

Authors: Stefan Grondelaers, Dirk Speelman  
University/Affiliation: Radboud University Nijmegen, University of Leuven  
Email addresses: [S.Grondelaers@let.ru.nl](mailto:S.Grondelaers@let.ru.nl), [dirk.speelman@arts.kuleuven.be](mailto:dirk.speelman@arts.kuleuven.be)

## **Constructional near-synonymy, individual variation, and grammaticality judgments. Can careful design and participant ignorance overcome the ill reputation of questionnaires?**

### **Background**

Few native speakers of Dutch would acknowledge any difference between (1) and (2):

(Dutch)

- (1) In de asbak lag *er* een hagelkorrel.  
“In the ashtray there was a hailstone”
- (2) In de asbak lag een hagelkorrel.  
“In the ashtray was a hailstone”

If people would not regard post-verbal *er* “there” in the locative inversion construction as totally superfluous for comprehension, they would have great difficulties glossing its precise contribution to the adjunct-initial sentence. Interestingly, most professional linguists have fared little better with *er* “there”, arguably one of the most troublesome words in the Dutch language ever since it was put on the linguistic agenda by Brill in 1854 (*er*’s equivalents in other languages have excited comparable controversy).

What 15 years of data-based investigation has taught us is that *er*’s distribution is multi-factorially and probabilistically motivated. As a result, our task as variation analysts has been to identify *meaningful subgroups* in the data, viz. subgroups of the locative inversion construction which trigger *er* (constructions with temporal adjuncts, with semantically vague locative adjuncts, or with taxonomically unspecific main verbs), but also those sub-varieties of Dutch in which *er* is more frequent (notably Belgian Dutch and informal Dutch). A series of corpus-based regression analyses (Grondelaers, Speelman, & Geeraerts, 2002; 2008; Grondelaers, Geeraerts & Speelman 2007) to which these group factors were added revealed that *er*-preferences in the locative inversion construction can be correctly modelled in about 85 % of all cases. Building on the fact that most *er*-determinants are low-predictability contexts, and inspired by Bolinger’s (1977: 92) observation that English *there* signals insufficient contextual anticipation, we conducted a series of self-paced reading and eye-tracking experiments which confirmed that *er* is an *inaccessibility marker*. *Er* is inserted to deactivate inferences which are incompatible with an upcoming low-predictability subject: *ashtray* and *lay* in (1) anticipate “smoked-up tobacco products”, not hailstones (Grondelaers, Brysbaert, Speelman & Geeraerts, 2002; Grondelaers, Speelman, Drieghe, Brysbaert & Geeraerts, submitted).

While this function-based predictive success is clearly incompatible with the prevailing idea that *er*’s post-verbal distribution cannot be modelled (De Rooij, 1991), it is interesting to notice that we have never been able to fit *er*’s distribution beyond the 85 % success rate cited above. Observe in this respect that the distribution of the impersonal *il* in French locative inversion constructions can be predicted nearly categorically along similar functional lines (with success estimates going up to 97 %). The inevitable conclusion is that there remains *er*-variation we have not been able to model, either because there are as yet unidentified subgroups of locative inversion constructions or speakers of Dutch which manifest a significantly higher or lower *er*-probability, or because there is individual bias. The

latter is not improbable. Individual variation in *er*-preferences was pre-empirically reported in Geerts et al. (1984), De Rooij (1991), Haeseryn et al. (1997), and Van Boxtel (2003). In addition, we have argued that *er* is chosen in Belgian Dutch on the basis of the speaker's *subjective* assessment of the subject's predictability (Grondelaers, Speelman & Geeraerts, 2008), which entails that what is predictable for one speaker or listener need not be predictable for another speaker or listener.

## **Aim**

This paper will, therefore, focus on individual variation, no matter how theoretically and operationally diffuse that concept is. In our regression-based variationist approach to *er*, individual factors are considered as unfittable "noise", as the complement of the group variation which *can* be modelled. In our function-based (psycho-)linguistic approach to *er*, by contrast, individual preferences could be considered as the motivated consequence of the fact that the border between predictable and unpredictable is fuzzy and subjective.

