, which is affected (among other things) by previous mentions. I am arguing, then, that referring expressions are not

4. Interestingly, I found an identical set of examples in conversational English—see (Ariel 2001).
chosen only by reference to their content which is supposed to ensure the addressee's ability to pick out the right referent (Fraurud's implicit assumption). Rather, the speaker has to signal in addition the degree of accessibility of the mental representation retrieved. Potential interpretative independence (Fraurud's definition for first-mentions) does not automatically entail being first-mention. It may merely involve a very low degree of accessibility. My claim has been that potentially independent as well as dependent retrievals are performed by the addressee based on the referring expression chosen by the speaker according to the degree of accessibility it codes. And definite descriptions code a relatively low degree of accessibility, whether anaphoric or first-mention. The classification does not seem to be significant for the use of definite descriptions.

Notice, in addition, that we cannot really generalize over the whole category of definite descriptions. Short and long (less vs. more informative) definite descriptions have radically different discourse profiles in fact. Consider the first-subsequent mention distribution for short (1–2 content words) vs. long (3+content words) definite descriptions:

<table>
<thead>
<tr>
<th>Definite description type</th>
<th>First-mention</th>
<th>Discourse-anaphoric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long definite description</td>
<td>96 (65.3%)</td>
<td>51 (34.7%)</td>
</tr>
<tr>
<td>Short definite description</td>
<td>84 (21.8%)</td>
<td>302 (78.2%)</td>
</tr>
</tbody>
</table>

While two thirds of the long definite descriptions are first-mentions, as predicted by Fraurud, an even larger majority of the short definite descriptions (78%) are discourse anaphoric. Under Accessibility theory, short definite descriptions indicate a higher degree of accessibility, and hence are on the whole more appropriate for subsequent mentions.

Another difference between my definition of discourse anaphoric and Fraurud's concerns inferred entities (newly introduced entities readily inferable on the basis of a previous mention of an entity closely associated with the inferred entity). While Fraurud takes inferables as new entities, I took them as discourse anaphoric. They are indeed an in-between type of entity. The following is a typical example, where the judge refers to an entity stereotypically associated with the federal court:

(4) The first case was tried in federal court in ... Their witnesses said awful things to convince the judge ... (136:206).

Inferred entities are somewhat of a hybrid (see also Prince 1992; Lyons 1999: 4). They crucially rely on a previous discourse representation (the anchor), and in this sense they are not first-mentions. But the specific entity itself has not previously been mentioned in the discourse, and should count as new, therefore. Lyons (1999: 52) notes that in Hausa, the anaphoric definite article (distinguished from the first-mention definite article) is used not only for previously mentioned entities, but also for referents related
to previously mentioned entities (inferred entities). Inferred entities are then classified with anaphoric entities, rather than with new discourse entities. In fact, in many cases it is very hard to decide whether or not discourse anaphora is actually involved (and I suspect that this is true even for Fraurud’s stricter notion of anaphora). The reason for my decision to count inferred entities as discourse anaphoric and not as first-mentions is based on their linguistic coding: A “true” first-mention of a specific judge would have demanded a more informative referring expression (i.e., one marking a lower degree of accessibility than the short the judge). As it is, the introduction of the court makes the referent ‘the judge’ easily inferable and accessible. O’Brien and Albrecht (1991) find that highly context-appropriate first-mentions are rather easily accessed, and in Ariel (1996) I found that frame-induced entities were only referred to by one content word definite descriptions (normally reserved for entities not entertained at a very low degree of accessibility). Entities retrieved purely from general knowledge, on the other hand, required more informative referring expressions for the most part (see also Gundel et al. 1993 for different forms for different inferred entities).

So we can now understand how the different definitions for being discourse anaphoric led different researchers to find different first/subsequent mention ratios for the definite descriptions in their data. I therefore suggest that the different counts presented in Ariel (1990), Fraurud (1992), Maes and Noordman (1995) and Gundel et al. (2001) are not actually important, because they do not necessarily constitute different discourse profiles for definite descriptions. In fact, as I have argued at length (Ariel 1994, 1996), the first vs. subsequent mention is not the determining factor here. Rather, it is the degree of accessibility with which the entity is entertained by the addressee, as judged by the speaker. Being discourse-anaphoric contributes towards a higher degree of accessibility of some mental representation, but neither it or its absence constitute necessary or sufficient conditions for the use of practically all referring expressions.

3. A grammar in every register?

Section 3 examines cases where different registers do have different discourse profiles for definite descriptions. I analyze the distribution of first versus subsequent retrievals in 3.1, the relative frequency of definite descriptions versus pronouns in 3.2, and the different referential distances for anaphoric definite descriptions in different registers in 3.3. While all these comparisons show different register profiles for definite descriptions, I will argue that the same discourse function can account for the apparent variability in the use of definite descriptions in different registers.

