A variationist account of constituent ordering in presentative sentences in Belgian Dutch

STEFAN GRONDELAERS and DIRK SPEELMAN

Abstract

This paper reports on a corpus-based analysis of constituent ordering in presentative er-*constructions in Belgian Dutch. Whereas in the majority of these sentences, the locative typically follows the indefinite subject (cf.* Er ligt een bompakket op de zesde verdieping *'There is a bomb on the sixth floor'), in a small number the locative precedes the subject, as in* Er zijn in Brussel geen getto's *'There are in Brussels no ghettoes'. In order to account for this hitherto unnoticed variation, we extracted 360* er-*initial sentences with one locative adjunct (either in final or penultimate position) from a corpus of written and spoken Dutch, and coded them for eight semantic, syntactic, and pragmatic variables. A stepwise logistic regression confirmed our hypothesis that* er*'s inaccessibility marking function (Grondelaers, Brysbaert, Speelman and Geeraerts 2002) is a factor which determines word order variation. At the same time, however, the regression analysis demonstrated that there is a more significant ordering motivation, viz. informational prominence. The finding that it is the more prominent constituent which tends to be sentence-final throws new light on the given-before-new principle (Gundel 1988), and rules out any analysis which restricts the constructional meaning of presentative sentences to "introducing an indefinite subject".*

Keywords: presentative sentences; word order variation; discourse analysis; variationist approach; regression analysis.

1. Introduction

This paper reports on the third phase of a large-scale variationist investigation into the constructional meaning of presentative sentences and presentative *er* 'there' in Modern Dutch. Examples (1)–(4) illustrate the standard type of presentative sentence in Dutch, which is to a high extent

isomorphous with English *there*-sentences. The sentence-initial *er* in this construction type has been dubbed "plaatsonderwerp" (topical subject), "repletive *er*", "presentative *er*" (Haeseryn et al. 1997: 464, 467), and "existential *er*":

(1) *Er bevonden zich geen paparazzi*
     Er were   no paparazzi
     [vlak voor, naast of achter de wagen]$_{LOC}$.
     right in front of, next to, or behind the car.
     'There were no paparazzi right in front of, next to, or behind the car.'
     (0.2571)

(2) *Er zijn [op dit moment]*$_{TEMP}$ *11.500 invalide zelfstandigen.*
     Er are at this moment   11.500 disabled businessmen
     'There are 11.500 disabled businessmen at this moment.'
     (0.1703)

(3) *Er is   [in mijn leven]*$_{LOC}$ *iets belangrijks   gebeurd.*
     Er has in my life   something important happened
     'Something important has happened in my life.'
     (0.0802)

(4) *Er zal verkeershinder   zijn [tijdens de spits]*$_{TEMP}$ *.*
     Er will traffic problems be   during the rush hour.
     'There will be traffic problems during the rush hour.'
     (0.0213)

In Dutch *er*-initial presentatives, locative adjuncts (LOC) typically follow the subject, whereas temporal adjuncts (TEMP) characteristically precede it. Alternative orderings as in (3) and (4) occur, but they represent the marked option, as indicated by the typicality rates[1]. Subject-preceding locative adjuncts as in (3) typically denote abstract or wide-scope locations (frequently referred to by toponyms), which are moreover characteristically coded by means of short NP's.

   A principal ambition of this paper is to demonstrate that word order variation in presentative sentences can be accounted for and predicted to a considerable extent. Our main objective, however, is to establish that the variationist methodology we use for that purpose can be fruitfully extended from sociolinguistics (the linguistic paradigm in which it is predominantly used) to areas of linguistic research in which it is not common, such as functionalist syntax and discourse analysis. Our research into presentative sentences will be shown to take on a *helix-like* structure of a gradual refinement of interpretations through the repeated

confrontation with variationist empirical data: a first set of multivariate analyses of the factors which determine the syntax of presentative sentences revealed discourse functions for these constructions, which are further tested and refined in this paper in a new series of multivariate analyses.

To be sure, variationist concern has rarely been a primary focus of linguists interested in functional or pragmatic explanations (see Geeraerts 2005 on the cognitive linguistic neglect of contextual factors). Conversely, sociolinguists mostly steer clear of pragmatic factors when charting the relation between language varieties: pragmatic variables are not only difficult to identify and measure (compared to phonological and morphological variables), their use as sociolinguistic indicators is severely hindered by the interference of various types of semantic variation (for an overview, cf. Milroy and Gordon 2003: 169−197).

Yet, we will argue that pragmaticians and discourse analysts can benefit from sociolinguistic experience and techniques in three important ways. In addition to overcoming their general reluctance towards empirical substantiation and quantification (see Sanders 1997; Geeraerts 2005), pragmaticians must be persuaded that the functional explanations they put forward for languages as a whole are more often than not *restricted to regional and social* varieties of those languages (cf. Grondelaers, Geeraerts and Speelman, in press), as a result of which it is essential to keep an eye on interfering *external* variation (viz. regional, stylistic, age, sex variation etc.) when carrying out investigations pertaining to *language-internal* structure, viz. morphology, lexicon, and syntax[2]. Sociolinguistic investigations into lexical or syntactic variation typically control for internal as well as external factors, and the main sociolinguistic research tool − multivariate statistical analysis − is specifically tailored to that purpose: techniques such as logistic regression (Rietveld and Van Hout 2005) enable the researcher to chart the co-variation of internal and external causes of variability. A regression analysis not only yields significance and impact values which reflect the relative *importance* of all the factors included in the analysis; it also returns estimates which reflect the *explanatory and predictive power* of the global set of factors, allowing the researcher to fine-tune his model by adding additional internal or external factors if necessary. Third, and most importantly, the research in this paper goes beyond predicting and modelling the observed variation. By identifying the underlying factors which motivate constituent ordering in *er*-initial sentences, we seek to access the constructional meaning of these sentences, in order to find out whether they code the same pragmatic function as the adjunct-initial variant with postverbal *er* (illustrated in (5b)−(8b)), *viz.* decontextualizing an inaccessible subject[3].

The material will be organized as follows. In the next section, we review some traditional linguistic approaches to presentative sentences, as well as our earlier variationist investigations which constitute the ground work for the analysis in this paper. In Section 3 we outline a testable functional hypothesis for the constituent ordering variation observed in (1)−(4). Section 4 concentrates on the corpus used for the analysis as well as on the characteristics of the dataset extracted from it. In Section 5 we introduce, implement, and provisionally test the hypothesis outlined in Section 3, as well as three other acknowledged determinants of word order variation. Section 5 ends in two conflicting hypotheses concerning the discourse function of *er*-initial presentative sentences, which are evaluated on the basis of the regression data collected in Section 6. Section 7 summarizes all findings, and reviews their most important theoretical and methodological consequences.

2. Earlier analyses

Presentative sentences lend themselves very well to a variationist approach. To begin with, relatively little is known about them, because research has typically focused on a limited set of fiercely debated questions, to the detriment of other, equally important matters. The principal bone of contention concerns the question whether or not presentative *er* is inherently meaningless. In the structuralist tradition, *er* is a semantically empty word which occupies subject position when the actual, ontological subject follows the main verb (Bech 1952: 13; Paardekoper 1963: 34; Nieuwborg 1968: 285; more recently: Leys 1979: 244; Swiggers and Van den Eynde 1985). Early generative analyses (Kraak 1966; Verkuyl et al. 1974; De Haan 1974; De Haan et al. 1975; Pollmann 1975; Van den Toorn 1976) subscribe to the same view, albeit with the modification that *er* is inserted to keep the indefinite subject away from sentence-initial position where it could receive an unintended generic interpretation (compare in this respect *Er loopt een jongen* 'there walks a boy' with *Een jongen loopt* 'a boy walks'[4]). Among the more sophisticated transformational analyses of presentative *er* is Bennis (1986), who accounts for the sentence-initial presence of *er* on the basis of an "Empty Presupposition Principle", which stipulates that pragmatically well-formed sentences mandatorily contain an overt constituent which is presuppositional (1986: 225); *er*, in Bennis' view, fills the presupposition when no other element is available to do so.

In sharp contrast with structuralist and generativist thought is a "school" of functional analyses which hold that *er* is not a meaningless subject but a locative adverb denoting the "the spatio-temporal context or circumstances surrounding the event and its participants" (Kirsner

1979: 71; similar views can be found in Van Es and Van Caspel 1971: 79; Schermer-Vermeer 1985: 67; Blom 1992: 22). All these analyses argue, in addition, that *er* is used to introduce and focus attention on new and important information (see De Schutter 1974: 349; Elffers 1977: 420; Geerts et al. 1984: 395; Schermer-Vermeer 1985: 80 and note 24). Kirsner (1979), arguably the most comprehensive study ever devoted to *er*, challenges the view that *er* should be an "introducer", and claims instead that presentative sentences are "background-participant sentences" in which "the central participant in the event is (for whatever reason) less in the foreground of attention than it might be (1979: 161, 142)."

