

# Multifunctional Dutch ‘er’

Stephen M. Jones

University of Groningen

Proceedings of the LFG’20 Conference

On-Line

Miriam Butt, Ida Toivonen (Editors)

2020

CSLI Publications

pages 148–168

<http://csli-publications.stanford.edu/LFG/2020>

Keywords: Dutch, pronoun, lexical specification, correspondence function, information structure

Jones, Stephen M. 2020. Multifunctional Dutch ‘er’. In Butt, Miriam, & Toivonen, Ida (Eds.), *Proceedings of the LFG’20 Conference, On-Line*, 148–168. Stanford, CA: CSLI Publications.



## Abstract

This paper presents an LFG analysis of two aspects of the Dutch pronoun *er*: its ability to provide multiple grammatical functions within a single clause; and the constraints on the position of *er* within a clause and the combinations of functions that are allowed in a given position. The analysis rests on interactions between string ordering, c-structure, f-structure and information structure constraints. The general lexical specification for *er* comprises a core together with optional subspecifications: each instance of *er* generates its own lexical specification to satisfy other constraints in the clause. The paper introduces the proposal that a c-structure node may project a set of f-structures, each of which shares its structure with a distinct element of the overall clausal f-structure.

## 1 Introduction

The Dutch pronoun *er* provides a challenge to resource-based grammar theories because of its ability to introduce potentially unlimited resources. For LFG there is a further challenge because of the one-to-many mapping between an instance of *er* and grammatical functions within a clause.

In this paper I present data on the distribution of *er*, illustrating its contribution to the meaning of a sentence and the interactions between these meanings and constraints on distribution. I then propose an account that addresses these challenges with two innovations: allowing a c-structure node to project a set of f-structures rather than a single f-structure, and using a template lexical specification for *er*, which is instantiated for a particular element of the string depending on the other properties of the clause.

*Er* has four distinct pronominal functions. In this paper, I follow Odijk's (1993) categorisation of them:

- i. existential  $er_X$  occurs with an indefinite subject or subjectless passive;
- ii. locative  $er_L$  is a locative adverbial pronoun;
- iii. prepositional  $er_P$  is a non-human prepositional object pronoun; and
- iv. quantitative  $er_Q$  is a partitive pronoun comparable with French *en*.

There are many descriptions of the distribution of *er*, and accounts of its syntactic constraints, including Bech (1952), Bennis (1986), van Riemsdijk (1978), Odijk (1993), Neeleman and van de Koot (2006), Donaldson (2008), Grondelaers et al. (2009), Klooster (2014), Webelhuth and Bonami (2019). Distributional constraints interact with the functions expressed by an instance of *er* within a given clause. A strong constraint is that *er* generally occurs only once, and maximally

---

<sup>†</sup>My thanks go to the numerous native speakers of Dutch who advised me on example sentences, and to the anonymous reviewers, whose comments and suggestions have improved this paper.

twice, within a single clause. This means that one instance of *er* may carry more than one function simultaneously, and as a pronoun may refer to more than one distinct antecedent. All pairwise combinations of functions have been observed in a single instance of *er*, and combinations of three or more functions are possible.

*Er* is a member of the family of what have been termed “R-pronouns” (van Riemsdijk 1978), which includes *daar* ‘there’, *hier* ‘here’, and *waar* ‘where’. Some of the functions of *er* can be provided by *daar* or *hier*. However, *er* is semantically less weighty, in that it does not contribute deictic information and cannot be phonologically emphasised.

Despite the numerous accounts of *er* in different theoretical frameworks, to date a treatment in LFG is lacking. In the remainder of the paper I discuss the constraints on the distribution of the functions of *er*, and then propose and test a lexical specification that can account for the distribution.

### 1.1 Dutch clause structure

In describing the structure of Dutch clauses I adopt the model provided by Haeseryn et al. (1997), in which there are two “poles” around which the other elements are ordered. Only one constituent can occupy the prefield and there are constraints on the types of constituents in the postfield. A diagram of the structure is given in Figure 1.