3.1 First versus subsequent mentions

I have argued above that the different ratios for first versus subsequent mentions for definite descriptions in Fraurud (1992) and in Ariel (1990) stem from our different cri-
Table 2. First-mention versus discourse anaphoric definite descriptions in four text types

<table>
<thead>
<tr>
<th>Text type</th>
<th>First-mention</th>
<th>Discourse anaphoric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper editorials</td>
<td>56 (43.75%)</td>
<td>72 (56.25%)</td>
</tr>
<tr>
<td>Semi-academic</td>
<td>51 (39.2%)</td>
<td>79 (60.8%)</td>
</tr>
<tr>
<td>News items</td>
<td>49 (33.6%)</td>
<td>97 (66.4%)</td>
</tr>
<tr>
<td>Fiction</td>
<td>24 (18.6%)</td>
<td>105 (81.4%)</td>
</tr>
<tr>
<td>Total</td>
<td>180 (33.8%)</td>
<td>353 (66.2%)</td>
</tr>
</tbody>
</table>

teria for what counts as discourse-anaphoric. However, other conflicting findings are not similarly reconcilable (e.g., Gundel et al. 2001: Table 2, where the Canadian Parliament transcripts contain three times as many first-mention entities as the spoken TRAINS corpus). Note first the data presented in Ariel (1990), based on four different sources of written materials. All were collected under the same (loose) definition of discourse anaphora. Table 2 is adapted from Table 1.1 in Ariel (1990: 35).

Assuming that my looser definition of discourse-anaphoric has some validity, are we really justified in claiming that about two thirds of definite descriptions are discourse anaphoric and a third are first-mentions (the total numbers)? I think not. Should we then say that the ratio of first versus subsequent mention definite descriptions is sub-genre dependent (editorials, news items, etc., are all sub-genres of the written genre)? After all, proportionately, there are 2.35 more first-mentions in the editorials than in the fiction, so that whereas the first/subsequent mention ratio for the editorials is almost equal (1.3 more subsequent mentions), there are almost 4.4 times more subsequent than first-mentions in the fiction. The answer is once again negative, I propose. Statistically significant differences are not necessarily grammatically significant.

In the narrative examined for this chapter I found an even lower rate of first-mention definite descriptions: 19 (9.3%).5 Moreover, in Ariel (1985) I compared between two types of news items: Items containing brand new information (e.g., on a robbery) and items containing news concerning an ongoing topic in the news of the time (The Lebanon war). It then turned out that the “sub-genre” of news items split into two sub-sub-genres re the first-subsequent mention ratios. The brand-new news items patterned with fiction, whereas the news currently on the agenda patterned with the editorials. These are no doubt distinct discourse profiles for definite descriptions in different text types. But I maintain that the numbers of first versus subsequent mention definite descriptions are unimportant in accounting for their discourse function (except as a basis for a rough comparison with other referring expressions, e.g., pronouns). The distinct discourse profiles above can all be seen to be consistent with an assumption that definite descriptions carry one discourse function: coding a variety of low degrees of mental

5. The reader should be advised that not every potential second mention was counted as such. Thus, the university's hallways, as well as the men's rooms, were mentioned a few times in different sections of this narrative, but since I doubt the addressees actually accumulate the information mentioned in their connection, I counted each as a new (inferred) mention.
accessibility for the entities retrieved. The differences should be attributed to the different contextual assumptions prototypical of the different genres.

I claim that at the basis of the different discourse profiles of definite descriptions above is a difference between the genres as to the common ground appropriately assumed as shared by the writer and the readers. Where the speaker and addressee share more common ground, more entities can be introduced initially as Given (using definite descriptions, for example). Where the speaker and addressee share less common ground, fewer entities can initially be presupposed to be Given. Probably the only similarity between fiction and brand-new news items (both have a low count for first-mention definite descriptions) is that the writer is not in a position to assume a large set of commonly presupposed entities with her addressee. She must therefore introduce most first-mentions as unfamiliar entities. Definite NPs are less suitable for this function. Editorials, on the other hand, mostly comment on topics already on the public agenda, and similarly, in repeated news items on the same topic (The Lebanon war), familiar entities play a prominent role. Such entities can then be taken as Given even in their initial introduction into the discourse, and definite NPs are appropriate. Thus, whereas different registers may very well prototypically involve discourse entities of different types (e.g., many initially shared entities in editorials, few initially shared entities in fiction), leading to different prototypical profiles for definite descriptions, this does not justify the conclusion that the use of definite descriptions is different in the different registers.

The same generalization is responsible for the use of definite descriptions in all the genres and sub-genres above, whether they are mostly first-mentions or mostly subsequent mentions: A relative low degree of mental accessibility of the mental representation retrieved, created by a variety of low accessibility factors. Whether this low degree of accessibility is due to the lack of previous mention of the entity in the current discourse (first-mentions) or to the occurrence of a great distance or paragraph or episode boundary between the previous and current mention (subsequent mentions—see 3.3 below) is immaterial. My point is then that statistically significant different counts for different registers do not automatically entail different referential or anaphoric styles for the same forms. While examining definite descriptions in the registers above with respect to first/subsequent mentions may point to different prototypical discourse profiles for definite descriptions in different registers, namely, predominantly anaphoric uses in fiction and brand-new news items, but almost equally anaphoric as first-mention for editorials and recurrent news items, I am claiming that the discourse function of definite descriptions remains constant in all these registers.

It has long been a crucial claim of Accessibility theory that the first-mention—anaphoric distinction, or discourse profile is not the relevant distinction for the coding of Given entities (see especially Ariel 1988, 1994, 1996). What matters is the discourse function(s) of the referring expressions (defined in relative degrees of accessibility). It is the degree of accessibility with which the representation is currently maintained that determines the referring expression selected to evoke it. Anaphoricity is not equivalent
to high accessibility, even though anaphoric references prototypically require a relatively high degree of accessibility, and are therefore often performed by pronouns. And first-mention of Given entities does not automatically entail a very low degree of accessibility, although of course, many first-mention definite descriptions do code discourse entities which are currently entertained with only a low degree of accessibility.