Examples (5)−(8) illustrate the adjunct-initial type of presentative sentence in Dutch. When an adjunct is topicalized in presentative sentences, *er* emerges postverbally, if it appears at all, for it has often been observed that it is in certain cases deleted in this context (see Haeseryn et al. 1997; De Rooij 1991). For reasons that will become clear below, separate typicality ratios[5] are given for Belgian (BEL) and Netherlandic (NET) adjunct-initial sentences:

(5) a. *[Vandaag]* TEMP *is ook een examen*[6].
         Today   is also an exam
         'Today there is also an exam.'
         (BEL = 0.0000; NET = 0.0000)

    b. *[In 1977]*TEMP *was* er *een fusie tussen Materne en Confilux.*
        In 1977   was *er* a merger between Materne and Confilux
        'In 1977 there was a merger between Materne and Confilux.'
        (BEL = 0.3658; NET = 0.2411)

(6) a. *[In 1963]* TEMP *startte een experimentele competitie.*
         In 1963   started an experimental competition
         'In 1963 there started an experimental competition.'
         (BEL = 0.1395; NET = 0.1189)

    b. *[Op die dag]*TEMP *komen* er *ook speciale acties.*
        On that day   come *er* also special actions
        'On that day there will also be special actions.'
        (BEL = 0.3224; NET = 0.2700)

(7) a. *[In Kabul]* LOC *zijn maar weinig buitenlanders.*
         In Kabul   are but few foreigners
         'In Kabul there are but few foreigners.'
         (BEL = 0.0279; NET = 0.0514)

    b. *[In Vlaanderen]* LOC *is* er *een echte warme gastvrijheid.*
        In Flanders   is *er* a real warm hospitality
        'In Flanders there is a real warm hospitality.'
        (BEL = 0.3255; NET = 0.0868)

(8) a. *[In het redactielokaal]*_{LOC} *staan enkele flessen wijn.*
In the newsroom   stand some bottles of wine
'In the newsroom there are some bottles of wine.'
(BEL = 1.116; NET = 0.7329)

b. *[In de woonkamer]* _{LOC} *ontstond* er *brand.*
In the living room   emerged *er* a fire
'In the living room (there) broke out a fire.'
(BEL = 0.1829; NET = 0.0643)

Er's complex postverbal distribution has given rise to a good deal of descriptive uncertainty in the linguistic literature. According to the *Algemene Nederlandse Spraakkunst* − the standard grammar of Dutch (Haeseryn et al. 1997) − no strict rules can be given for the presence or absence of postverbal *er*: "it can be optional, there may be semantic or stylistic differences involved, and there is a lot of individual and sometimes also regional variation in its use" (1997: 473). Only four reliable factors have hitherto been identified, mostly on intuitive grounds. Although these factors converge well with native speakers' judgements (and also, on the whole, with the typicality rates in (5)−(8)), it is equally clear that they interact in various ways which go far beyond human observation. It is, to begin with, well-known (cf. Van Es and Van Caspel 1971: 79−82; De Rooij 1991: 124) that the preference for *er* correlates negatively with the taxonomical specificity of the main verb: the more specific the verb, the less *er* (the typicality rates unanimously support this observation). In addition, *er* is undeniably less frequent in sentences with a fronted locative than in other cases (Bech 1952: 18; Van Es and Van Caspel 1971: 79; De Schutter 1974: 347; Haeseryn et al. 1997: 477), but according to the typicality rates, the impact of adjunct type is largely constrained to sentences with a taxonomically specific main verb: *er*-preferences are more or less comparable in (5) and (7) − except in Netherlandic clauses with *zijn* 'to be', in which *er* is far less frequent when preceded by a locative adjunct than is the case in Belgian clauses (compare Netherlandic and Belgian preferences in 5b and 7b) −, but they are markedly different in sentences (6) and (8). To complicate matters further, there is much more regional impact on *er's* distribution than the cited divergence between Belgian and Netherlandic usage in (7b): Belgian Dutch is known to be more tolerant *in globo* towards *er* than Netherlandic Dutch (cf. De Rooij 1991: 127; Haeseryn et al. 1997: 477), though this bias is more noticeable in clauses with a fronted locative than in clauses with a fronted temporal adjunct. Observe, finally, that *er* is attested significantly more often in *informal* register (Haeseryn et al. 1997: 477)[7].

In spite of its complexity, the postverbal distribution of *er* not only constitutes the strongest evidence against traditional approaches to *er*; it also represents an ideal starting-point for our variationist investigation into *er*. Observe, first, that the typicality rates for the b-variants of (5)−(8) invalidate all traditional approaches to *er*. Proponents of the dummy subject-approach have a hard time finding an answer to the question why *er* is attested so often in sentences with a grammatical subject and a subject position which is filled by an adjunct (a position, in other words, which does not need a dummy subject[8]). In the same vein, it is unclear how sentences with an initial locative should benefit from the "zooming out" to the very general situational reference allegedly coded by *er*. More particularly, if locative adjuncts are indeed a suitable "'handle' on an event which otherwise would be difficult to picture" (Kirsner 1979: 103−104), there is no reason why an even more suitable handle according to Kirsner − presentative *er* − should be realized postverbally when an appropriate handle is already available. Kirsner seems to be aware of this shortcoming when he observes that "while it is also possible for such 'fronted locatives' to introduce 'expletive' *er*, sentences such as *In de tuin blaft er een hond* 'In the garden there barks a dog' are quite rare in discourse". This observation, however, is not only contradicted by the typicality rates in (7b) and (8b), but also by the corpus data in Grondelaers, Speelman and Geeraerts (2002).

Adjunct-initial sentence pairs such as (5)−(8) represent an appropriate starting-point for our variationist alternative to existing work on *er*, for they have the crucial methodological advantage − first observed in Bolinger's thought provoking study on the distribution of *there* in English adjunct-initial presentatives (1977: 93) − that any syntactic, semantic or functional difference between variants with and without *er* can only be attributed to the presence or absence of *er*. By identifying such syntactic, semantic and functional differences between variants with and without *er*, one can "tease out the true sense of" *er* (Bolinger 1977: 93).

Confronted with all the distributional complexity and scant evidence, how can we proceed to investigate *er*'s distribution in an empirically more responsible way? In the first phase of our investigation (see Grondelaers 2000; Grondelaers, Speelman and Carbonez 2001; Grondelaers, Speelman and Geeraerts 2002, 2003), we reconsidered the factors reviewed in the paragraph following (5)−(8) in an empirically more dependable way, in order to find an answer to the following questions. First, are the four factors really significant determinants of *er*'s distribution? All of them are compatible with speaker judgements, but none of them constitute clear-cut rules. Second, what is their respective impact on *er*'s distribution? In view of the methodologically different status of the language-structural factors (ADJUNCT TYPE and VERBAL SPECIFICITY)

and the contextual factors (REGION and REGISTER), it is essential that we learn which have the greatest impact. Third, how much variation can be explained and predicted on the basis of these four factors? Are they already sufficiently powerful predictors, or do we need additional features? A regression analysis of the 1905 adjunct sentences from the Con-Div-corpus[9] (which were coded for the relevant variables) was subsequently carried out. This revealed, first, that in spite of the traditional "no strict rules" pessimism, *er*'s postverbal distribution could in 85% of all cases be predicted on the basis of the four variables, which turned out to be significant determinants of this distribution (albeit that the language-structural factors had a considerably higher impact than the contextual factors). More importantly, the regression analysis confirmed that there are outspoken proportional and structural differences between Netherlandic and Belgian models of *er* (see Grondelaers, Speelman and Geeraerts 2002 for an elaborate overview).

In the next phase of the research (see especially Grondelaers, Brysbaert, Speelman and Geeraerts 2002), we factored out all regional variation, and focused our exclusive attention on the search for additional factors which may drive *er*'s distribution in *Belgian* locative adjunct sentences (*er*'s distribution in temporal adjunct sentences is almost fully predictable). These additional factors no longer came from the linguistic literature on *er*, but represented corpus parameters of an encompassing functional hypothesis about the adjunct-initial sentence type and *er*'s role therein. Adjunct sentences, we argued, constitute a conventionalized means of inferential access into the upcoming unfamiliar subject referent. Following Altmann and Kamide (1999), we suggested that in order to facilitate the processing of that upcoming subject, comprehenders use lexical, syntactic, and world knowledge associated with the initial adjunct and the immediately following main verb to project *predictive inferences* with respect to the subject, as a result of which the latter can be activated well in advance of its actual realization. Compatibility between the projected and the intended subject, we hypothesized, would significantly enhance the latter's processing. Now, if this hypothesis is correct, locative adjuncts will be better subject predictors than temporal adjuncts, because they confine their subjects to a spatial setting which generates more useful inferences (concerning its concreteness, size, etc.). In the same vein, concrete locative adjuncts will be better subject anticipators than abstract locative adjuncts.

The finding that *er* is typically preferred in the context of what we claim to be "poor" subject predictors − temporal adjuncts and abstract locative adjuncts − leads us to the hypothesis that *er* is a "diagnostic indicator" of limited inferential access into the upcoming subject; *er*, we propose, advizes the hearer to block all ongoing inferencing so as not to

waste processing resources on a subject s/he can insufficiently anticipate. Following Ariel (1990), who argues that one of the main functions of natural language is to code a referent's degree of *accessibility*, we will provisionally refer to *er* as an *inaccessibility marker* (Bolinger 1977: 92 also mentions the absence of proper contextual anticipation of the sub-ject as a *there*-triggering factor). Unfortunately, the processing speed all these hypotheses draw on is not a phenomenon which can be measured *directly* in an offline context: in a corpus design, there is no dependent variable which allows us to determine whether the hearer needs more time for the processing of constituent *y* in context *a* than in context *b*. The self-paced reading (or subject-paced reading) technique we used in Grondelaers and Brysbaert (1996) and Grondelaers (2000), by contrast, offers reliable online indications of processing speed. A first experiment (Grondelaers and Brysbaert 1996; Grondelaers 2000) confirmed not only that subject referents are processed significantly faster in the context of concrete location adjuncts than in abstract location adjuncts; more importantly, *er* significantly enhanced the processing of subjects follow-ing abstract but *not* concrete locations. In a second experiment (Gronde-laers, Brysbaert, Speelman and Geeraerts 2002), we manipulated subject referents to find out whether contextually accessible subjects take longer to process than contextually inaccessible subjects, and whether *er* facili-tates the processing of the inaccessible materials. The analysis of the reading data confirmed both expectations[10]. Hence, all available data converge on the idea that *er* is an *inaccessibility marker* which deactivates context-based inferencing with respect to the subject, as a result of which the subject is *decontextualized*.