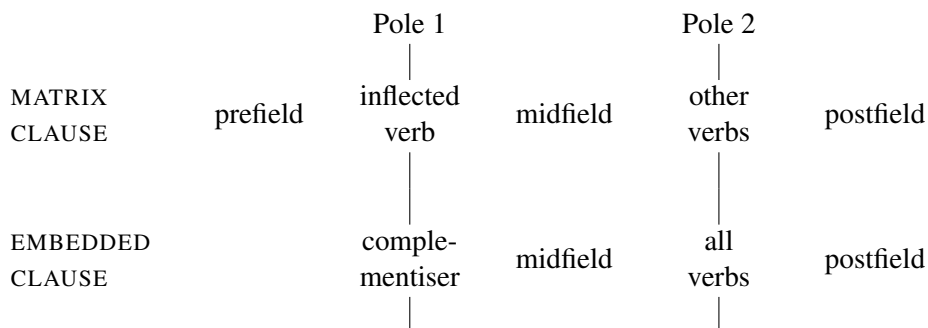


Figure 1: Assumed Dutch clause structure (Haeseryn et al. 1997)

This topological description of phrase structure is similar to other Germanic languages, including LFG discussions of Icelandic and Swedish (e.g. Sells 2001, 2005, Booth and Schätzle 2019). At a more detailed level, there are differences between Dutch and these other languages with respect to the behaviour of expletives and the ordering constraints in the midfield, which are outside the scope of this paper. For the purposes of accounting for the constraints on *er*, I make the following assumptions.

In terms of phrase structure, I assume that Pole 1 for matrix clauses is the head of IP, which is occupied by the inflected verb. The prefield, which I assume to be

the specifier of IP, Spec-IP, is reserved for a constituent that is prominent either syntactically or at information structure. Canonically it is the subject position, but information structural constraints frequently result in either the focus or topic of an utterance occupying the prefield, with the subject occurring in the midfield. The order of constituents in the midfield is determined by interactions between category, grammatical function, and information structure considerations, which are not discussed in this paper. Also, the lack of case marking on Dutch nouns means that grammatical functions may not be fully determined by the syntax. As a simplification of c-structure, I assume that the midfield is contained within a VP daughter of I', and that midfield dependents of the verb attach at V'. All dependents of the verb, whether daughters of V' or Spec-IP, carry the functional constraint ( $\uparrow$  GF) =  $\downarrow$ . For numerically quantified noun phrases, I follow the grammar for German available from INESS XLE-Web<sup>1</sup> (Rosén et al. 2012) in assuming that numbers project a NumP, which occupies the specifier of NP and contribute a NUM feature at f-structure via the constraint ( $\uparrow$  NUM) =  $\downarrow$ .

## 2 The distribution of *er*

In this section I describe the four core functions of *er* and their distribution in sentences where there is one instance of *er* that carries only one function. I then discuss the constraints that apply where a single instance of *er* carries more than one function.

### 2.1 Core functions of *er*

#### 2.1.1 Presentative *er<sub>X</sub>*

Presentative *er<sub>X</sub>* appears in the prefield of a matrix clause where the subject is indefinite (1) or as the subject of an impersonal passive sentence (2) where there is no other prominent constituent in the clause. Where the prefield is occupied by a non-subject constituent, or in embedded clauses, *er<sub>X</sub>* is optional if there is an explicit subject (3) but must appear where there is no other expressed subject (4). Grondelaers et al. (2009) identified a processing advantage for sentences with midfield *er<sub>X</sub>* where the subject is semantically unexpected in context.