Findings in Sanford and Garrod (1981) illustrate this point rather well. In one experiment, subjects read sentences which contained a definite description (e.g., the lawyer). This definite description presented either a new or a second mention entity, and it occurred in a context in which the entity is easily or not easily inferable (in texts entitled "In court" versus "Telling a lie" respectively). Now, if the anaphoric versus subsequent mention is the crucial distinction, we would expect reading times to be distinct for these two conditions. They are indeed different in the nonpredictive context ("Telling a lie"). However, reading times for first versus subsequent mention are not significantly different in the predictive context ("In court"). Crucially, subjects took longer reading the sentence with the anaphoric reference in the nonpredictive context than the first-mention in the predictive context. In other words, predictable first-mention entities are more accessible than second-mention unpredictable entities. Along the same lines, Yekovich and Walker (1986) found that false positive responses to probes not presented in the experimental materials depended on how script-based the concepts were. In other words, having been mentioned does not automatically guarantee a high degree of accessibility, and being a first-mention does not automatically entail a very low degree of accessibility. High inferability means a relatively high degree of accessibility despite the newness. All in all, a previous mention is just one factor contributing towards a higher degree of accessibility. It does not exhaust it.

In addition, had the anaphoric-first-mention distinction been a crucial linguistic distinction, one would expect that different referring expressions would specialize for one or the other use. This is not the case for the most part (see Ariel 1988 and onwards). Definite descriptions and names can be and are discourse subsequent-mention sometimes, and even pronouns can be used to initially retrieve a discourse entity. This is why whereas one can certainly characterize the prototypical discourse profiles of definite descriptions in terms of first-mention versus discourse anaphoric, a better account for their use is provided by their discourse function—marking relatively low degrees of mental accessibility. Whether discourse anaphoric or first-mention, definite descriptions (of various kinds) retrieve mental entities entertained at (a variety of) low degrees of accessibility (for distinctions among different definite descriptions see Ariel 1990, 1996, Table 1, above, and below).

Consider anaphoric references first. If one looks at anaphoric interpretations as retrievals of entities which vary in the degree of their accessibility to the addressee, one would expect different referring expressions to specialize for different cases of anaphora, specifically, for contexts of low versus high degrees of accessibility. Indeed, whereas 81.3% of the anaphoric pronouns in the data quoted in Ariel (1990): 18 referred to antecedents mentioned in the same or the previous sentence, 83.1% of the anaphoric defin-
ite descriptions referred to antecedents mentioned at least two sentences away, or even across the paragraph. Taking the anaphoric/first-mention criterion as distinguishing between referring expressions cannot distinguish between definite descriptions and pronouns, since obviously both are used (also) anaphorically. This, however, does not mean that they are interchangeable. Accessibility theory can distinguish between them, because the two expressions differ in the degree of accessibility they code.

Next, consider first-mentions of Given NPs. Again, if the anaphoric-first-mention distinction were the crucial one, then the results presented in Ariel (1996) would seem accidental. In Ariel (1996) I argued that first-mentions do not constitute a single category: they too are entities entertained at different degrees of accessibility. The forms used to introduce them vary, and again, it is degree of accessibility which determines their selection. Thus, definite descriptions containing one content word (relatively high accessibility) retrieved first-mentions which were relatively more accessible (84.6%): These were either “permanent generics”, such as the public, or frame induced, i.e., easily inferable from the specific context (e.g., ‘guests’ at a party). On the other hand, definite NPs which contained relative clauses averaged 6.8 content words (very low accessibility markers), and retrieved first-mentions which were not so easily accessible to the addressee (82.3%): Entities which required inferencing based on general knowledge not directly represented in the discourse itself, or entities normally stored differently in the addressee’s memory (for details see Ariel 1996).

Thus, I believe that for the form and function(s) of definite descriptions, it is degree of mental accessibility which is the crucial factor, and not the first-mention/discourse anaphoric distinction, nor the discourse profile. Both first-mentions and discourse anaphoric entities manifest a variety of degrees of mental accessibility. From a discourse function point of view, it is therefore not at all important to decide whether the prototypical discourse profile for definite descriptions is as Fraurud (1992) claims it to be (first-mention), or as my data has shown it to be (discourse anaphoric). It is the discourse function they code which determines their proper use, both when they are first-mentions and when they are discourse anaphoric. This, I have argued, is a relatively low degree of accessibility.

3.2 Register differences for definite descriptions vs. pronouns

In section 3.2 we look at register differences regarding the use of definite descriptions and pronouns. We examine both the question of frequency and of referential distance. Fox (1987) found a difference in the frequency of definite descriptions and pronouns between conversations and written texts. Whereas the number of pronouns and lexical NPs is almost equal in the written texts (53% pronouns, 47% lexical NPs), there are about 3.5 pronouns per one lexical NP in the spoken data (78% pronouns, 22% lexical NPs). In addition, whereas the referential distance between a previous and a current mention of a pronoun antecedent is 2.5 clauses in her spoken data, it is only 1.2 clauses
in the written data. Fox then assumes that there are different patterns of anaphora for different genres.

