Syntactic Haplology and the Dutch Proform *er*

Gert Webelhuth and Olivier Bonami

Dutch has four proforms spelled *er* whose surface distribution is unusual and intriguing. In some cases, more than one *er* can be overt but in others a single overt *er* expresses several functions at once, even though they may be syntactically (e.g. argument vs. modifier) and/or semantically (e.g. referential vs. non-referential) incompatible. Transformational analyses have typically taken recourse to deletion operations to capture this distribution. The purpose of our talk is to demonstrate that the behavior of *er* can be derived non-transformationally through the interaction of constraints on argument structure and linear structure.

Broekhuis (2013, p. 338) states that on *er* "conflicting judgments can be found in the literature." Due to space limitations on the abstract, we decided to limit ourselves to a single paradigm of judgments, namely that in Broekhuis' lengthy and systematic study. Unless stated otherwise, all examples below originate in this work. We may have added emphasis or modified the emphasis in the original for expository purposes.

We build on the insights of previous work on *er* in HPSG, but extend and in this process also depart from it. We go beyond the treatment in Bouma (2001), whose theory does not cover quantitative *er*. The theory in Campbell-Kibler (2001) is meant to account for different judgments than those considered here. Since the author draws data from different works in the literature, it is unclear that the data reflects the judgments of any single native speaker.

1 The four *er*s

(1) gives an example of each *er*:[1]

(1) a. *Er*$_X$ loopt een man op straat.
there walks a man in the.street

b. Jan wacht *er*$_P$ al tijden **op**.
Jan waits there for ages for

c. Jan staat *er*$_L$ al.
Jan stands there already

d. Jan heeft *er*$_Q$ [NP **drie** [*e*]]
Jan has there three

The *er* in (1a) is existential. This is the only *er* which is permitted to occur in the first position of a declarative main clause.[2] In (1b), we see the pronominal *er*, which serves as the referential argument of the sentence-final preposition *op*. The *er* in the third example is a locational proform corresponding to the English adverb *there*. (1d), finally, contains the quantitative proform *er*. It is a partitive element related to the following NP, similar to Italian *ne* and French *en*.

Dutch is a verb-second language in declarative main clauses and a verb-final language in embedded clauses. For the purposes of this abstract, we will say that the first constituent of (declarative) main clauses occupies the prefield and the constituents following the finite verb in main clauses or the complementizer in embedded clauses up to the sentence-final verb or verb particle occupy the middle field:

(2) a. Main clause: Prefield V_{finite} Middle field (Verb(s))

b. Subordinate clause: Complementizer Middle field Verb(s)

[1] In the examples, we label each *er* with a subscript, as follows: X = existential, P = pronominal, Q = quantitative, and L = locational.

[2] The presence of expletive *er* correlates with a definiteness effect. This will play no role in our account.

2 The Data

Due to space limitations, we will only deal with finite sentences. In this section we show that clauses with existential *er* in the prefield behave slightly differently from clauses which do not have a prefield or where some element other than existential *er* occupies the prefield. We begin with the latter because the data is somewhat less complex.

2.1 Clauses without existential *er* in the prefield

We will illustrate the behavior of *er* in clauses without an existential *er* in the prefield with examples from complementizer-introduced subordinate clauses which lack a prefield altogether. The sentences in (3) illustrate in exemplary fashion that sentences with two overt *er*'s in the middle field are ungrammatical, no matter which types of *er* cooccur:[3]

(3) a. * dat er_X er_L gedanst wordt.
that there there danced is
Intended reading: 'People are dancing there.'

b. * dat Jan er_P er_L *over* praatte.
that Jan there there about talked
'that Jan talked about it there.'

c. * dat Jan er_Q er_L $[_{NP}$ twee [*e*]] gezien heeft.
that Jan there there two seen has
'that Jan saw two [e.g., rats] there.'

However, there are sentences with an overt *er* where a second *er* is implicit, as in (4b)-(4d):[4]

(4) a. dat er_X gisteren drie potloden op tafel lagen.
that there yesterday three pencils on the.table lay
'that there were three pencils lying on the table yesterday.'

b. dat er_{XP} gisteren drie potloden **op** lagen.
that there yesterday three pencils on lay
'that there were three pencils lying on it yesterday.'

c. dat er_{XQ} gisteren **drie** op tafel lagen.
that there yesterday three on the.table lay
'that there were three lying on the table yesterday.'

d. dat er_{XL} veel mensen **wonen**.
that there many people live
'that many people live there.'