The present paper reports on the third phase of the investigation. In subsequent sections, the following questions will be addressed: can the inaccessibility theory of *er*'s postverbal distribution in adjunct-initial sen-tences be extended to presentative sentences in which *er* is mandatorily present in initial position? And do constructions types which always con-tain *er* inherit their constructional meaning from the function of *er*? In the variationist framework outlined in the previous paragraphs, these questions can be operationalized as follows: can the inaccessibility theory of *er* account for the constituent ordering variation observed in (1)−(4)? Conversely, does word order in *er*-initial presentative sentences reflect *er*'s function?

3. Inaccessibility theory and constituent ordering in *er*-initial presentative sentences

How do we proceed to find out whether *er*-initial sentences instantiate postverbal *er*'s inaccessibility marking and decontextualizing function?

We hypothesize that that function is perceptibly reflected in their syntax. The distribution of adjuncts over *er*-initial presentatives is crucial in this respect, so let us return to the constituent ordering patterns illustrated in (1)−(4). Building on corpus-based typicality rates, we found that locative adjuncts are characteristically sentence-final in Belgian Dutch *er*-initial sentences, whereas temporal adjuncts nearly always precede the subject. Now, if sentence-initial *er* is indeed an inference deactivator, it makes sense that concrete locative adjuncts − a major source of inferential access into the subject (cf. previous section) − should be in a position where they have *no impact* on the processing of the subject. It would indeed be ambiguous to announce *decontextualization* of the subject by means of *er*, and at the same time *recontextualize* the subject by means of a subject-preceding concrete locative. In the same vein, it is perfectly logical why temporal adjuncts and abstract location-denoting adjuncts − both of which were found to generate few useful inferences with respect to the upcoming subject − *can* precede the subject without creating an ambiguous signal.

In Section 5 we present corpus evidence in support of the impact of ADJUNCT CONCRETENESS on constituent ordering in presentative sentences; in addition, we introduce, operationalize, and provisionally test three other potential determinants of word order in *er*-initial sentences. Let us first, however, take a closer look at the materials on which the analysis is based.

4. Materials

This investigation builds on three components of the ConDiv corpus − an extensive electronic corpus of written Dutch compiled for the investigation of linguistic variation (cf. Grondelaers, Deygers, Van Aken, Van Den Heede and Speelman 2000) − and Release 5 of the newly compiled *Corpus Gesproken Nederlands* (Corpus of Spoken Dutch, henceforward CGN). Corpus components were selected in function of a *register factor* we wish to enter into the analysis, to corroborate our intuition that the marked word order variant (adjunct-subject) is much more characteristic of carefully edited written Dutch than of colloquial Dutch. So, whereas the CGN contains moderately to very colloquial spoken Belgian Dutch[11], *Het Belang van Limburg* (HBVL) and *De Gazet van Antwerpen* (GVA) are regional popular newspapers containing informal written Dutch for an uneducated audience, and *De Standaard* (STA) is a high-brow quality newspaper.

From these sources we extracted all *er*-initial active main clauses with one locative adjunct, which either preceded or followed the subject. All locative adjuncts consisted of a prepositional phrase (PP) which was

Table 1.   *Corpus size and number of attestations per corpus component*

	CGN	HBVL	GVA	STA
n tokens	3,151,535	3,012,330	3,068,405	3,228,960
n attestations	75	92	101	92

never attested between punctuation marks: the movement of isolated adverbs such as *ergens* 'somewhere' or *hier* 'here' is more restricted in Dutch than that of PP's; punctuated adjuncts, by contrast, move much more freely. Only 360 compliant presentative sentences were attested. Table 1 diagrams the size (in n of tokens) per corpus component, as well as the number of *er*-initial active main clauses with one locative adjunct extracted from each corpus component.

Observe that we did not include *er*-sentences with temporal adjuncts in the analysis, for the practical reason that it is impossible to code time adverbials for some of the parameters (notably ADJUNCT CONCRETENESS) necessarily included in the analysis.

5. Parameters and proportions

If we wish to establish that ADJUNCT CONCRETENESS has a discernible effect on constituent ordering in *er*-initial presentative sentences, we will have to show that pre-subject position in these sentences is typically reserved for non-concrete location denoting adjuncts. To that effect, we classified all adjuncts in our dataset on the basis of the *salience taxonomy* developed in Grondelaers (2000: 176−177), an operational parameter of ADJUNCT CONCRETENESS which breaks up into five distinct values. The maximal value is assigned to adjuncts which refer to the physical aspects of the speech situation, by definition the most salient and stable reference point available to the language user. Value 2 is assigned to concrete two- or three-dimensional objects, and value 3 to unbounded physical spaces. Value 4 is reserved for locations which cannot be straightforwardly categorized as either concrete or abstract. In a sentence such as *Bij de BOB is er een gebrek aan personeel* 'In the Belgian Criminal Investigation Department there is a shortage of staff', *de BOB* 'The Belgian Criminal Investigation Department' refers to a metaphorical location which can nevertheless be metonymically interpreted as referring to 'the building in which the BOB resides'. Value 5, finally, is assigned to abstract locations which allow no spatial interpretation whatsoever.

Table 2 charts absolute and relative frequencies of the five adjunct types distributed over both word order variants:

Table 2.   *Raw and relative frequencies of* ADJUNCT CONCRETENESS *as a function of word order type (Exact p = 0.0001; Cramer's V = 0.289)*

	adjunct-subject		subject-adjunct	
	abs	Rel	abs	rel
speech situation	1	1.35	2	0.7
2 and 3-dim.	6	8.11	98	34.27
unbounded spatial	18	24.32	89	31.12
vaguely spatial	43	58.11	85	29.72
Abstract	6	8.11	12	4.2
Σ	74		286	

Two preliminary remarks have to be made in connection with Table 2. Observe first that these data confirm the marginal status of the subject-final word order variant: on average, the adjunct precedes the subject in only 20.56 % of all attestations. It should be noted, however, that the proportion of marked word order varies with *register*: the distribution of the two word orders over the selected source types corroborates our expectation that marked word order decreases with increasing formality ($\chi^2 = 12.2$; *df* = 3; *p* < 0.01): in the CGN, only 9.33 % of *er*-initial presentatives have the subject-final order, whereas HBVL and GVA respectively have 16.3 % and 22.77 % of subject-final presentatives. In STA, the proportion of marked word order rises to no less than 31.52 %, which suggests that the reversal of subject and adjunct in *er*-initial presentative sentences is very much a characteristic of high-brow, edited newspaper language.

More importantly, the data in Table 2 confirm the impact of ADJUNCT CONCRETENESS on constituent ordering variation in presentative sentences. Especially eye-catching is the pre-subject paucity of adjuncts denoting concrete, bounded objects (dubbed *2 and 3-dim.* in the Table). Recall that such adjuncts significantly improved subject processing in our self-paced reading designs, which show that they provide considerably better inferential access into the subject than vaguely spatial or abstract location adjuncts. It is revealing that it is above all this "giant" among inference-generators which is barred from pre-subject position. Observe furthermore that in five out of six exceptions to this generalization, the adjunct is tagged as a two- or three-dimensional bounded object for reasons of methodological consistency, although it is in all likelihood interpreted otherwise:

(5) De openheid van die groep komt tot uiting in de inrichting van de dagelijkse leefruimte: zitgedeelte, eetgedeelte en keuken vormen één geheel. *In de open keuken* is het lang niet altijd piekfijn in orde, want *er* gebeurt *in de keuken* meer dan koken alleen.

'The openness of this component is reflected in the layout of the living quarters: sitting area, eating area and kitchen form one whole. In the open kitchen not everything is always in the proper order, for more takes place in the kitchen than only cooking (literally: there occurs more in the kitchen than only cooking).'

(6) Er *dreigde   tot in mijn ooghoeken*
Er threatened in the corner of my eyes
een onwelvoeg-lijke ontroering.
an indecent poignancy
'In the corner of my eyes (there) threatened an indecent poignancy.'

In (5), *in de open keuken* 'in the open kitchen' refers to a three-dimensional material enclosure, and we cannot exclude that *in de keuken* 'in the kitchen' in the *er*-sentence does so too. It is also possible, however, that the latter more generally denotes 'the section of a house dedicated to the preparation of food, as opposed to the dining area and the sitting area', in which case it does *not* designate a bounded space (in Section 7, we will elaborate on this example). Since, however, such post-hoc considerations endanger responsible coding, *keuken* 'kitchen' is interpreted as a concrete three-dimensional enclosure in any context which allows that sense. The same goes for *ooghoeken* 'corners of the eye' in (6), although here too a figurative reading is probably intended.