Toole (1996) convincingly argued against Fox’s (1987) conclusions. While she too found different discourse profiles for referring expressions in the same genres examined by Fox (Toole’s data is Australian English, however), she nonetheless concluded that the same (Accessibility) discourse functions characterize the use of both pronouns and definite lexical NPs in all the genres examined. By taking degree of accessibility as a complex concept, she was able to prove that full NPs retrieve entities of a low degree of accessibility, and pronouns retrieve mental representations entertained at a high degree of accessibility, regardless of the discourse profiles they participate in. Rather than distance alone (Fox’s measure), Toole combined a few measurements of degree of accessibility, distance being one of them, but also competition and the number of previous mentions. The result was that while the statistical frequency of different referring expressions in different genres was indeed different, the same accessibility discourse function could account for all the findings: the correlation between degree of accessibility and choice of referring expression was significant across all genres at the .01 level.

Toole also examined each case which seemed to pose an exception to the assumption of an accessibility discourse function (i.e., full NPs in high accessibility contexts, pronouns in low accessibility contexts). For example, Fox (1987) found a large number of definite descriptions retrieving discourse entities previously mentioned in the previous clause (a high accessibility context in terms of distance): 38% and 11% of the definite descriptions in the written and conversational data respectively. Toole examined all such cases in her own data. She found that she could explain the majority of these cases within Accessibility theory (many of them were clarifications, for example) in that what seemed to be a high accessibility context turned out to be a low accessibility context after all. The same applies to cases where pronouns were used when the context seemed to imply a low degree of accessibility. In other words, what Toole’s research demonstrates is that when the discourse function of some form is properly defined (in this case, degree of accessibility carefully evaluated), significantly different discourse profiles can be shown to follow from the very same discourse functions.

Most intriguing is a comparison of Fox’s results with the findings in Biber et al. (1999). While they agree about the predominance of pronouns in conversations (80% of the anaphoric expressions are personal pronouns in their data), Biber et al.’s counts of personal pronouns in newspapers and academic writings (similar texts to Fox’s choice of written English) show quite a different ratio for pronouns vs. lexical NP anaphors. Their corpus shows fewer pronouns (30% on average, as compared with Fox’s 53%). In addition, whereas Fox finds that the referential distance for pronouns is twice as large in conversations than in her written materials, Biber et al. find the opposite. The referential distance for pronouns is larger in their written materials than in their conversations (by 33.3%). Note, however, that whereas Fox counted clauses, Biber et al. counted number of words. They find that the average distance in terms of words is 15 for conversations and 20 for all their written sources. Comparing Biber et al.’s counts for definite
descriptions, the average referential distance for conversations is 32.5 words, and 36.25 (11.5% more) in the comparable written materials. These differences are not nearly as impressive as the differences noted by Fox (not to mention the fact that one of them goes in the opposite direction). I propose that the different results for spoken vs. written registers are mainly due to the well known difference in length between spoken and written clauses. Written clauses are significantly longer (see e.g. Chafe 1982), and this may be why in terms of clauses, the referential distance for written English pronouns seems so much shorter to Fox. Apparently, written and spoken clauses don’t constitute comparable accessibility factors because of the differences in length and complexity.

However, it is not at all my goal to doubt the generalizability of Fox’s data. In fact, she and Biber et al. did not really even count precisely the same forms: I’m not clear on whether Biber et al. counted anaphoric names, for example. Fox only looked at third-person human references. My point is that different registers may very well have different goals and therefore different entities to introduce, and hence significantly different ratios of referring expressions. Unlike Fox, however, I maintain that the principle governing their use remains constant. To see this, compare Tables 3 and 4, adapted from Biber et al. Note how different the proportions of pronouns and lexical anaphors are across the different registers (Table 3), but at the same time how rather similar the referential distance is per referring expression across the very same registers (Table 4).

Note that pronouns vary between being rather marginal in academic writings (20%) to being predominant in conversations (80%, a 4:1 ratio). Repeated (identical) definite descriptions vary between 5% for conversation and 40% for academic writings (a ratio of 8:1), and synonymous definite descriptions have an approximate 10:1 ratio (I estimate) in newspapers vs. conversations. These are no doubt highly impressive differences. Now, contrast these large differences in Table 3 with the much more mild differences in Table 4.

Table 3. The percentage of various expressions out of all anaphoric expressions (L.T.2.5% (Less than 2.5%)

<table>
<thead>
<tr>
<th>Ref. exp.</th>
<th>Conversation</th>
<th>Fiction</th>
<th>News</th>
<th>Academic writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronoun</td>
<td>80%</td>
<td>75%</td>
<td>40%</td>
<td>20%</td>
</tr>
<tr>
<td>Def., repeated</td>
<td>5%</td>
<td>10%</td>
<td>35%</td>
<td>40%</td>
</tr>
<tr>
<td>Def., synonym</td>
<td>L.T.2.5%</td>
<td>5%</td>
<td>20%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Adapted from Biber et al’s Table 4.2, p. 237

Table 4. The referential distance for various referring expressions (in number of words)

<table>
<thead>
<tr>
<th>Ref. exp.</th>
<th>Conversation</th>
<th>Fiction</th>
<th>News</th>
<th>Academic writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronoun</td>
<td>15</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Def., repeated</td>
<td>30</td>
<td>45</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Def., synonym</td>
<td>35</td>
<td>35</td>
<td>30</td>
<td>35</td>
</tr>
</tbody>
</table>