(4a) is an existential sentence with er_X as the first constituent of the middle field. In (4b) the preposition *op* needs a complement and (4c) contains the elliptical NP *drie* which requires a quantitative *er* to be present. Finally, the verb *wonen* in (4d) needs a locational complement. Yet, in each case only a single *er* surfaces.

Additional combinations of explicit and implicit *er* are possible. In (5), there can only be one overt pronominal *er*, even though each of the two stranded prepositions *mee* and *uit* needs a complement:

(5) a. Jan heeft de sleutel **met** een tang **uit** het slot gehaald
Jan has the key with a pair.of.tongs out.of the lock taken
"Jan took the key out of the lock with pliers."

b. Jan heeft er_P de sleutel **mee uit** gehaald.

[3] Neeleman and van de Koot (2006) accept sentences with two *er*s in the middle field as long as the *er*s are not adjacent. Hans Broekhuis considers these examples ungrammatical (personal communication). The theory presented below is only meant to cover Broekhuis' judgments!

[4] The notation "er_{XP}" in (4b) labels an overt er_X in a clause with an implicit er_P, and the same for any two-letter subscript.

c. * Jan heeft er_P er_P de sleutel **mee uit** gehaald.

Example (6a) shows that it is also possible to conflate two instances of quantitative *er*:

(6) Iedere student heeft een onvoldoende gekregen ...
every student has an unsatisfactory mark gotten
"Every student got an unsatisfactory mark ..."
a. ... en [$_{NP}$ **drie** *e*] hebben er_Q zelfs [$_{NP}$ **twee** *e*].
and three have there even two
"... and three even got two."

Given that various binary combinations of *er*s can be represented by a single overt *er*, it is perhaps not surprising that more than one *er* can remain implicit in a sentence. Broekhuis (2013, p. 362) gives the examples in (7) and comments on them as follows [we have adjusted his example numbering to ours]:

> In [(7a)], *er* only has an expletive function. In [(7b)], there are two quantitative noun phrases, so *er* simultaneously performs the expletive function once and the quantitative function twice. In [(7c)], R-pronominalization has been applied to the PP *uit de boekenkast*, so that *er* performs the pronominal function on top of the other functions in [(7b)].

(7) a. dat er twee studenten drie boeken uit de boekkast gehaald hebben.
that there two students three books out.of the bookcase fetched have
b. dat er [NP **twee** e] [NP **drie** e] uit de boekenkast gehaald hebben.
c. dat er [NP **twee** e] [NP **drie** e] **uit** gehaald hebben.

2.2 Clauses with existential *er* in the prefield

Main clauses without an *er* in the pre-field show the same pattern that we saw above: the same combinations of *er*s can cooccur as in subordinate clauses, but only one overt *er* is permitted in the middle field.

Main clauses with an existential *er* in first position display a slightly different pattern. Recall that existential *er* but none of the other *er*s can occupy the prefield. Er_X in the prefield can cooccur with all the others, so the only question that arises is whether the latter are overt or not.

In this respect, locational and pronominal *er* differ from quantitative *er*. The first two must remain implicit. Both of the sentences below become grammatical if the *er* in the middle field is left out:

(8) a. * Er_X wordt er_L morgen gedanst.
there is there tomorrow danced
b. * Er_X wordt er_P morgen *over* gesproken.
there is there tomorrow about spoken

With quantitative *er*, the opposite is true. It must be spelled out separately in the middle field:

(9) Er_X zijn er_Q gisteren [$_{NP}$ **twee** [*e*]] gestolen.
there have.been there yesterday two stolen

There is a slight wrinkle: Hans Broekhuis (p.c.) has informed us that when there are two quantitative NPs in a clause with existential *er* in the prefield, then it is possible (in fact, necessary) for only a single quantitative *er* to be spelled out in the middle field. The er_Q in the second half of the clause below is related both to *een paar* and to *twee*.

(10) [Er hebben veel studenten een onvoldoende gekregen] en [er hebben er_Q
there(E) have many students an unsatisfactory_mark gotten and there(E) have there(Q)
[**een paar** *e*] zelfs [**twee** *e*] gekregen.
a couple even two gotten

3 Analysis

We are in need of a grammatical mechanism that makes it possible for an overtly expressed *er* to force the absence of additional overt *er*s.

We begin by postulating a general lexical identifier *er* for the *er*-family of words and a specific subtype for each different *er*: *er-X*, *er-Q*, *er-P*, and *er-L*.