What, then, is the proportion between motivated group variation, motivated individual variation, and non-motivated individual variation (noise)? More specifically, can the proportion of unaccountable "noise" be reduced by a more careful analysis of individual *er*-preferences? A corpus-based answer to this question requires materials in which idiosyncratic uses of *er* (resulting from non-standard assessments of subject predictability) are not eliminated, a condition which excludes virtually all newspaper materials. Since, in addition, sample sizes for spontaneous written or spoken data are too small for reliable analysis, and reliable demographic information is rarely available for these materials, we have no choice but to abandon responsibly collected corpus data, and elicit grammaticality judgments to measure preferential differences between individual listeners.

While introspective judgments continue to represent the standard data collection technique in generative linguistics, they have been under constant attack in other linguistic disciplines for their unreliability and instability (see Schütze 1996; Labov 1996, and Sampson 2007), and for the fact that they are almost never collected according to the standard methodology of psycholinguistic experiments (what raises most concern is the fact that participants are rarely ignorant of the research hypothesis, cf. Wasow & Clark 2005: 1483). The latter authors, however, have convincingly argued that a valid questionnaire design can overcome many of the criticisms against grammaticality judgments. Can we, therefore, develop a rating experiment which makes "predictions about usage which coincide perfectly (...) with what speakers are observed to utter and not to utter in spontaneous speech" (Sampson 2007: 188)? More specifically, can this experiment be designed so carefully that motivated group or individual *er*-preferences are not drowned in unaccountable noise?

## **Design**

A sizeable pool of native speakers (n = 181) rated the grammaticality of 12 short passages containing a locative inversion construction on a 7-point scale. All locative inversion constructions were presented in 2 versions, with and without *er*; participants rated either the version with or the version without *er*. In the 12 critical sentences, three low-predictability factors were orthogonally varied (*temporal* vs. *locative adjunct*, *vague locative* vs. *specific locative adjunct*, and *main verb "zijn"* vs. *more specific main verb*). In contrast to previous questionnaire-based approaches to *er*'s distribution (De Rooij 1991 and Van Boxtel 2003), we elicited ratings pertaining to the *global* grammaticality of the passages (not to the appropriateness of *er*), in order to direct attention away from the research question. Critical

passages were presented in two orders to gauge the impact of context on subject predictability and *er*-use.

To check the stability of the ratings, an identical copy of the original questionnaire was administered to the same participants three weeks later. At the end of the second trial, we explicitly asked what participants thought the scientific goal of the experiment had been: 75 % reported ignorance or failed to identify our interest in *er*. Since only 0,94 % of the participants correctly identified *er*'s post-verbal distribution as the exact goal of our enquiry, we can safely state that the absolute majority of participants was ignorant of our research hypothesis.

## Results & discussion

**Table 1: Linear regression on grammaticality judgments**

	Estimate	p-value
Intercept	6,02783	< 2e-16 ***
Adj_vagueloc	-0,41293	1.72e-05 ***
Adj_temp	-0,97596	< 2e-16 ***
Verb_zijn	-1,05573	< 2e-16 ***
Er1	-0,54229	1.17e-05 ***
Prov_antw	-0,12269	0.146156
Prov_limb	-0,32831	0.000299 ***
Prov_ovl	0,08039	0.486234
Prov_wvl	-0,39114	0.000533 ***
Secondtrial	-0.21140	0.000957 ***
Intentionunderstood	-0.08525	0.254584
Adj_vagueloc:zijn	0.61970	2.35e-08 ***
Adj_temp:zijn	-0.32073	0.003831 **
Er:adj_vagueloc	0.39810	0.000330 ***
Er:adj_temp	0.94517	< 2e-16 ***
Er:zijn	0.81111	< 2e-16 ***
Er:prov_antw	-0.02136	0.857994
Er:prov_limb	0.24744	0.053939
Er:prov_ovl	-0.56238	0.000578 ***
Er:prov_wvl	0.06754	0.672438
Er:secondtrial	0.17754	0.049817 *
Er:intentionunderstood	0.29690	0.005046 **
multiple R-Squared	0.1533	