Although the data in Table 2 provide a provisional[12] confirmation of the impact of ADJUNCT CONCRETENESS on constituent ordering in *er*-initial sentences − which suggests that the *er*-initial construction is indeed an inference-blocking device − not all our problems are solved. The annoying syntactic consequence of this type of decontextualization is a high frequency of attestations with adjuncts which are sentence-final solely because they cannot precede the subject. These adjuncts, in addition, are in apparent violation of the GIVEN-BEFORE-NEW principle (see a.o. Gundel 1988: 229), a well-known determinant of constituent ordering (cf. Arnold et al. 2000). Being nearly always definite in our sample[13], adjuncts code older information than the indefinite subject, which should therefore be sentence-final according to the GIVEN-BEFORE-NEW principle. The following examples, however, suggest that unmarked word order in some *er*-initial presentative sentences *does* comply with that principle:

(7) a. *... met de melding dat er een bom zou ontploffen. "Dit keer is het geen grap.* Er *ligt een bompakket op de zesde verdieping", zei de onbekende man.*

'... with the announcement that a bomb would explode. "This time it's no joke. There is a bomb on the sixth floor", said the unknown man.'

b. *Dit keer is het geen grap: het bompakket ligt op de zesde verdieping.*

'This time it's no joke: the bomb is on the 6th floor.'

(8) a. *Het concert dat Bryan Adams op zaterdag 31 mei in Flanders Expo in het kader van zijn "18 till I die"-tour geeft, is al uitverkocht.* Er *komt een extra show op vrijdag 30 mei.*

'The concert Bryan Adams is scheduled to give in Flanders Expo on Saturday May 31, has already sold out. There will be an extra show on Friday May 30.'

b. *De extra show vindt plaats op vrijdag 30 mei.*

'The extra concert will take place on Friday.'

(9) *[De persfotografen]$_i$ hebben het ongeval, volgens de anonieme politie-bron, dus niet veroorzaakt. Ze speelden slechts een tweederangsrol in het drama. Er bevonden zich [geen paparazzi]$_i$ vlak voor, naast of achter de wagen.*

'The press photographers are not, according to the anonymous police source, responsible for the accident. They only played a minor role in the drama. There were no paparazzi right in front of, alongside, or behind the car.'

In (7)−(9), the syntactically definite sentence-final adjunct codes "newer" information than the indefinite subject. Crucially, each of the subjects in (7)−(9) is contextually available or inferable at the moment of creation: in (9), the subject has already been referred to by means of a definite determiner (cf. [ ]$_i$), and in (7)−(8), it can be marked definite in the non-presentative paraphrases in b. The adjunct, by contrast, is neither available nor inferable: in (7), *verdieping* 'floor' can be metonymically inferred from 'building', but *zesde* '6th' is new; the new information contributed by the presentative in (8) is the exact date of the extra concert, and (9) conveys the exact position of the paperazzi *vlak voor, naast of achter* 'right in front of, alongside, or behind' with respect to the contextually available car. Thus, if we can accept that accessibility differences between subject and adjunct are not always straightforwardly reflected in the structural characteristics of the encoding NP's − information marked definite can apparently be newer in some cases than information marked indefinite − we can similarly accept that the canonical

ordering "subject-adjunct" in *er*-initial presentative sentences is function-
ally motivated. "Relative adjunct accessibility", therefore, is a factor
which should be taken into account in our analysis of word order varia-
tion in presentative sentences.

This factor, however, resists reliable implementation and quantifica-
tion. No available given-new taxonomy − a diagnostic tool to classify
NP referents in terms of their degree of contextual accessibility − would
accommodate such reversal of accessibility marking as illustrated in (7)−
(9), with indefinite and definite NP's resp. conveying available and non-
available information. In order to be operational, most existing taxono-
mies uniquely position referents on the basis of their "conventionalized
linguistic dress" (Ariel 1990: 82), viz. the formal characteristics of refer-
ring NP and context. Since indefiniteness is a major formal parameter
of contextual newness, referents which are typically marked indefinite
structurally occupy the lowest ranks of the taxonomy, well below entities
which are normally marked definite (cf. Prince 1981; Gundel et al. 1993;
Grondelaers and Heylen 2003). As a consequence, there is no structural
motivation to assign a higher accessibility rank to the subjects in (7)−
(9) than to the adjuncts[14]: if we wish to implement the factors SUBJECT
NEWNESS and ADJUNCT NEWNESS in any reliable way, definite adjuncts will
always receive a lower newness value than indefinite subjects[15]. As a
consequence, it is impossible to include both SUBJECT NEWNESS and AD-
JUNCT NEWNESS in the analysis in a non-ad hoc way, however logical and
desirable it would be to do so, for what is at stake here is *relative* new-
ness, not absolute newness: sentence-final position in *er*-initial sentences
appears to be reserved not simply for a *new* entity, but for the *newest* en-
tity[16].

Fortunately, the adjuncts in (7)−(9) communicate not only the *newest*
information in the sentence, but also the *focus* information (it bears testi-
mony to the organic relation between the notions "newness" and "focus"
that they are frequently interchanged and even confused in the pragmatic
literature on information distribution; cf. Schmid 1999, chapter 3 for an
overview). One pivotal advantage of the notion *focus* in the sense that it
is used here, is that it can be reliably quantified, irrespective of the defi-
niteness or indefiniteness of the referring NP. The best-known operation-
alization of focus or informational prominence is found in Givon (1983:
14), who defines the notion "persistence" as "a reflection of the topic's
importance in the discourse"; persistence is quantified as the number of
propositions to the right of element x, in which x "continues an unin-
terrupted presence as a semantic argument of the clause, an argument of
whatever role and marked by whatever grammatical means". We fol-
lowed Givon (1983) and Brown (1983: 318) in their insistence on referent
identity when measuring topic continuity, so just like Givon and Brown

we excluded hyponymous or hyperonymous reference to x as the same referent. We did, however, consider synonymous reference as identical, provided that the synonym did not add information which had not previously been available (unlike Givon and his colleagues, who illustrate their persistence calculation with colloquial, spoken language materials, our analysis builds for the most part on written newspaper language, in which reference to the same entity is intentionally varied for stylistic purposes). Another deviation from Givon's topic continuity statistic is our decision to relax his stipulation that continuous reference should be *uninterrupted*: we allowed interruption of topic continuity by maximally one clause (mainly to avoid situations whereby persistent reference to x is not calculated as such because of intermittent modal statements such as "as the reader will recall").

Table 3 charts the impact of ADJUNCT PERSISTENCE on constituent ordering in *er*-initial presentative sentences. On the horizontal axis, the two word order variants are contrasted, whereas the vertical axis lists the number of clauses in which the adjunct entity is in some way referred to in the subsequent context (for convenience, the persistence data are presented in three frequency categories):

Table 3.   ADJUNCT PERSISTENCE *as a function of word order type ($\chi^2 = 11.84$; df = 2; p = 0.003; Cramer's V = 0.181)*

	adjunct-subject		subject-adjunct	
	abs	rel	abs	rel
0	59	79.73	166	58.04
1 or 2	11	14.86	84	29.37
3 or > 3	4	5.41	36	12.59
Σ	74		286	

Again, it is not very intuitive to focus on this parameter in an absolute way: the fact that the adjunct persists for a given number of sentences is not very revealing if the subject is continued longer in discourse. We therefore also calculate SUBJECT PERSISTENCE in Table 4. More generally, all parameters assumed to have an impact on constituent ordering will, whenever possible, be implemented for both subject and adjunct, and added to the regression analysis so that their *relative* impact can be monitored.

Tables 3 and 4 provisionally confirm the impact of ADJUNCT PERSISTENCE and SUBJECT PERSISTENCE: for both constituents goes that they are significantly more persistent when sentence-final. Informational prominence, hence, is a positive ordering motivation: adjuncts not only follow

Table 4.   SUBJECT PERSISTENCE *as a function of word order type* $(\chi^2 = 13.05; df = 2; p = 0.001; Cramer's V = 0.19)$

	adjunct-subject		subject-adjunct	
	abs	rel	abs	rel
0	21	28.38	127	44.41
1 or 2	28	37.84	113	39.51
3 or > 3	25	33.78	46	16.08
Σ	74		286	

the subject when they cannot precede it; they are also sentence-final when they encode relatively more important information the speaker wishes to come back to later. Observe that it is not possible at this stage to answer the question which of both factors is the primary mover of the variation observed. That will have to wait till the regression analysis in the next section.

Let us, before moving on to that analysis, focus on one more parameter. The importance of HEAVINESS or LENGTH as a determinant of word order variation has frequently been recognized. Arnold et al. (2000) cite it as a major motivator of the heavy NP-shift alternation and dative shift: the heavier (viz. the longer) the constituent, the higher the chance that it will be delayed. We follow Arnold et al. (2000) in their decision to operationalize LENGTH in terms of number of syllables (implementation in terms of number of words would obscure the actual length difference between such pairs as *in him* and *on Madagascar*). To appreciate the importance of the relative impact of LENGTH, the vertical axis of Table 5 features three relative length categories, viz. attestations whereby the adjunct is longer than the subject, attestations whereby adjunct and subject are equally long, and attestations whereby the adjunct is shorter than the subject.

The data in Table 5 confirm the impact of relative LENGTH on constituent ordering. A somewhat closer look reveals that it is in particular SUBJECT LENGTH which plays a decisive role in constituent placement: whereas relatively longer subjects end up in sentence-final position in 75.68 % of all cases, there is no marked preference for sentence-final longer adjuncts.