Second, we associate the *er*s with linear options: all of them can occur in the middle field; existential *er* can additionally occur in the prefield:

(11) **Linear constraints on the 4 *er*s**

$$\begin{bmatrix} \textit{word} \\ \text{SYNSEM} \begin{bmatrix} \text{FLD} \ \textit{pre-fld} \vee \textit{mid-fld} \end{bmatrix} \end{bmatrix} \qquad \begin{bmatrix} \textit{word} \\ \text{SYNSEM} \begin{bmatrix} \text{FLD} \ \textit{mid-fld} \end{bmatrix} \end{bmatrix}$$

Existential *er* **Non-existential *er*s**

We assume that each of the *er*s treated in this abstract occurs on the ARG-S list of the finite verb, either directly, through argument extension (locational *er*), or through argument raising (quantitative and pronominal *er*).[5] Moreover, we adopt a bifurcation between canonical synsems, which can be realized as an actual sign, and noncanonical synsems, which may occur on ARG-S but are not realized explicitly in local syntax (Miller and Sag (1997)). Since *er*-X is always overt, it must be selected as a *canon-synsem*, whereas the remaining *er*s can alternatively be selected as *canon-synsem* and *pro-synsem*.

With these preliminaries taken care of, we now formulate three simple constraints on the expression of *er*s on the ARG-S and COMPS lists of finite verbs.

The first constraint requires that at least one *er*-argument of a finite verb be expressed:

(12)
$$\begin{bmatrix} \text{HEAD} & \text{V}[\textit{fin}] \\ \text{ARG-ST} & \left\langle \begin{bmatrix} \textit{synsem} \\ \text{LID} \ \ \textit{er} \end{bmatrix} \right\rangle \bigcirc \textit{list} \end{bmatrix} \longrightarrow \begin{bmatrix} \text{ARG-ST} & \left\langle \begin{bmatrix} \textit{canon-synsem} \\ \text{LID} \ \ \textit{er} \end{bmatrix} \right\rangle \bigcirc \textit{list} \end{bmatrix}$$

This constraint is epistemologically plausible, as it entails that every verb with one or more implicit *er*s must signal this state of affairs by overtly realizing one *er*.

The second constraint expresses the generalization [exemplified by the contrast between (8) and (9)] that if a verb expresses two *er*s, then one of them needs to occupy the prefield (which only *er*-X is permitted to do) and the other one must be quantitative *er*:

(13)
$$\begin{bmatrix} \text{HEAD} & \text{V}[\textit{fin}] \\ \text{COMPS} & \left\langle \begin{bmatrix} \textit{synsem} \\ \text{LID} \ \ \textit{er} \end{bmatrix} \right\rangle \bigcirc \left\langle \begin{bmatrix} \textit{synsem} \\ \text{LID} \ \ \textit{er} \end{bmatrix} \right\rangle \bigcirc \textit{list} \end{bmatrix} \longrightarrow \begin{bmatrix} \text{COMPS} & \left\langle \begin{bmatrix} \textit{synsem} \\ \text{LID} \ \ \textit{er} \\ \text{FLD} \ \ \textit{pre-fld} \end{bmatrix} \right\rangle \bigcirc \left\langle \begin{bmatrix} \textit{synsem} \\ \text{LID} \ \ \textit{er-Q} \end{bmatrix} \right\rangle \bigcirc \textit{list} \end{bmatrix}$$

In conjunction with the lexical entries and constraints introduced above, the third and final constraint produces the haplological effect that we illustrated in section 2. It simply states that a finite verb at most selects one *er*-complement in the middle field:

(14)
$$\begin{bmatrix} \text{HEAD} & \text{V}[\textit{fin}] \\ \text{COMPS} & \left\langle \begin{bmatrix} \textit{synsem} \\ \text{LID} \ \ \textit{er} \\ \text{FLD} \ \ \textit{mid-fld} \end{bmatrix} \right\rangle \bigcirc \textit{list} \end{bmatrix} \longrightarrow \begin{bmatrix} \text{COMPS} & \left\langle \begin{bmatrix} \textit{synsem} \\ \text{LID} \ \ \textit{er} \\ \text{FLD} \ \ \textit{mid-fld} \end{bmatrix} \right\rangle \bigcirc \textit{list}(\neg \begin{bmatrix} \textit{synsem} \\ \text{LID} \ \ \textit{er} \\ \text{FLD} \ \ \textit{mid-fld} \end{bmatrix}) \end{bmatrix}$$

[5] With locational *er*, we adopt the treatment in Bouma (2001). The author argues against the argument raising approach of pronominal *er*, but we believe that this approach can be made to work.