- (1) *Er<sub>X</sub> staan nog teveel mensen aan de zijlijn.*  
 ER stand yet too.many people on the sideline  
 “There are still too many people standing on the sidelines.”<sup>2</sup>

<sup>1</sup><http://clarino.uib.no/iness>

<sup>2</sup><https://www.rwm.nl/kringloop/hetgoed> (all URLs accessed on 2020-10-31)

- (2) *Er<sub>X</sub> werd gedanst en gedronken.*  
ER became danced and drunk.  
“There was dancing and drinking.”<sup>3</sup>
- (3) *...aan de haak hing (er<sub>X</sub>) een peer...*  
...on the hook hung (ER) a pear...  
“...on the fishhook (there) hung a pear...”<sup>4</sup>
- (4) *Waar wordt \*(er<sub>X</sub>) deze winter gebaggerd?*  
Where PASS.PRS \*(ER) this winter dredged  
“Where will there be dredging this winter?”<sup>5</sup>

Existential *er<sub>X</sub>* does not alternate with *daar/hier*. Where *daar* occupies the prefield, *er<sub>X</sub>* is still possible in the midfield (5).

- (5) *Daar<sub>Loc</sub> kwam er<sub>X</sub> al een wet in 2006.*  
There came ER already a law in 2006  
“A law had already been passed *there* (NL) in 2006.”<sup>6</sup>

### 2.1.2 Locative *er<sub>L</sub>*

Locative *er<sub>L</sub>* replaces a prepositional, nominal, or adverbial locative phrase (6) whose grammatical function can be either an argument or an adjunct. It can be replaced by *daar/hier* (7). If *er<sub>L</sub>* is the only function of *er*, it cannot occur in the prefield (8). However, locative *daar* is possible in the prefield (9).

- (6) *Ik ben er<sub>L</sub> nooit geweest, en het trekt me ook niet.*  
I am ER never been, and it attracts me also not  
“I’ve never been there, and it doesn’t attract me either.”<sup>7</sup>
- (7) *Ik ben daar nooit geweest, en het trekt me ook niet.*  
I am there never been, and it attracts me also not  
“I’ve never been there, and it doesn’t attract me either.”
- (8) *\*Er<sub>L</sub> ben ik nooit geweest, en het trekt me ook niet.*  
ER am I never been, and it attracts me also not  
(intended) “I’ve never been there, and it doesn’t attract me either.”
- (9) *Daar ben ik nooit geweest, en het trekt me ook niet.*  
there am I never been, and it attracts me also not  
“I’ve never been *there*, and it doesn’t attract me either.”

<sup>3</sup><https://dorpskrantdeknipe.nl/vier-vijf-mei>

<sup>4</sup>Grondelaers et al. (2009)

<sup>5</sup><https://www.waterschaprivierenland.nl/waar-wordt-er-deze-winter-gebaggerd>

<sup>6</sup><https://nl.wikipedia.org/wiki/Gerechtstolk>

<sup>7</sup><https://nl.toluna.com/opinions/2513744/Libelle-Zomerweek>

### 2.1.3 Pronominal *er<sub>P</sub>*

Pronominal *er<sub>P</sub>* appears if the sentence requires a pronominal non-human prepositional object (10): *het* ‘it’ following the preposition is ungrammatical here (11). In Netherlands Dutch, the preposition associated with *er<sub>P</sub>* usually appears at the end of the midfield.<sup>8</sup> Similarly to *er<sub>L</sub>*, *er<sub>P</sub>* can be replaced by *daar/hier* (12). In the prefield, *er<sub>P</sub>* as the sole function of *er* is unacceptable (13), but *daar/hier* is possible here (14).

- (10) *Ja soms kan je er<sub>P</sub> trots op zijn.*  
Yes sometimes can you ER proud on be  
“Yes, sometimes you can be proud of it.”<sup>9</sup>
- (11) \**Ja soms kan je trots op het zijn.*  
Yes sometimes can you proud on it be  
(intended) “Yes, sometimes you can be proud of it.”
- (12) *Ja soms kan je daar trots op zijn.*  
Yes sometimes can you there proud on be  
“Yes, sometimes you can be proud of it.”
- (13) \**Er<sub>P</sub> kan je trots op zijn.*  
ER can you proud on be  
(intended) “You can be proud of it.”
- (14) *Daar kan je trots op zijn.*  
There can you proud on be  
“You can be proud of that.”