Adapted from Biber et al’s Table 4.3, p. 239
In Table 4 the difference between the referential distances of pronouns in different discourse varieties is quite smaller (33.3% more for the written genres), repeated definite descriptions have a maximal 50% difference (between conversation and fiction), and synonymous definite descriptions have a rather marginal difference between the newspapers and the three other genres (16.7%). Thus, in terms of types of preferred referring expressions, there is great variability between the different sources (the gaps in Table 3 vary between 300% and 900%), but in terms of referential distance, the differences vary between 16.7% and 50%. These findings attest that different registers have clear preferences for certain types of entities: conversations prototypically involve high-accessibility entities, whereas academic writings tend to involve low-accessibility entities. But the conditions under which pronouns and definite descriptions are used are quite similar, despite the very large gap in their frequency. The referential distance remains quite constant across genres per referring expression.6

Definites are not only contrasted with pronouns. They can also be compared to indefinites. These ratios are also genre-sensitive. Assayag (1999) finds that first-person narratives and expository texts (both spoken monologues) differ in their definite versus indefinite subject NP ratios. The former introduce (proportionately) many more definite than indefinite NPs (61 vs. 25, 2.4 times more), whereas the latter introduce more indefinite than definite NP subjects (76 vs. 48, 1.6 times more). This is a clear register difference, which is dictated by the differential nature of the entities referred to in the two registers. In the narratives, speakers talked about the actions of the protagonists, one of these being the speaker herself. These entities are concrete, very often human, and there are only a few of them. They can easily become even highly accessible, therefore. In contrast, the expository text contained complex entities, very often abstract and generic concepts, and there were many of them. Hence, there were more first-mentions in the expository texts. For example, the concept of violence, which repeatedly appeared in the texts (it was the topic the subjects were asked to discuss), remained indefinite even when it was not a first-mention. Given the difference in the entities referred to, it is not surprising to find that the referring expressions employed differ in the two registers.

Finally, note that statistically significant genre differences may result from the (intended) violation of Accessibility Theory/The Givenness Hierarchy predictions in one but not in another genre, thus creating a genre distinction. Nesher (2002) examined the first 50 references to first-mention human characters in stream of consciousness novels (SOC, henceforth) and in detective novels in English and Hebrew novels by male and female authors. The expectation is that first-mention entities will be introduced either by indefinites (because the entities are inaccessible) or by low accessibility markers (because the entities are inferable based on general knowledge), but not so much by high accessibility markers (demonstrative and personal pronouns). The detective novels fulfill these predictions: 39% of the entities (78) are introduced by some indefinite NP,

6. And I maintain that most of the differences present in Table 4 will evaporate once degree of accessibility is calculated by reference to a multiplicity of accessibility factors.
55% (110) are introduced by names and definite descriptions (low accessibility markers), and only 4.5% (9) are introduced by high accessibility markers. The SOC novels, however, show the same rate only for low accessibility marking 56% (112). It has fewer indefinites (29%, 58), but more high accessibility markers (15%, 30). Incidentally, the rate of definite descriptions is quite similar in these two very different genres: 25% in the detective novels and 29.5% in the stream of consciousness novels.

Now, rather than conclude that detective stories and SOC novels have different conventions for using pronouns and indefinite NPs, Nesher concludes that the discourse function of all referring expressions remains constant across the two genres. In SOC novels, however, there is another factor involved, which is in conflict with Accessibility theory. The author in such novels is supposed to directly reflect one’s consciousness (Prince 1987). In other words, the writer of such novels pretends to be a less cooperative addressee, taking the addressee’s state of mind less into consideration. Given these goals of the genre, it is not surprising that SOC novels show fewer indefinites (the entities introduced are not accessible to the addressee, but they are, of course, accessible to the writer) and more pronouns (the entities introduced are not highly accessible to the addressee, but they are so to the writer). In other words, SOC novels show a systematic violation of Accessibility Theory (as well as the Givenness Hierarchy) for special pragmatic effects. Indeed, both theories leave room for such violations. But the main point is that once again, we do not need to directly connect between referring expressions and genre. There are no special conventions for the use of referring expressions in SOC novels. The conventions are the same, but they are violated at higher rates due to the special goals of the genre.

Indeed, other findings point to the same “uncooperative” direction for SOC novels. Nesher found that the detective novels introduced 3.6 new characters per page on average, whereas SOCs introduced 4.8 new characters (30.7% more). She also examined the syntactic position used to introduce the new character. As is well known, subject position tends to be reserved for continuing discourse topics. Indeed, whereas 22.5% of the new entities were introduced in subject position in the detective novels 35% of the entities were introduced in subject position in the SOCs (55% more). Finally, a majority of the “uncooperative” high accessibility markers (62.5%) in the detective novels were anchored, i.e., they were inferable based on a previously mentioned salient entity. This was only true for 10.7% of the counterpart cases in the SOCs (5.8 times less). SOC novels then consistently violate Accessibility Theory in order to create an impression that it is the writer’s and not the addressee’s consciousness that lies behind referential choice.

For another register-related consistent violation of Accessibility theory for a special effect see Kronrod and Engel (2001), who found that newspaper headlines use too high accessibility markers (intermediate, rather than low accessibility ones) when first introducing entities, in order to both save on space and arouse readers’ curiosity.

Finally, Kumpf (2003) finds a consistent register-dependent counterexample to the expectation that Given entities be nonlexical (her claims are couched within Chafe’s (1994) theory). More than a third of the Given entities in her data were coded by lex-
tical NPs (185/545, 33.9%). This is a surprisingly high rate, and Kumpf proposes that it is the register she is examining which is responsible for this exceptional rate. Kumpf’s data comes from high school science classroom discourse. Indeed, the great majority of these lexical NPs refer to science terms. Kumpf reasons that teachers tend to repeat the science terms they introduce “to make sure that the full referent is available to all students”, (p. 121). Repeated lexical NPs help teachers maintain the attention of their students, a primary goal for a teacher facing some 30 high schoolers.