4 Illustration

Below, we give the relevant parts of the lexical entry of the finite verb *wordt* in sentence (3a):

(15)
$$\begin{bmatrix} \textit{word} \\ \textsc{head} \quad \text{V}[\textit{fin}] \\ \textsc{comps} \quad \boxed{1}\left\langle \text{V}[\textit{pastp}],\ \begin{bmatrix} \textit{canon-synsem} \\ \textsc{lid} \quad \textit{er-X} \\ \textsc{fld} \quad \textit{mid-fld} \end{bmatrix}, \begin{bmatrix} \textit{canon-synsem} \\ \textsc{lid} \quad \textit{er-L} \\ \textsc{fld} \quad \textit{mid-fld} \end{bmatrix} \right\rangle \\ \textsc{arg-s} \quad \boxed{1}\left\langle \text{V}[\textit{pastp}],\ \begin{bmatrix} \textit{canon-synsem} \\ \textsc{lid} \quad \textit{er-X} \\ \textsc{fld} \quad \textit{mid-fld} \end{bmatrix}, \begin{bmatrix} \textit{canon-synsem} \\ \textsc{lid} \quad \textit{er-L} \\ \textsc{fld} \quad \textit{mid-fld} \end{bmatrix} \right\rangle \end{bmatrix}$$

In order to realize the two overt *er*s in (3a), the word must have two canonical *er*-complements. However, the COMPS list of (15) violates two constraints at once: (i) it violates (13), because this constraint requires one of two overt *er*s to be in the pre-field and the other one to be *er*-Q—neither condition is fullfilled; and (ii) it violates (14), because it has two mid-field *er*-complements, when the constraint permits at most one. Hence, (15) is not a legitimate word, which derives the ungrammaticality of (3a).

The verb *wonen* of sentence (4d) is required to have the following structure in this sentence:

(16)
$$\begin{bmatrix} \textit{word} \\ \textsc{head} \quad \text{V}[\textit{fin}] \\ \textsc{comps} \quad \left\langle \boxed{1} \begin{bmatrix} \textit{canon-synsem} \\ \textsc{lid} \quad \textit{er-X} \\ \textsc{fld} \quad \textit{mid-fld} \end{bmatrix} \right\rangle \\ \textsc{arg-s} \quad \left\langle \boxed{1} \begin{bmatrix} \textit{canon-synsem} \\ \textsc{lid} \quad \textit{er-X} \\ \textsc{fld} \quad \textit{mid-fld} \end{bmatrix}, \begin{bmatrix} \textit{pro-synsem} \\ \textsc{lid} \quad \textit{er-L} \\ \textsc{fld} \quad \textit{mid-fld} \end{bmatrix} \right\rangle \end{bmatrix}$$

(16) satisfies all our constraints: (i) *er*-X is permitted to occur in the middle field by its lexical entry in (11), (ii) by overtly expressing one of its *er*-arguments it obeys (12), (iii) it obeys (13) vacuously, because it doesn't have two *er*-complements, and, finally, (iv) it satisfies (14), as it expresses no more than one *er*-complement in the middle field. No constraints are violated in (4d) and the sentence is correctly predicted to be grammatical.

In sum: by formulating relatively simple constraints on the ARG-S and COMPS lists of finite verbs in Dutch, we have formally captured the essence of *er*-haplology in this language. It remains to be seen whether other cases of syntactic haplology can be handled in a similar fashion.

References

Bouma, G. (2001). 'Argument Realization and Dutch R-Pronouns: Solving Bech's Problem without Movement or Deletion'. In R. Cann, C. Grover, and P. Miller (eds.), *Grammatical Interfaces in HPSG*. Stanford: CSLI Publications.

Broekhuis, H. (2013). *Syntax of Dutch: Adpositions and Adpositional Phrases*. Amsterdam: Amsterdam University Press.

Campbell-Kibler, K. (2001). 'Bech's Problem, Again: Using Linearization on Dutch R-Pronouns'. In F. V. Eynde, L. Hellan, and D. Beermann (eds.), *The Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*. Stanford: CSLI Publications, 87—102.

Miller, P. and Sag, I. A. (1997). 'French clitic movement without clitics or movement'. *Natural Language and Linguistic Theory*, 15:573–639.

Neeleman, A. and van de Koot, H. (2006). 'Syntactic Haplology'. In M. Everaert and H. van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, vol. 4. London: Blackwell, 685—710.