### 2.1.4 Quantitative *er<sub>Q</sub>*

Quantitative *er<sub>Q</sub>* appears with headless quantified (15) or restricted noun phrases (16). It cannot be replaced by *daar/hier* (17) and must appear in the midfield (18).

- (15) *De speler van Veenhuizen maakte er<sub>Q</sub> drie.*  
The player of Veenhuizen made ER three  
“The Veenhuizen player scored three.”<sup>10</sup>
- (16) *Bovendien zijn er<sub>Q</sub> die wél de titel maar geen Michelinster hebben.*  
Furthermore are ER who certainly the title but no Michelin.star  
have  
“Then there are those who do have the title but no Michelin star.”<sup>11</sup>

<sup>8</sup>[https://taaladvies.net/taal/advies/vraag/1340/er\\_op\\_erop/](https://taaladvies.net/taal/advies/vraag/1340/er_op_erop/)

<sup>9</sup><https://uitleganimatie.studiosteenproducties.nl/blog/trots-op-je-pot>

<sup>10</sup><https://dekrantnieuws.nl/topscorers-moes-maakt-er-drie/>

<sup>11</sup><https://www.bndestem.nl/moerdijk/chefkok-vista-in-willemstad-krijgt-meestertitel-ik-wil-het-hoogst-haalbare~a5428451/>

- (17) \* *De speler van Veenhuizen maakte daar drie.*  
 The player of Veenhuizen made there three  
 (intended) “The Veenhuizen player scored three.”
- (18) \* *Daar/er<sub>Q</sub> maakte de speler van Veenhuizen drie.*  
 There/ER made the player of Veenhuizen three  
 (intended) “The Veenhuizen player scored three.”

### 2.1.5 Summary: single-function *er*

In summary, when *er* fulfils a single function, only *er<sub>X</sub>* is possible in the prefield, but all functions are possible in the midfield. Of the four functions, only *er<sub>L</sub>* and *er<sub>P</sub>* can be substituted by *daar* or *hier*: in these cases, *daar/hier* may occupy the prefield.

## 2.2 Single instances of *er* serving multiple functions

Where possible, a single instance of *er* in a clause provides all the functions. However, constraints apply to the prefield such that not all function combinations are possible there.

### 2.2.1 *Er* in the prefield

Where *er<sub>X</sub>* occurs in the prefield, it must also provide the functions for *er<sub>L</sub>* (19) and *er<sub>P</sub>* if these are present in the clause (20). Here, a second instance of *er* in the clause is ungrammatical. However, *er<sub>Q</sub>* is not compatible with prefield *er<sub>X</sub>* and must be expressed separately (21). This the only acceptable case for prefield *er* and midfield *er* in the same clause.

- (19) a. *Er<sub>XL</sub> woont ook vrijwel niemand.*  
 ER lives also almost niemand.  
 “Pretty much nobody lives there.”<sup>12</sup>
- b. \* *Er<sub>X</sub> woont er<sub>L</sub> ook vrijwel niemand.*  
 ER lives ER also almost niemand.  
 (intended) “Pretty much nobody lives there.”
- (20) a. *Er<sub>XP</sub> heeft iemand over nagedacht voor ons.*  
 ER has someone over thought.about for us.  
 “Someone has thought that through for us.”<sup>13</sup>
- b. \* *Er<sub>X</sub> heeft er<sub>P</sub> iemand over nagedacht voor ons.*  
 ER has ER someone over thought.about for us.  
 (intended) “Someone has thought that through for us.”