Given the same set of form-function correlations (be they Accessibility Theory or Givenness Hierarchy ones), the linguistic forms actually selected by addressees may well be consistently different across registers if the functions to be coded (the type of discourse entities) are consistently different. In each case we have seen that it is the difference in the nature of the entities referred to, or else, the special goals of addressees which are responsible for the selection of different forms. And while the nature of the entities is directly dictated by the type of register, there is no direct, conventional association between the specific register and the forms frequently figuring in it. In each register, it is the same Accessibility Theory/Givenness Hierarchy principles which mediate between the register expectations (re entities) and the resulting linguistic expressions (types of referring expressions). Thus, the different frequencies with which definite descriptions are used in different registers can directly be accounted for by their conventional discourse function, and no appeal need be made to the register they occur in.

3.3 Register differences re referential distance of definite descriptions

We have seen above that the referential distance of definite descriptions is consistently larger than that of pronouns. This is only to be expected if definite descriptions are low accessibility markers and pronouns are high accessibility markers. Recall further that the referential distance for definite descriptions was not too variable in different registers in Biber et al.’s data (see again Table 4). As a last example of potentially different discourse profiles for definite descriptions in different registers, consider the findings in Table 5 regarding the distance between the last and the current mention of the discourse entity for anaphoric definite descriptions. The first three sources were quoted in Ariel (1990:42), the fourth presents the results for Morris (1994), which we here focused on. Concentrating on magazine article I and the short story, one could argue for a register difference between newspaper articles and short stories. Thus, whereas same

<table>
<thead>
<tr>
<th>Source</th>
<th>Same paragraph</th>
<th>Previous paragraphs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magazine article I</td>
<td>36 (50%)</td>
<td>36 (50%)</td>
</tr>
<tr>
<td>Magazine article II</td>
<td>35 (90%)</td>
<td>5 (10%)</td>
</tr>
<tr>
<td>Short story</td>
<td>27 (64.3%)</td>
<td>15 (35.7%)</td>
</tr>
<tr>
<td>Edited narrative</td>
<td>100 (53.8%)</td>
<td>86 (46.2%)</td>
</tr>
</tbody>
</table>
paragraph antecedents are equally frequent as previous paragraph antecedents for the magazine article, there are almost two same-paragraph antecedents per one previous-paragraph antecedent in the short story. However, once we also consider magazine article II and the narrative, the register generalization evaporates: The edited spoken narrative resembles magazine article I (rather than the short story), but the latter does not resemble magazine article II, where same-paragraph antecedents are the overwhelming majority.

While the discourse profiles represented in Table 5 are quite different from each other, just like for first vs. subsequent mention, distance alone does not determine referential form. Rather, it is degree of accessibility as a complex notion. In order to see this let us examine more carefully the data in Morris (1994). Such an examination reveals that the differences re distance from the antecedent are not by themselves the key to understanding the proper conditions placed on the use of definite descriptions. It is the cases of same-paragraph antecedents (100 cases) which seem puzzling under an assumption that definite descriptions code a low degree of accessibility, because of the apparently short distance from the antecedent. These cases break down to at least three types, only one of which is a potential problem for the assumption that definite descriptions mark a low degree of accessibility. We examine them all.

First, 29 (29%) of the definite descriptions which have a paragraph-internal antecedent refer to that antecedent in a different form than it was initially introduced (a), or else the coreference established is only partial (b), where the department refers to more than its faculty members. As such, their antecedents cannot really be said to be entertained at a high degree of accessibility, of course, despite the relatively short distance:

(5) a. At this time, something had just appeared in the Chronicle of Higher Education about ... So I called several faculty women mentioned in the article ... (134: 148).

b. Several faculty members wrote the dean ... the fact that he was disrupting the department (133: 97).

Second, 44 (44%) of the paragraph-internal definite descriptions actually constitute partially new discourse entities, since they are inferred based on a discourse entity previously mentioned (in the same paragraph in this case). As argued above, these cases are hybrids, and cannot be counted as cases constituting a very high degree of accessibility for the referent. Here is a typical example (and so is the judge in (4) above):

(6) They came to Greenwood, Mississippi, and opened a grocery store in the black section of town (129: 9).

Last, 27 definite descriptions (27%) retrieve entities previously mentioned in the same paragraph in roughly the same form as they were previously mentioned. These

7. See Lord and Dahlgren (1997: 337) for a similar finding. And see the different referential distances for repeated vs. synonymous definite descriptions in Biber et al.'s data quoted in Table 4.
should be examined individually, for they constitute a potential counterexample to the claim that all definite descriptions in all registers retrieve mental representations of a rather low degree of accessibility. First, it is noteworthy that all of them were quite short and therefore do not constitute very low accessibility markers: 23 (85.2%) contained one content word, and four (14.8%) contained two content words. Second, while they all retrieve entities previously mentioned in the same paragraph, for many of them (12, 44.4%) the antecedent appears at least four clauses away (the average distance being seven clauses back). Others are used for clarification, due to competing antecedents or because of a combination of accessibility factors (e.g., distance + competition). One is an intended violation for the generation of a Gricean implicature. All in all, three examples seem not to be subsumed under the standard Accessibility explanations. One of them is (2) above, where we still see that the second mention was shortened. Note also that the second mention occurs in a sentence which steps back in time to provide a background for the first-mention sentence (a lower cohesion between the sentences reduces the accessibility of the antecedent—see Ariel 1990: ch. 7, and references therein). Finally, two of these three cases (including 2) refer to very transient discourse entities, and this is their second and last mention. Their antecedents, although close by, do not constitute salient (i.e., highly accessible) discourse entities, as is also attested by their non-subject role. The third case is quoted below:

(7) Privately, we were told that the university, didn’t want to pay … so they, were trying … They, said …, because however small the chances of their, winning, there was that chance and I could lose everything. So the university, … (138:306).