In the regression analysis, to which we turn next, eight factors were eventually entered: the contextual factor REGISTER, and the language-structural factors ADJUNCT LENGTH, SUBJECT LENGTH, ADJUNCT PERSISTENCE, SUBJECT PERSISTENCE, ADJUNCT CONCRETENESS, SUBJECT CONCRETENESS[17] and ADJUNCT NEWNESS. Data were collected and analyzed in order

Table 5. *Relative* LENGTH *of adjunct and subject as a function of word order type*
         ($\chi^2 = 25.79$, df = 2; $p < 0.001$; Cramer's V = 0.267)

	adjunct-subject		subject-adjunct	
	abs	rel	abs	rel
adj > subj	11	14.86	120	41.96
adj = subj	7	9.46	43	15.03
adj < subj	56	75.68	123	43.01
Σ	74		286	

to answer three sets of questions. First, can we explain and predict the constituent ordering variation at issue here on the basis of these eight factors? And if we can, are all those factors statistically significant determinants of that variation? Crucially, does ADJUNCT CONCRETENESS have a significant impact on it? If yes, we have a reliable indication that inaccessibility marking and decontextualization are not only the functional motor of the syntax of adjunct-initial presentative sentences but also of *er*-initial presentatives. Third, do all factors contribute in the same way to the variation observed, or are some factors more outspoken determinants of the variation at issue? The current state of affairs − all factors except ADJUNCT NEWNESS have been found to be significant determinants of the variation, but it is as yet unclear *how* significant their impact is − allows for two opposite models:

− if the regression analysis confirms that ADJUNCT CONCRETENESS is not only a significant factor, but also the most important driving force of constituent ordering, then inaccessibility theory *motivates* the syntax of *er*-initial sentences;
− if the regression returns significance and impact estimates which indicate that ADJUNCT CONCRETENES is a significant factor, but not the main determinant, then inaccessibility theory merely conditions the operation of more important word order motivations.

6. Regression

In the next phase of the research, we carried out a series of stepwise logistic regression analyses, which select and order factors in function of their impact on the dependent variable, which is the POSITION OF THE ADJUNCT in *er*-initial presentative sentences (preceding the subject or following the subject). A stepwise analysis only retains those factors which

Table 6. *Input model with p-values for eight possible determinants of constituent order-ing in er-initial presentative sentences*

	p-value
register1: STA − GVA	0.273
register2: STA − HBVL	0.455
register3: STA − CGN	0.052
adjunct concreteness	0.029
adjunct persistence	0.002
adjunct length	0.005
adjunct newness	0.043
subject concreteness	0.116
subject persistence	0.005
subject length	0

significantly increase model fit. In order, however, to determine the signi-ficance of *all* the factors we entered into the analysis, we first analysed the input model containing the eight factors discussed in Section 5. Table 6 contains p-values, which reflect the statistical significance of each factor as a determinant of constituent ordering variation in *er*-initial presentative sentences. Observe that REGISTER was not entered in the analysis as an ordinal variable: we used dummy coding to enter REGISTER as a nominal variable, with the most formal materials − *De Standaard* − as the reference value. As a result, the statistical software returns a p-value for three register comparisons, viz. between *De Standaard* and *Ga-zet van Antwerpen*, between *De Standaard* and *Belang van Limburg,* and between *De Standaard* and the spoken materials from the *Corpus Gesp-roken Nederlands.*

The principal conclusion to be drawn from the input model is that not even REGISTER3, the level where the most extreme register differences should obtain given the stylistic divergence between the very formal *Standaard*-materials and the spoken CGN-data, reaches full statistical significance ($p = 0.052$) in this analysis. To be sure, REGISTER is not rejected on purely technical grounds. In order to test for multicollinear-ity − the situation whereby two factors correlate to such an extent that one of both becomes statistically superfluous, no matter how important from a linguistic point of view − we computed Variance Inflation Factors (VIF's) for all the factors in the input model. Since none of those exceeds 1.5427 (the threshold for multicollinearity is 4), there is no reason to expect purely statistical distortions.

The genuine reason for the limited significance of REGISTER appears to lie elsewhere. Although the CGN is identical to the other corpus compo-nents in terms of size (compare Table 1 in Section 4), it contains less *er-*

Table 7.   *Stepwise model with p-values and Odds Ratios for five determinants of word order variation in er-initial presentative sentences*

	p-value	Odds Ratio
subject length	0	0.405
adjunct persistence	0.003	2.015
adjunct concreteness	0	0.543
subject persistence	0.004	0.642
adjunct length	0.014	1.452

initial sentences than the other sources in the corpus ($n = 75$ vs. $n = 92$, 92 and 101[18]). In addition, the CGN-materials consist for the most part of read aloud written text (news reports and novels written for the blind): the set of *er*-initial sentences attested in truly unscripted, spontaneous conversation represents no more than 25.3% of the CGN-dataset. However, given the already limited size of that dataset, and in order to retain workable frequencies, we initially decided to include *all* spoken materials, scripted or not. The limited impact of REGISTER on constituent ordering variation in *er*-initial sentences may therefore well be a consequence of the design of the CGN-corpus and should, as a consequence, be treated with caution.

Let us next focus on the model returned by the stepwise procedure, which selects and orders factors in terms of model fit. In addition to p-values, Table 7 contains *Odds Ratios* which reflect the relative importance of each factor's impact on the variation at issue. Odds Ratios should be interpreted as follows: the Odds Ratio 2.015 for ADJUNCT PERSISTENCE indicates that the odds for adjunct-final word order increase 2.015 times with each next persistence level. Conversely, the Odds Ratio 0.405 for SUBJECT LENGTH reveals that the odds for adjunct-final word order *decrease* 2.47 times as the length of the subject increases.

In order to obtain the best model fit, the regression originally retained all the factors except REGISTER[19]. We then omitted the non-significant factors SUBJECT CONCRETENESS ($p = 0.1349$), and ADJUNCT NEWNESS ($p = 0.504$), an operation which did not weaken the model in any considerable sense (compare AIC 267.45 and AIC 269.04). Again, the non-significance of the rejected factors is no artefact. Since their VIF's all range between 1.071 and 1.409, there is no multicollinearity. Neither does any of the three rejected factors enter into significant higher-order interactions.

Before we go into the significance and impact of the retained factors, let us first concentrate on some estimates of the predictive and explanatory quality of the stepwise model. The fact that only 20.56% of the

attestations in our dataset have subject-final word order makes for a high predictive power of the naive model without factors (79.4%). The higher success rate (88.89%) of the fitted model suggests that what little variation is initially left unaccounted for, is explained to a considerable extent by the added factors. Reliable statistical evidence for the explanatory power of the fitted model can be obtained by computing the *lack-of-fit* between the fitted model and a "saturated" model, in which any possible interaction − no matter how *ad hoc* − is included to reach maximal power. The p-value returned for this test ($p = 0.71$) indicates that no lack-of-fit can be established and that, as a consequence, there is no indication that our model lacks crucial additional factors.

The five significant factors that were eventually retained are listed in Table 7 (in their order of addition to the forward stepwise model). Crucially, the high significance of ADJUNCT CONCRETENESS confirms our hypothesis that an adjunct's position in an *er*-initial sentence correlates with its degree of concreteness. Hence, we now have conclusive evidence that *er*-sentences do not favour incompatible signals: only pre-subject adjuncts with a low inferential potential appear to be compatible with the decontextualization effected by *er*.

The Odds Ratio 0.543 for ADJUNCT CONCRETENESS indicates that the odds for pre-subject adjuncts will increase 1.85 times as adjuncts get more abstract. The Odds Ratios for the other factors, however, suggest that concrete location adjuncts are not simply postponed because they cannot precede the subject. In fact, the data indicate that adjuncts are predominantly sentence-final in two conditions: when they are *longer* than the subject (SUBJECT LENGTH Odds Ratio = 0.405; ADJUNCT LENGTH Odds Ratio = 1.45; see Arnold et al. 2000 for similar evidence), and, more importantly, when they are more *prominent*, viz. continued longer in discourse (ADJUNCT PERSISTENCE Odds Ratio = 2.015; SUBJECT PERSISTENCE Odds Ratio = 0.642). On the basis of these data, we propose that INFORMATIONAL PROMINENCE is the main motivation for delaying materials in *er*-initial presentative sentences. CONCRETENESS and LENGTH merely *condition* the operation of the "delay important materials" principle: unimportant constituents which are undeserving of sentence-final position can only precede the subject when they are not too complex, and when they do not generate too many predictive inferences with respect to the upcoming subject.

Although we have restricted the regression analysis to locative adjuncts (whose CONCRETENESS and NEWNESS can be measured in a more reliable way than that of temporal adjuncts), it should be noticed that it also accounts for the distribution of temporal adjuncts. In (4), the temporal adjunct *tijdens de spitsuren* 'during the rush hour' communicates new as well as central information, and it is accordingly sentence-final.

In (2), by contrast, *op dit moment* 'at this moment' is not very prominent information-wise: since it is neither too complex nor too rich in predictive inferences with respect to the subject, it can precede the latter. The fact that temporal adjuncts rarely generate useful inferences about upcoming materials explains why they typically precede the subject, and why, hence, the canonical word order for *er*-initial sentences with one temporal adjunct is the one illustrated in (2).

Summarizing: although ADJUNCT CONCRETENESS is not the primary determinant of constituent ordering in *er*-initial presentative sentences, the regression findings prove that these sentences inherit their constructional meaning from *er*'s function. The *decontextualization* effected by *er* is therefore reflected in the syntax of *er*-initial presentatives. As a consequence, we can safely extend the inaccessibility theory of sentence-internal *er* to sentence-initial *er*.

7. Conclusions

This paper has been dedicated to outlining the merits of a variationist approach to linguistic phenomena traditionally considered to be in the margin of theoretical linguistics, viz. pragmatics and discourse analysis. This is what our regression analysis of eight determinants of constituent ordering in 360 *er*-initial sentences, attested in three registers of written and spoken Belgian Dutch, revealed.