<sup>12</sup><https://www.weerwoord.be/m/2582768>

<sup>13</sup>[https://gathering.tweakers.net/forum/list\\_messages/1894879](https://gathering.tweakers.net/forum/list_messages/1894879)

- (21) a. \**Er<sub>XQ</sub> waren twee (in de zaal)*  
 ER were two (in the room).  
 (intended) “There were two (of them in the room).”<sup>14</sup>
- b. *Er<sub>X</sub> wonen er<sub>Q</sub> 53 in Kortrijk.*  
 ER live ER 53 in Kortrijk  
 “53 (of them) live in Kortrijk.”<sup>15</sup>

### 2.2.2 Midfield *er* carrying two functions

In clauses where *er* occurs only in the midfield, it carries all the functions required by the clause. Bennis (1986) demonstrates this using lexical substitutions and valency constraints for the combinations *er<sub>XL</sub>*, *er<sub>XP</sub>*, and *er<sub>XQ</sub>*, where *er* has only one pronominal antecedent. However, clauses where a single midfield *er* has two or more distinct antecedents are also possible, and the corresponding clauses with multiple instances of *er* in the midfield are almost always rejected (22). Corpus evidence suggests that a second midfield *er* is observed infrequently where it provides *er<sub>P</sub>* for a subsequent clausal antecedent, and where the *er* is written as a single word with its governing preposition. This phenomenon is the subject of ongoing research and for the purposes of this paper, it is assumed that a second midfield *er* is ungrammatical.

- (22) a. *Er<sub>L</sub> and er<sub>P</sub>*  
*De student wacht er (\*er) nu (\*er=)op*  
 The student waits ER (\*ER) now (\*ER=)on  
 “The student is waiting there for it now.”
- b. *Er<sub>L</sub> and er<sub>Q</sub>*  
*Merel heeft er (\*er) vijf gegeten*  
 Merel has ER (\*ER) five eaten  
 “Merel ate five there.”
- c. *Er<sub>P</sub> and er<sub>Q</sub>*  
*Suus heeft er (\*er) drie (\*er=)op neergezet*  
 Suus has ER (\*ER) three (\*ER=)on put.down  
 “Suus put three down on it.”

It is also possible for a single instance *er* to provide multiple instances of the same function with different antecedents (23).

- (23) a. *Jan heeft de sleutel met een tang<sub>i</sub> uit het slot<sub>j</sub> gehaald.*  
 Jan has the key with a tongs<sub>i</sub> out the lock<sub>j</sub> taken  
 “Jan took the key out of the lock with pliers”

<sup>14</sup>Odijk (1993)

<sup>15</sup><https://www.standaard.be/cnt/g0lsk35f>



- b. *Jan heeft er<sub>ij:PP</sub> de sleutel mee<sub>i</sub> uit<sub>j</sub> gehaald.*  
 Jan has ER<sub>ij</sub> the key with<sub>i</sub> out taken.  
 “Jan took the key out of it with them.”<sup>16</sup>

Sentences with a single *er* providing four functions with three distinct antecedents are also possible (24).

- (24) a. ... *dat er twee studenten<sub>i</sub> drie boeken<sub>j</sub> uit de boekenkast<sub>k</sub> gehaald hebben.*  
 ... that ER two students three books out the bookcase fetched have.  
 “... that two students fetched three books out of the bookcase.”
- b. ... *dat er<sub>ijk:XQQP</sub> twee<sub>i</sub> drie<sub>j</sub> uit<sub>k</sub> gehaald hebben.*  
 ... COMP ER<sub>ijk</sub> two<sub>i</sub> three<sub>j</sub> out<sub>k</sub> fetched have.  
 “... that two (of them) fetched three (of them) out of it.”<sup>17</sup>

### 2.2.3 Summary: multifunctional *er*

All functions of *er* are compatible with each other in the midfield. While *er<sub>L</sub>* and *er<sub>P</sub>* cannot occupy the prefield if they provide the sole function of *er*, they must be provided by prefield *er* if this is licensed by *er<sub>X</sub>*. However, any instances of *er<sub>Q</sub>* can never be provided by prefield *er*, instead requiring an instance of *er* in the midfield.

## 3 Accounting for multifunctionality

Most other accounts rely on syntactic deletion rules (e.g. Bennis 1986, Neeleman and van de Koot 2006). Webelhuth and Bonami (2019) propose an account within HPSG which relies on the optional non-expression of *er* in phrase-structure, the expression being determined by interactions of constraints that relate specifically to the prefield and the midfield. Again, information is contributed to the analysis by an element that is invisible in the string. Phonological deletion is another possible cause, but as Dutch allows the repetition of other unstressed pronouns (25) this explanation is also unsatisfactory, and is also rejected by Neeleman and van de Koot (2006).