Note that the last reference to the university switches to a definite description after a chain of pronominal references to it. I believe speakers switch to a low accessibility marker from time to time to reassure their addressees that they have the right referent in mind.8

Thus, once we calculate degree of accessibility based on a number of factors, rather than on the basis of distance alone, certainly when distance is measured in terms of paragraphs, we see that virtually all definite descriptions refer to discourse entities entertained at (varying) degrees of low accessibility. This relatively low degree of accessibility can be due to the nature of the anaphor, the intended referent, or the relation between them. If the anaphor is not a repeated NP, or if the intended referent is not identical in reference to the antecedent (partial coreference, inferred reference), the degree of accessibility is lowered. Finally, even within a paragraph, the referential distance may be large, creating a relatively low degree of accessibility, especially if there are other inter-

8. Also, according to Gernsbacher and Shroyer (1989), reference by a low accessibility marker boosts the accessibility of the entity for future references. So some low accessibility markers are chosen because of their cataphoric contribution rather than the anaphoric history of the discourse entity involved.
venering referents. Section 3 has argued that definite descriptions manifest a variety of discourse profiles, if discourse profile is characterized either by reference to first versus subsequent mention (even if we adopt a single, loose definition for discourse anaphora, 3.1), by frequency (relative to pronouns, 3.2), or by reference to the referential distance (for anaphoric definite descriptions, 3.3). If, however, we analyze definite descriptions as low accessibility markers, the different discourse profiles do not translate to register-dependent use conditions. Rather, the different discourse profiles all result from the application of one and the same discourse function, which happens to produce different discourse profiles under different circumstances. These differences are only apparent. First versus subsequent mention and short versus long referential distance demonstrate that accessibility is simply not reducible to single criteria, and register differences only show that the amount of presupposed entities the speaker and addressee share initially consistently varies from one register to another. These different circumstances do affect discourse profiles, but they do not necessitate our positing different discourse functions for definite descriptions in different registers.\(^9\) Surface register differences in the use of referring expressions can therefore be compatible with the assumption of a single discourse function, because the register differences are only indirectly correlated with referential style.

4. Conventional register differences for referring expressions

Biber (1995: 10) notes that associations between linguistic expressions and situations may be either functionally motivated, or they may be conventional. I have argued that the cases we have reviewed so far have all been cases where the statistically significant differences among registers are functional, and do not entail register-specific conventions for referring expressions. This does not mean that all such differences are register-insensitive.

Comparing between the findings for possessive NPs in the same narrative (Morris 1994—see Ariel 2002) and definite descriptions (section 3), I see a potentially different role for the prototypical discourse profile I presented for each. Possessive NPs across languages do show the relevance of their prototypical discourse profiles in their grammaticization patterns. To the best of my knowledge, definite descriptions (and other referring expressions) do not tend to grammaticize in a fashion compatible with their prototypical discourse profiles, as these have been described in the literature above. Referring expressions do not specialize for first versus subsequent mentions, nor for

---

\(^9\) Note that although Lord and Dahlgren (1997) argue for genre differences for referential form and usage, they actually provide explanations similar to mine for the apparent different referential styles, noting the different characteristics of various genres (e.g., the fact that unlike newspaper articles, conversations do not very often have one global discourse topic).
shorter versus longer referential distances per se. Although some languages have definite articles which are apparently restricted to anaphora versus general knowledge, the “anaphoric” articles can refer to situational and to inferred entities (both first-mention according to Fraurud 1992), and the “general knowledge” articles can refer anaphorically as well (Lyons 1999). Section 4, however, briefly surveys a few cases where I believe there to be conventional associations between referring expressions and registers. It is beyond the scope of this article to examine the fascinating question of which discourse profiles are prone to lead to grammaticization and which are not. It would seem to me that register-related distributional patterns are less prone to grammaticization than cross-register discourse profiles, but this question must await further research.

In Ariel (1983), I have discussed the discourse function of one specific apposition in Hebrew (mi she . . . ‘who that . . . ’), when adjacent to a proper name, as in:

(8) doctor henri kisinger, mi she+ haya sar ha+xuc shel Dr Henry Kissinger, who that was the minister of foreign affairs of arcot ha+brit, hibia etmol et daato . . . the United States, expressed yesterday acc his opinion (Yedioth Ahronot, 11.2.1978) ‘Dr Henry Kissinger, former US secretary of state, expressed his opinion yesterday . . . ’

I argued that this construction is reserved for first-mention V.I.Ps. Moreover, the construction is restricted to journalistic Hebrew. We here have an example of a referring expression which is restricted (a) to first-mention, and (b) to a specific register.