The main determinant of word order in *er*-initial sentences is informational prominence: the constituent which makes the greatest informational contribution to the clause – be that the subject or the adjunct – tends to be sentence-final, whereas less prominent materials are "backgrounded" to penultimate position. This implies that one of the major functions of *er*-initial presentative sentences is focus attribution. Our data reveal that there is a *low-level* as well as a *high-level* constraint on constituent reordering: non-prominent adjuncts can only be backgrounded into pre-subject position when they are not too complex, and when they are not too concrete. The statistical significance of the latter constraint confirms our hypothesis that adjunct-initial sentences with postverbal *er* and *er*-initial sentences share a commitment towards inference-deactivation and decontextualization. The determined avoidance in *er*-initial sentences of *recontextualization* by means of a concrete locative only makes sense in a construction type which is concerned with *decontextualization*.

A major question which remains to be solved in this light is *how* exactly decontextualization is instantiated in a sentence type which is not so much concerned with the creation of new entities, as with the attribution of focus to new and non-new entities. Why should focus attribution

benefit from decontextualization? Though space limitations preclude an elaborate answer to this question, we will hazard a provisional explanation (or, more correctly, we will outline the direction in which that answer can eventually be found). Let us first define context as a database, or rather a set of databases, pertaining to the state of affairs in the discourse model of a text (we use the term "discourse model" as in Prince 1981). Context keeps a "track record" of each nominal entity and process in which − in the case of nominal entities − the most recent information about *type* (designated by the lexical head of the referring NP), *subtype* (designated by adjectival qualifiers), *quantity*, and *newness* (both designated by the determiner system) is stored and continuously updated (see Kamp 1988 for a related view of context). Example (10) is an interesting case in point of how the context, or more particularly the *kitchen*-record, determines how reference to that entity is made in unfolding discourse, and where it is positioned in successive clauses:

(10) a. *De openheid van die groep komt tot uiting in de inrichting van de dagelijkse leefruimte: zitgedeelte, eetgedeelte en keuken vormen één geheel. In de open keuken is het lang niet altijd piekfijn in orde, want* **er** *gebeurt* **in de keuken** *meer dan koken alleen.*
'The openness of this component is reflected in the layout of the living quarters: sitting area, eating area and kitchen form one whole. In the open kitchen not everything is always in the proper order, for more takes place in the kitchen than only cooking (literally: there occurs more in the kitchen than only cooking).'

     b. *De openheid van die groep komt tot uiting in de inrichting van de dagelijkse leefruimte: zitgedeelte, eetgedeelte en keuken vormen één geheel. In de open keuken is het lang niet altijd piekfijn in orde, want* **er** *gebeurt meer dan koken alleen.*
'The openness of this component is reflected in the layout of the living quarters: sitting area, eating area and kitchen form one whole. In the open kitchen not everything is always in the proper order, for more takes place there than only cooking (literally: there occurs more than only cooking).'

When *keuken* enters discourse, a "folder" is opened which specifies a number of basic parameters of that entity, in this case the fact that one specific member of the general type "keuken" has been entered in discourse[20], and is available for further reference. Connected with these parameters, we suggest, is a complex of experientially derived pragmatic and syntactic implications which are proportionally updated as the basic parameters are adjusted to unfolding context. Such implications pertain

to the relation between newness and *prominence* ("the newer, the more important"), *sentence position* ("the newer, the more sentence-final"), and *type of referring expression* ("the newer, the lower the Accessibility Marker" in the sense of Ariel 1990); it follows that when an entity gets more activated in discourse (as a result of repeated mention), it also becomes less *prominent*, less *sentence-final*, and it tends to be designated by *higher Accessibility Markers*, *viz.* demonstratives and, eventually, pronouns. The lexical type parameter enters not only into frame-based implications − such as "buses have drivers", as in *When I got on the bus, the driver was drunk* − but also in more general inferential implications such as "the more concrete, the more predictive potential".

In the sentence preceding the first mention of *keuken* 'kitchen' in (10), *dagelijkse leefruimte* 'daily living quarters' has entered discourse with the frame-based implication that living quarters typically comprise a sitting area, an eating area, and a kitchen. As a result, *keuken* in the following clause is sufficiently accessible to be part of the sentence-initial constituent, and to be designated by a short definite NP rather than a long description. The next clause further activates the kitchen in discourse and adds subtype information − the "openness" of the kitchen − to its contextual passport. By then, the kitchen is so available to the comprehender that (10b) would have been perfectly acceptable, featuring the sentence-initial anaphoric variant of *er* (equivalent to English referential *there*) which designates *de open keuken* 'the open kitchen'. In (10b), this referential *er* is a High Accessibility Marker which would be fully compatible with the discourse-conditioned topicality of *keuken*. This, however, is not what happens in attested discourse: the *er*-sentence ignores all the implications of the discourse status of *keuken*, and unexpectedly continues reference to the latter with a Low Accessibility Marker, a non-initial, definite NP. This seeming violation of accessibility marking has two immediately discernible effects. It not only casts doubt on the coreferentiality of *de open keuken* and *de keuken*; its pragmatic effect is quite considerable: *in de keuken* no longer is a fully accessible, easily retrievable topic, but a piece of information which has become worthy of attention itself.

We therefore hypothesize that the decontextualization established in *er*-initial sentences amounts to the "switching off" of some or all of the *implications* of specific parameter adjustments in the track records of entities and processes in the model of the foregoing discourse, so that materials included in the *er*-sentence can be stripped off of contextually accumulated features the speaker wishes to change or reset for some purpose. In (10), the speaker wants to re-focus on an entity which is so highly activated in discourse that reference to it by anything else than a sentence-initial Extremely High Accessibility Marker would be awkward

and confusing to the comprehender, unless this refocusing takes place in a setting which allows the latter to *ignore* the contextual history of that entity, to temporarily *deactivate* his experience-based knowledge that topics are sentence-initial and non-prominent. So, whereas postverbal *er* in adjunct sentences decontextualizes a subject by deactivating all frame-based predictive inferences with respect to that subject, *er*-initial sentences decontextualize presented materials by erasing their accumulated contextual history. The difference between the decontextualization established in adjunct sentences with *er* and the decontextualization coded by *er*-initial sentences therefore amounts to a difference in direction: whereas the former turn off *forward* predictive assumptions, the latter operate *backwards*, by expunging contextual records.

Our findings have three important theoretical consequences. First, they rule out any analysis which reduces the function of *er*-initial sentences to foregrounding an important new subject. The tenacity of the latter view is an artefact, we believe, of the one-on-one relation tacitly assumed between newness and informational prominence. We hope to have demonstrated that adjuncts need not be new in order to be the most prominent element in an *er*-initial clause. Second, since prominence was implemented in this study in terms of Givon's (1983) "persistence", our data contradict Kirsner's (1979: 161) statement that *er*-initial sentences should not be dubbed "presentative" in the sense of Hetzron (1971), viz. "introducing an entity in discourse so that it may be referred to later": since the sentence-final constituent in *er*-initial constructions is typically the one which persists longer in discourse, these constructions are eminently presentative in whatever sense. Kirsner's view, again, is the result of exclusive reliance on the subject when dealing with the function of *er*-sentences. If, third, informational prominence is the main motivation for delaying materials in *er*-initial presentative sentences, then the typical indefinite-subject-before-definite-adjunct ordering of *er*-initial sentences does not go against the better-known GIVEN-BEFORE-NEW word ordering motivation. Since informational prominence appears to overrule newness as an ordering determinant in some contexts, GIVEN-BEFORE-NEW had better be rephrased as UNIMPORTANT-BEFORE-IMPORTANT. Although newness is a major parameter of prominence, it is obviously not the only one. And the correspondence between newness and indefiniteness is too partial to have absolute faith in purely syntactic determinants of informational importance.

On the methodological level, we hope to have demonstrated that by gaining insight into the factors which determine constituent ordering in *er*-initial sentences, we were able to access the constructional meaning of these sentences. This allowed us, in turn, to settle some longstanding syntactic and semantic issues in connection with *er*, one of the tiniest but

linguistically most troublesome morphemes in the Dutch language. The main advantage of this empirical approach to functional analysis is the guarantee that it provides a reliable factual basis for a discipline which is rightly criticized as being "woolly" sometimes. Functional linguistics can therefore become more cumulative: when linguistic hypotheses can be tested against a shared basis of corpus data, they will become more comparable than is currently the case, with many theories existing in parallel, without sufficient common ground for a stringent comparison of competing models.

Bionotes

Stefan Grondelaers is currently affiliated with the Institut des Langues Vivantes of the Université Libre de Bruxelles, and in the Research Unit Quantitative Lexicology and Variational Linguistics of the University of Leuven. In 2000 he received his PhD at the University of Leuven, with a thesis focusing on the sociolinguistics, the pragmatics, and the psycholinguistics of Modern Dutch presentative *er* "there". Stefan Grondelaers investigates lexical and syntactic variation, building on corpus-linguistic and psycholinguistic research methods.

   Dirk Speelman is Associate Professor at the Department of Linguistics (Research Unit Quantitative Lexicology and Variational Linguistics) of the University of Leuven. In 1997 he received his PhD at the University of Leuven, with a thesis on the cross-fertilization between prototype-theoretical concepts and empirical methods from quantitative corpus linguistics. Much of his work focuses on methodology, and on the application of statistical and other quantitative techniques to the study of language.