- (25) a. *Opdat je je bruiloft keer op keer opnieuw kunt beleven.*  
 so.that 2 2.POSS wedding time on time again can experience  
 “So that you can relive your wedding time and time again.”<sup>18</sup>

<sup>16</sup>Webelhuth and Bonami (2019, exx. 6a,6d)

<sup>17</sup>Webelhuth and Bonami (2019, exx. 8a,c)

<sup>18</sup>weddingreport.nl

- b. *Herinner je je je verjaardag?*  
 remember 2 2.REFL 2.POSS birthday  
 “Do you remember your birthday?”<sup>19</sup>

However, accounts based on deletion are unsatisfactory: the required deletion of *c*-structure elements means that an analysis is no longer monotonic. This causes problems computationally and, for LFG, contravenes one of the underpinning assumptions of the theory. It is also unclear how empirical psycholinguistic evidence in support of a deletion-based account might be gathered.

The proposal here is based on interactions between positional and functional constraints, builds on Asudeh (2009) in relating *f*-structure to the string. Rather than remove elements from *c*-structure by deletion, the account assumes that the lexical specification for *er* includes optional resources that can be included as required to satisfy constraints introduced elsewhere in the string. Similar to Webelhuth and Bonami (2019), the *c*-structure constraints distinguish between the prefield and the midfield. The finer-grained constraints on the position of *er* within the midfield are left for future work.

### 3.1 Constraints and interactions

A lexical specification for *er* must reflect constraints at both *f*-structure and *c*-structure. At *f*-structure, a single instance of *er* must correspond to a single *f*-structure via the correspondence function  $\phi$ , whilst potentially providing content, including distinct PRED values, to multiple *f*-structures. At *c*-structure, the functions expressed by a single instance of *er* constrain its distribution.

#### 3.1.1 C-structure distributional constraints

A sole *er* in the midfield is grammatical whatever the combination of functions it carries. This provides evidence that *er* is a single lexical item that can provide more than one PRED value into *f*-structure. It also demonstrates that the four functions  $er_X$ ,  $er_L$ ,  $er_P$ , and  $er_Q$  are not intrinsically incompatible, and that the constraints on particular combinations of function associated with specific *c*-structure positions arise from interactions between constraints within the lexical specification and constraints within phrase structure rules.

A sole *er* in the prefield is only grammatical when  $er_X$  is present, and is never grammatical where the clause has an instance of  $er_Q$ . This requires the specification for  $er_X$  to satisfy *c*-structural constraints on the Spec-IP position, and the specification for  $er_Q$  to be incompatible with those constraints. It further suggests that  $er_L$  and  $er_P$  are underspecified with regard to the Spec-IP constraints, allowing them to occupy Spec-IP where  $er_X$  is present, but preventing them appearing in Spec-IP without  $er_X$ .

---

<sup>19</sup>[taalthuis.com/theory/pronouns](http://taalthuis.com/theory/pronouns)

Two *er* in the string are ungrammatical if the clause has  $er_X$  and either  $er_L$  or  $er_P$ . This suggests that there is a string ordering constraint on  $er_L$  and  $er_P$  such that they must be carried by the leftmost instance of *er*.

### 3.1.2 Functional assumptions

Existential  $er_X$  does not contribute a semantic form to f-structure. If it is present together with an indefinite subject, that provides the PRED value. If it is present in the impersonal passive construction, the subject is athematic and therefore a value of SUBJ PRED would result in an incoherent f-structure.

However, each instance of the functions  $er_L$ ,  $er_P$ , and  $er_Q$  contributes the constraint PRED = ‘pro’ to an f-structure within the clause, and these pronouns may have different antecedents. Except in cases such as (21b), these multiple PRED values are provided by one instance of *er* in the string and therefore must correspond to a single f-structure through the  $\phi$ -function. This is problematic because of the PRED uniqueness constraint on f-structures.