A linear regression analysis on the ratings confirms that all interactions between *er* and the low-predictability factors are highly significant: *er* considerably reduces the ungrammaticality experienced when locative inversion constructions contain temporal adjuncts, semantically vague locative adjuncts, or taxonomically unspecific main verbs. While these findings confirm the correctness of the research hypothesis, the interaction “Er:secondtrial”, which indicates that *er* is preferred significantly more often ( $p = 0.049$ ) in the second trial of exactly the same questionnaire, strongly suggests that some *er*-variation is not functionally motivated. The low R-Squared (0.1533) raises even more reasons for concern: the evident correctness of the research hypothesis and the careful design of the questionnaire cannot prevent that only a minimal percentage of variation in the grammaticality judgments is motivated by our manipulations. A reliability analysis on the ratings further indicates that a satisfactory

Cronbach's Alpha (> .9) is reached only when all raters (> 40 in each of the 4 conditions) are included in the analysis, which suggests massive individual variation.

The only valid conclusion that can be drawn at this moment is that even when participants are ignorant of the research hypothesis, grammaticality judgments are "too shifty and variable (both from speaker to speaker and from moment to moment)" (Schutze 1996: 3) to reveal much beyond what we already know from other data-collection techniques. Although we have not yet fully analyzed the effect of presentation order – we are currently experimenting with predictability estimates (n-gram probability) to gauge the extent to which the preceding context in the different presentation orders makes subjects more or less predictable –, and although Belgian Dutch is known to manifest more individual variation than Netherlandic Dutch on account of its delayed standardization (Grondelaers et al.: 2008), we fear that the deluge of individual variation observed is technique-related: the inevitable conclusion is that reliable *er*-intuitions cannot properly be elicited in a grammaticality judgment experiment.

## References

- Bolinger, Dwight. 1977. *Meaning and Form*. London: Longman.
- De Rooij, J. 1991. Regionale variatie in het gebruik van *er* III. *Taal en Tongval* 43, 113-136.
- Geerts, G., W. Haeseryn, J. de Rooij & M.C. van den Toorn. 1984. *Algemene Nederlandse Spraakkunst*. Groningen/Leuven: Wolters-Noordhoff.
- Grondelaers, Stefan, Marc Brysbaert, Dirk Speelman & Dirk Geeraerts. 2002. *Er* als accessibility marker: on- en offline evidentie voor een procedurele interpretatie van presentatieve zinnen. *Gramma/TTT* 9/1, 1-22.
- Grondelaers, S., D. Geeraerts & D. Speelman (2007). A case for a Cognitive corpus Linguistics. In Gonzales-Marques, M., I. Mittelberg, S. Coulson & M. J. Spivey (eds.), *Methods in Cognitive Linguistics*, 149-169. Amsterdam/Philadelphia: John Benjamins.
- Grondelaers, Stefan, Dirk Speelman & Dirk Geeraerts. 2002. Regressing on *er*. Statistical analysis of texts and language variation. In Annie Morin & Pascale Sébillot (eds.), *Proceedings of the 6th International Conference on the Statistical Analysis of Textual Data*, 335-346. Rennes: Institut National de Recherche en Informatique et en Automatique.
- Grondelaers, Stefan, Dirk Speelman, Denis Drieghe, Marc Brysbaert & Dirk Geeraerts (submitted). Indefinite reference processing: Converging on- and offline evidence for predictive inferencing and remedial cueing. Submitted to *Acta Psychologica*.
- Grondelaers, Stefan, Dirk Speelman & Dirk Geeraerts. 2008. National variation in the use of *er* "there". Regional and diachronic constraints on cognitive explanations. To appear in Gitte Kristiansen & René Dirven (eds.), *Cognitive sociolinguistics: Language variation, cultural models, social systems*. Berlin: Mouton de Gruyter.
- Labov, W. 1996. When intuitions fail. In L. McNair, K. Singer, L. Dolbrin, and M. Aucon (eds.), *Papers from the Parasession on Theory and Data in Linguistics*, 77-106. Chicago: Chicago Linguistic Society.
- Sampson, Geoffrey R. 2007. Grammar without grammaticality. *Corpus Linguistics and Linguistic Theory* 3/1, 1-32.
- Schütze, Carson T. 2006. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: Chicago University Press.
- Van Boxtel, S., P.-A. Coppen & T. Bongaerts (2003). Veel is (er) nog onduidelijk gebleken. Factoren in de keuze voor vervangende subjecten in het Nederlands. *Nederlandse Taalkunde* 8, 181-198.