Notes

1. In (1)−(4), typicality rates are computed as the relative frequency per 10,000 words of the *er*-initial sentence type in which the locative adjunct follows the subject (as in 1), the *er*-initial sentence type in which the temporal adjunct precedes the subject (2), the *er*-initial sentence type in which the locative adjunct precedes the subject (3), and the *er*-initial sentence type in which the temporal adjunct follows the subject (4). The corpus on which all Belgian typicality rates are based consists of materials from the Belgian popular newspaper *Het Belang van Limburg* (n = 1,561,362) and the Belgian quality newspaper *De Standaard* (n = 1,665,144). Our preference for typicality rates is motivated by the notorious unreliability of grammaticality judgements of presentative *er*-sentences (cf. De Rooij 1991; Grondelaers 2000).
2. Unfortunately, few functionalists are aware of these limitations. Among the notable exceptions are Andersen (2001) − commendable in his alertness to region- and

age-constraints on the use of discourse markers in English and Norwegian −, Szmrecsanyi (2005), who argues that persistence in spoken discourse is subject to contextual determinants such as turn-taking, and Gries (see especially 2003) who includes contextual constraints in his functional analyses of word order variation.

3. An excellent example of multivariate work in functionalist syntax and pragmatics is Bresnan et al. (in press), in which "a model [is constructed] that can predict the choice of dative structures with 94% accuracy" (p. 29). Bresnan, however, does not go beyond predicting the choice of alternative structures.

4. In the Government and Binding stage of generative syntax, *er*'s distribution continues to be governed by a mechanical insertion rule (Den Besten 1981, 1982, 1983; Reuland 1983; Hoekstra 1984), although the endocentric X-bar syntax and the curtailed transformational component impose more stringent restrictions on the descriptions.

5. Typicality rates for the constructions in (5)−(8) are computed as the relative frequency per 10,000 words of the presentative construction type with an initial temporal adjunct and a form of *zijn* 'to be' which does not contain *er* (5a), and which contains *er* (5b), the presentative construction type with an initial temporal adjunct and a more specific verb than *zijn* 'to be' which does not contain *er* (6a), and which contains *er* (6b), the presentative construction type with an initial locative adjunct and a form of *zijn* 'to be' which does not contain *er* (7a), and which contains *er* (7b), and the presentative construction type with an initial locative adjunct and a more specific verb than *zijn* 'to be' which does not contain *er* (8a), and which contains *er* (8b). The corpus on which all Netherlandic typicality rates are based consists of materials from the Dutch popular newspaper *De Telegraaf* (n = 1,590,581) and the Dutch quality newspaper *NRC Handelsblad* (n = 1,520,064).

6. Constructions with a sentence-initial temporal adjunct and a form of *zijn* 'to be' which do not contain *er* are extremely rare in written standard Dutch. The example in (5a) comes from the highly informal internet chat channel *#belgië* (recorded on 13/3/1999).

7. In order to avoid an inconveniently large number of example sentences, all clauses in (5)−(8) − except (5a) − were taken from formal sources. As a consequence, the impact of register on *er*-preferences is not demonstrated in (5)−(8), as a result of which there are no typicality rates for register effects.

8. To our knowledge, no generative linguist has ever tried to answer that question. Yet, the tenacity of the dummy subject-view transpires from the fact that the sporadic analyses of *er*'s postverbal distribution (Van Es and Van Caspel 1971; De Rooij 1991) invariably take the form of a list of contexts in which *er* is *weglaatbaar* 'deletable' or *facultatief* 'optional', instead of an active search for factors which motivate why *er* has to be present in certain contexts. The underlying rationale seems to be that since *er* cannot possibly have a postverbal function, it is useless to try and determine its distribution there.

9. An extensive digital text corpus of written Dutch, structured to accommodate regional variation − Belgium vs. The Netherlands − and register variation − newspaper vs. Internet language. See especially Grondelaers, Deygers, Van Aken, Van Den Heede and Speelman (2000) for additional information on the design of the corpus.

10. In Grondelaers, Speelman, Drieghe, Brysbaert and Geeraerts (in press), we replicated the experimental findings from Grondelaers, Brysbaert, Speelman and Geeraerts (2002) in an eyetracking design.

11. It should be noticed that a substantial portion of the CGN is read aloud speech, based on carefully edited prose (component 13 − news reports − and component

    14 − novels for the blind). For reasons of data scarcity, we initially included all the CGN-data in the analysis; in Section 6 we will come back to the consequences of that decision.

12. A more dependable confirmation will have to wait until Section 6, in which the relative impact of ADJUNCT CONCRETENESS vis-à-vis other predictors of constituent ordering is determined in a regression analysis.

13. Only 4.72 % of locative adjuncts in our sample are syntactically indefinite.

14. An additional problem is the fact that given-new taxonomies are presented as one-dimensional scales (from minimally to maximally accessible), on which all accessibility-inducing dimensions (memory location, referential distance, etc.) are mapped, so that accessibility statuses can be uniquely positioned and quantified. In spite of its convenience, this one-dimensional operationalization obscures the multidimensional and non-binary character of the notion "newness". Observe, for instance, that Bryan Adams is discourse-new in (8) − this is the first mention of the Canadian rock star − but encyclopaedically given, because the speaker assumes that the comprehender has a mental representation for Bryan Adams, even though it has not been accessed in the foregoing context. Even more difficult are inferables such as *de zesde verdieping* "the 6th floor" in (7), which have to be created (and, as a consequence, are new) on the basis of accessible information metonymically associated with old entities (in this case: buildings have floors).

15. In Givon (1983), the newness status of a referent is operationalized in terms of "frequency of mention" and "distance to last mention". While we agree that these are relevant parameters of an entity's degree of contextual accessibility, they do not adequately discriminate between fully new entities (which are marked indefinite) and encyclopaedically available entities (Ariel 1990) which are accessible without having been mentioned in the foregoing context (a case in point are referents of proper names such as *In Belgium* or *In the eastern-most village of Flanders*). This is not only infelicitous from a purely cognitive point of view: encyclopaedically available materials are much more accessible than fully new referents. A methodological disadvantage of Givon's metrics is that they cannot be invoked to explain why encyclopaedically available adjuncts frequently precede the adjunct in our data (as in *Er zijn in Vlaanderen teveel universiteiten* 'There are in Flanders too many universities'), while indefinite adjuncts almost never do (a sentence type such as *Er zijn in een voetballeven minder prettige dingen dan te beginnen in een WK-duel tegen Nederland.* 'There are in a football player's life less nice things than to start against The Netherlands in a World Cup duel.' is attested only twice in our data; it moreover sounds highly artificial to native speakers).

16. We have tried to compensate to some extent for this methodological drawback by tagging all adjuncts for NEWNESS, building on the operational taxonomy reported in Grondelaers (2000) and revised in Grondelaers and Heylen (2003), which classifies all adjuncts in ten ranks from "maximally accessible" to "completely new and unanticipated". For statistical convenience, the ten ranks were later reduced to five: "new", "discourse-new", "available in the remote linguistic context", "available in the recent linguistic context" and "referring to the ground".

17. Subject referents were classified into 1 = animate, 2 = bounded material object, 3 = unbounded material object, 4 = vaguely spatial entity and 5 = abstract entity.

18. A possible explanation for the avoidance of *er*-initial structures in spoken register could be the following. If sentence-internal *er* "corrects" ongoing inferencing by deactivating those inferences projected by the preceding adjunct and the verb, then sentence-initial *er* will block any inference made available by the foregoing context. As a result, all the immediately following materials (not only the subject) have to be processed without the support of contextual preparation. Now, while

such a temporary suspension of context may be beneficial for data introduction in scripted text (in which the intrinsic and contextual accessibility of new materials can be carefully monitored and manipulated), it may be harmful in colloquial interaction where no such safety net is available, as a result of which communication problems may much easier arise (a more elaborate analysis along those lines can be found in Grondelaers, Speelman and Geeraerts: in press).

19. We computed forward and backward stepwise analyses, which returned exactly the same model.

20. It should be noticed that although *keuken* is discourse-new when first mentioned, it is syntactically definite: whereas substitution of *de keuken* 'the kitchen' for *keuken* is possible, substitution of *een keuken* 'a kitchen' for *keuken* yields an unacceptable sentence. In this case, definiteness is probably triggered by the referent's *unicity*, viz. the fact that houses typically have only one kitchen (Langacker 1991: 104).