Clancy (1982), analyzing referential forms in spoken and written Japanese narratives, also finds a few differences which seem conventional, rather than pragmatically motivated (although, some differences she does explain as deriving from the different circumstances obtaining in speaking versus writing). For example, put into Accessibility Theory terminology, second mention entities don’t yet merit reference by high accessibility markers (zeros or pronouns) in both registers. However, whereas written Japanese tended to repeat the informative low accessibility marker used for the first mention, spoken Japanese preferred a shorter, less informative lexical NP (e.g., sono hito ’that person’ in second mention, as opposed to the lengthier otoko no hito ‘male person’). Another distinction noted by Clancy is that third-person pronouns were never used in the spoken narrative. The reason is that in conversation (and only in conversation) personal pronouns “imply a personal relationship between the speaker and the referent” (p. 64).

Another type of case where a linguistic form may be register-specific is due to the gradual nature of linguistic change. Most linguistic changes occur in spoken discourse, and hence, innovations may occur with a higher, even absolute frequency in informal

---

10. Longacre (1979) is the only one who has argued that some languages have pronouns which are restricted to paragraph scope.
talk. Biber et al. mention the semi-modal verbs (e.g., *have got to*) which are more common in conversations in this connection. I have noted that spoken Hebrew first/second person future verbs only marginally permit zero subjects (0% in one conversation, 11.3% in another—see Ariel (1990, 2000) for statistics). This is not the case for written Hebrew, where the majority of such cases did have a zero subject (76.5%). As I argued in Ariel (1990, 2000), future tense verbal agreement markers are being reanalyzed as nonreferential, hence the need for an overt (subject) referring expression. This reanalysis is primarily felt in the spoken register. In this case, then, we can say that written Hebrew has a high accessibility referring expression which spoken Hebrew no longer has, namely, the person verbal agreement in future tense.

There are, then, cases where a referring expression is restricted to a specific register in a conventional manner. These are cases where the differential distribution cannot be explained on the basis of the extralinguistic, communicative goals of the addressees in the specific register. Note that we cannot say that Hebrew conversationalists never introduce a V.I.P. into their discourse. They most certainly do. When they do, however, they simply lack the conventional means to indicate the V.I.P. ness of the character conventionally. They make do with other means available for first mention entities, namely, very informative low accessibility markers. Similarly, it’s not the case that conversationalists cannot repeat full NPs, nor that they only have a personal relation to referents in written Japanese. Finally, it is not the case that speakers and addressees (the referents of first- and second-person verbal forms) are consistently less accessible or less frequent in spoken Hebrew than in written Hebrew. If anything, we would expect the opposite, namely that the speaker and addressee should be even more accessible in face-to-face conversations. The difference between spoken and written Hebrew (re the person agreement markers in future tense) must, then, be grammatically specified in this case. First- and second-person verbal agreement markers are a (very high accessibility) referring expression in written Hebrew, but are nonreferential agreement markers in spoken Hebrew. The register differences briefly discussed in section 4 all necessitate specific conventions which are register-dependent. This is not the case for the variety of differences noted in sections 1–3.

5. Conclusions

Register differences are real and pervasive. I would argue, however, that for the most part, such differences mainly concern addressees’ goals, and only indirectly manifest themselves formally. We therefore need not assume separate grammars for separate registers (in the majority of the cases). We have here found confirmation for Swales’ 1990 argument that genre cannot be reduced to statistical counting of formulas. Genres define communicative events, each with its recurrent communicative purposes. It is the goals and rationale of the discourse which informs the choice of content and style. Naturally, these have implications for the selection and use of linguistic expres-
sions because the latter are associated with certain discourse functions, but it does not usually justify the assumption of conventional associations between linguistic expressions and specific registers. The specific register distributional patterns fall out naturally from one and the same grammar because grammar is built to serve many purposes. If we assume a grammatical mechanism whereby linguistic forms are selected by reference to the function intended by the speaker, then the same set of conventions will yield the differential distributional patterns, since we can trust the addressee to let her communicative goals guide her linguistic choices among the alternatives offered by the grammar.

In a more philosophical tone, one can say that languages simply cannot afford to have a grammar in every register. It is therefore the cases where we do find form-register conventional associations which are in need of explanation. In fact, one may wonder what is to be considered a discourse profile. Surely had we characterized the discourse profile of definite descriptions by reference to degree of accessibility, I have argued that we would have come up with a much more homogenous pattern for all the registers (provided we controlled for accessibility differences among different definite descriptions). Clearly, we have to follow the actual language user here. It seems that some discourse profiles carry psychological reality (those discussed in section 4), while others (those discussed in sections 1–3) do not. Future research should address itself to the intriguing question of which discourse profiles may acquire a distinct salient status, such that might justify a distinct grammatical convention, and which discourse profiles may not trigger register-distinct grammars.

References


MA thesis submitted to Tel Aviv University.


*Longman Grammar of Spoken and Written English*. Essex: Longman.

Chafe, Wallace. 1982. "Integration and Involvement in Speaking, Writing, and Oral Literature". 
In *Spoken and Written Language: Exploring Orality and Literacy*, Tannen, Deborah (ed.). 


Ablex. 55–76.


Fowler, Carol A. and Housum, Jonathan. 1987. "Talkers' Signaling of "New" and "Old" words in 
speech and listeners' perception and use of the distinction". *Journal of Memory and Language* (26), 489–504.

Cambridge: Cambridge University Press.


Amsterdam: John Benjamins.