References

Altmann, Gerry T. M., and Yuki Kamide
 1999 Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73(3), 247−264.
Andersen, Gisle
 2001 *Pragmatic Markers and Sociolinguistic Variation*. Amsterdam/Philadelphia: John Benjamins.
Ariel, Mira
 1990 *Accessing Noun-Phrase Antecedents*. London/New York: Routledge.
Arnold, Jennifer. E., Tom Wasow, Ryan Ginstrom and Teresa Losongco
 2000 Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76(1), 28−55.
Bech, Günnar
 1952 Über das Niederländische Adverbialpronomen *er. Travaux du cercle linguistique de Copenhague* 8. Kopenhagen/Amsterdam, 5−32.
Bennis, Hans
 1986 *Gaps and Dummies*. Dordrecht: ICG Printing.
Besten, Hans den
 1981 Government, syntaktischen Struktur und Kasus. In Kohrt, Manfred and Jürgen Lenerz (eds.), *Sprache: Formen und Strukturen. Akten des 15. Linguistischen Kolloquiums*. Tübingen: Niemeyer, 97−107.
 1982 Some remarks on the ergative hypothesis. *Groninger Arbeiten zur Germanistischen Linguistik* 21, 61−82.
 1983 On the interaction of root transformations and lexical deletive rules. In Werner Abraham (ed.), *On the Formal Syntax of the Westgermania. Papers from the 3rd Groningen Grammar Talks*. Amsterdam/Philadelphia: John Benjamins, 47−131.
Blom, Alied
 1992 Wat gebeurde er in Lobith? In Schermer-Vermeer, Everdina C., Wim G. Klooster and Arjen Florijn (eds.), *De Kunst van de Grammatica, Artikelen Aangeboden aan Frida Balk-Smit Duyzentkunst*. Amsterdam: Universiteit van Amsterdam, Vakgroep Nederlandse Taalkunde, 15−26.
Bolinger, Dwight L.
 1977 *Meaning and Form*. London: Longman.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina and Harald Baayen
  2007   Predicting the dative alternation. In Bouma, Gerlof, Irene Krämer and
              Joost Zwarts (eds.), *Cognitive Foundations of Interpretation*. Amsterdam:
              Royal Netherlands Academy of Science, 69−94.
Brown, Cheryl
  1983   Topic continuity in written English narrative. In Givon, Talmy (ed.),
              Topic Continuity in Discourse. A Quantitative Cross-language Study. Am-
              sterdam/Philadelphia: John Benjamins, 313−337.
Elffers, Els
  1977   *Er*-verkenningen. *Spektator* 6, 417−422.
Es, Gustaaf A. van, and Paulus P. J. van Caspel.
  1971   De patronen van de zinspotente groepen; grondtype A en zijn varianten
              II. *Publicaties van het archief voor de Nederlandse syntaxis*. Groningen:
              Rijksuniversiteit.
Geeraerts, Dirk
  2005   Lectal variation and empirical data in Cognitive Linguistics. In Fran-
              cisco Ruiz de Mendoza Ibañez and M. Sandra Peña Cervel (eds.), *Cogni-
              tive Linguistics. Internal Dynamics and Interdisciplinary Interaction*. Ber-
              lin/New York: Mouton de Gruyter, 163−189.
Geerts, Guido, Walter Haeseryn, Jaap de Rooij and Maarten C. van den Toorn
  1984   *Algemene Nederlandse Spraakkunst*. Groningen/Leuven: Wolters-Noord-
              hoff.
Givon, Talmy
  1983   Topic continuity in discourse. An introduction. In Givon, Talmy (ed.),
              Topic Continuity in Discourse. A Quantitative Cross-language Study. Am-
              sterdam/Philadelphia: John Benjamins, 1−43.
Gries, Stefan Th.
  2003   *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Place-
              ment*. London/New York: Continuum Press.
Grondelaers, Stefan
  2000   De Distributie van Niet-Anaforisch *er* buiten de Eerste Zinsplaats. Socio-
              lectische, Functionele en Psycholinguistische Aspecten van *er's* Status als
              Presentatief Signaal. Unpublished doctoral dissertation, University of
              Leuven.
Grondelaers, Stefan and Marc Brysbaert
  1996   De distributie van het presentatieve *er* buiten de eerste zinsplaats. *Neder-
              landse Taalkunde* 1, 280−305.
Grondelaers, Stefan, Marc Brysbaert, Dirk Speelman and Dirk Geeraerts
  2002   *Er* als accessibility marker: On- en offline evidentie voor een procedurele
              interpretatie van presentatieve zinnen. *Gramma/TTT* 9, 1−22.
Grondelaers, Stefan, Katrien Deygers, Hilde Van Aken, Vicky Van den Heede en
Dirk Speelman
  2000   Het CONDIV-corpus geschreven Nederlands. *Nederlandse Taalkunde* 5,
              356−363.
Grondelaers, Stefan, and Kris Heylen
  2003   Towards a "new" given-new taxonomy. Paper presented at the 8th In-
              ternational Cognitive Linguistics Association Conference, Logroño (La
              Rioja).
Grondelaers, Stefan, Dirk Speelman, and An Carbonez
  2001   Regionale variatie in de postverbale distributie van presentatief *er*. On-
              line publication on *www.neerlandistiek.nl* 1.4.

Grondelaers, Stefan, Dirk Speelman, Denis Drieghe, Marc Brysbaert and Dirk Geeraerts
   in press   Modelling indefinite reference processing: Converging on- and offline evidence for predictive inferencing and remedial cueing. *Discourse Processes*.

Grondelaers, Stefan, Dirk Speelman and Dirk Geeraerts
   2002   Regressing on *er*. Statistical analysis of texts and language variation. In Morin, Annie and Pascale Sébillot (eds.), *Proceedings of the 6th International Conference on the Statistical Analysis of Textual Data*. Rennes: Institut National de Recherche en Informatique et en Automatique, 335−346.
   2003   De distributie van *er* in het gesproken Nederlands. Paper presented at the workshop *Spraakmakende Spraak* (Corpus Gesproken Nederlands), Radboud University Nijmegen.
   in press   National variation in the use of *er* "there". Regional and diachronic constraints on cognitive explanations. To appear in Kristiansen, Gitte and René Dirven (eds.), *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*. Berlin/New York: Mouton de Gruyter.

Gundel, Jeanette K.
   1988   Universals of topic-comment structure. In Hammond, Michael, Edith Moravcsik and Jessica Wirth (eds.), *Studies in Syntactic Typology*. Amsterdam/Philadelphia: John Benjamins, 209−239.

Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski
   1993   Cognitive status and the form of referring expressions in discourse. *Language* 69(2), 274−307.

Haan, Germen J. de
   1974   On extraposition. *Spektator* 4, 161−183.

Haan, Germen J. de, Gerard A. T. Koefoed and Louis Des Tombe
   1975   *Basiskursus Algemene Taalwetenschap*. Assen: Van Gorcum.

Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij, and Maarten C. van den Toorn
   1997[2]   *Algemene Nederlandse Spraakkunst*. Groningen and Deurne: Martinus Nijhoff − Wolters Plantyn.

Heylen, Kris, José Tummers and Dirk Geeraerts
   in press   Methodological issues in corpus-based Cognitive Linguistics. To appear in Kristiansen, Gitte and René Dirven (eds.), *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*. Berlin/New York: Mouton de Gruyter.

Hetzron, Robert
   1971   Presentative function and presentative movement. *Studies in African Linguistics*. Supplement 2, 79−105.

Hoekstra, Teun
   1984   *Transitivity. Grammatical Relations in Government-Binding Theory*. Dordrecht: Foris.

Kamp, Hans
   1988   Discourse Representation Theory: What it is and where it ought to go. In Bläser, Albrecht (ed.), *Natural Language and the Computer*. Berlin: Springer, 84−111.

Kirsner, Robert S.
   1979   *The Problem of Presentative Sentences in Modern Dutch*. Amsterdam: North Holland.

Kraak, Albert
   1966   *Negatieve zinnen; een Methodologische en Grammaticale Analyse*. Hilversum: W. de Haan.

Lakoff, George
   1987   *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
Langacker, Ronald W.
   1991   *Foundations of Cognitive Grammar. Volume 2: Descriptive Application*. Stanford: Stanford University Press.
Leys, Odo
   1979   De bepaling van het voornaamwoordelijk bijwoord en de systematisering van het Nederlands *er. De Nieuwe Taalgids* 72, 240−246.
Milroy, Lesley and Matthew Gordon
   2003   *Sociolinguistics. Method and Interpretation*. Oxford: Blackwell.
Nieuwborg, Eli
   1968   *De Distributie van het Onderwerp en het Lijdend Voorwerp in het Huidige Geschreven Nederlands in zijn A.B.-vorm*. Antwerpen: Plantijn.
Paardekoper, Piet
   1963   *Beknopte ABN-syntaksis*. Den Bosch: Malmberg.
Pollmann, Thijs
   1975   Oorzaak en handelende persoon: de beschrijving van passieve zinnen in de Nederlandse grammatica. Unpublished doctoral dissertation, University of Nijmegen.
Prince, Ellen
   1981   Towards a taxonomy of given-new information. In Cole, Peter (ed.), *Radical Pragmatics*. New York: Academic, 223−255.
Reuland, Erik
   1983   Movement vs. merger: Relations between inflection and verb. *Proceedings N.E.L.S.* XIII.
Rietveld, Toni and Roeland Van Hout
   2005   *Statistics in Language Research: Analysis of Variance*. Berlin/New York: Mouton de Gruyter.
Rooij, Jaap De
   1991   Regionale variatie in het gebruik van *er* III. *Taal en Tongval* 43, 113−136.
Sanders, Ted
   1997   Psycholinguistics and the discourse level: Challenges for cognitive linguistics. Review of Gernsbacher, Morton Ann (ed.), *Handbook of psycholinguistics*, and Gernsbacher, Morton Ann and Talmy Givon (eds.), *Coherence in spontaneous text*. *Cognitive Linguistics* 8(3), 243−265.
Schermer-Vermeer, Everdina C.
   1985   De onthullende status van *er* in de generatieve grammatica. *Spektator* 15, 65−84.
Schmid, Monika
   1999   *Translating the Elusive. Marked Word Order and Subjectivity in English-German Translation*. Amsterdam/Philadelphia: John Benjamins.
Schutter, Georges de
   1974   *De Nederlandse zin. Poging tot Beschrijving van zijn Struktuur*. Brugge: De Tempel.
Szmrecsanyi, Benedikt
   2005   Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Journal of Corpus Linguistics and Linguistic Theory* 1(1), 113−149.
Swiggers, Piet, and Karel Van den Eynde
   1985   Distributie- en combinatiemogelijkheden van Nederlands *er*: Een studie in syntactische classificatie. *Linguistics in Belgium* 7, 67−86.

Toorn, Maarten C. van den
   1976   *Nederlandse Grammatica*. Groningen: Tjeenk Willink.
Verkuyl, Henk J., Geert E. Booij, Els H. C. Elffers-Van Ketel, Wim G. Klooster, Johannes H. J. Luif and Everdina C. Schermer-Vermeer
   1974   *Transformationele taalkunde*. Utrecht: Het Spectrum.