To resolve this problem, I propose to amend the definition of the  $\phi$ -function such that it is possible to project a single set of f-structures. For *er* this set is defined as  $E$ . Each f-structure within the set  $E$  then shares its structure with a grammatical function in the clause. These individual f-structures have specific constraints, not only functionally but also relating to c-structure, linear precedence in the string, and information structure.

As a result, the lexical specification for *er* must be described in general terms, with a specific instantiation for each appearance of *er* in a string. These instantiations must include the constraints relating to at least one function of *er*, but the exact composition is dependent on the content of the whole clause.

The question then arises as to where in f-structure the set  $E$  sits. The structure shared by the individual f-structures within  $E$  relating to  $er_P$ ,  $er_Q$ , and non-passive  $er_X$ , must also contain material contributed by other c-structure elements, which may be non-adjacent to *er* in the string. This can be seen as a dislocation within the clause, but because an instance of *er* is not necessarily associated with a prominent element of information structure such as topic or focus, it is not appropriate to use the f-structural discourse functions TOPIC or FOCUS (Bresnan and Mchombo 1987, and others). Instead, I follow Dalrymple et al. (2019, p. 38), who propose the overlay function DIS to represent dislocation or long distance dependency, and who include the discourse functions TOPIC and FOCUS in the separate i-structure level of representation. Accordingly, I propose that the set  $E$  is the value of the overlay function DIS.

## 3.2 Building a lexical specification

The lexical specification for *er* consists of a core specification together with four subspecifications that each relate to one of the functions of *er*. The specification results in a set of f-structures, each of which shares structure with another f-structure

or grammatical function in the clause. The subspecifications each follow a similar template, including functional constraints, any constraints on the number of f-structures of a given ERTYPE that may be present, the path constraint for structure sharing, and a c-structure precedence condition that constrains the number of instances of *er* in the s-string. The feature ERTYPE is used for  $er_X$  and  $er_L$  to preclude situations where an infinitely large set  $E$  could be generated.  $er_P$  and  $er_Q$  are not specified for the feature because their presence is constrained by completeness and coherence constraints dependent on other words in the sentence. The subspecifications relating to  $er_X$  and  $er_Q$  also reflect the relationship of these functions to the c-structure Spec-IP position.

For a particular instance of *er*, the core specification is always present, and copies of the subspecifications are added to satisfy the requirements of the sentence. Thus the exact composition of the set  $E$  depends on the presence of other elements in the clause (e.g. an indefinite subject for  $er_X$ , an objectless preposition for  $er_P$ , a number without a specific noun for  $er_Q$ , a location required by valency or context for  $er_L$ ) to satisfy constraints.<sup>20</sup> If it is not possible to build a lexical specification for a particular instance of *er*, or if the generated specification results in feature clashes, the sentence is ungrammatical.

### 3.2.1 The core specification

The core specification for *er* is given at (26).

$$(26) \quad er \quad N \quad (DIS \uparrow) \\
\{E: \%ER_1, \dots, \%ER_n\}, |E| \geq 1 \\
\%ER_i = \{ER_X \mid ER_L \mid ER_P \mid ER_Q\} \\
\%ER_i = ((DIS \uparrow) ERPATH_i)$$

The first line constrains the information from *er* to be added to the value of the clause's overlay function. The second and third lines define this information as a non-zero set of f-structures, each represented by indexed local variable  $\%ER_i$ . Each instance of  $\%ER_i$  is further constrained to be one of four subspecifications  $ER_X$ ,  $ER_L$ ,  $ER_P$ ,  $ER_Q$  which correspond to the four functions of *er*. There may be more than one instance of  $er_P$  or  $er_Q$  in a clause, and so it is assumed that there is no upper limit on the size of set  $E$ . The fourth line specifies that each instance of  $\%ER_i$  shares its structure with an f-structure along the path  $ERPATH_i$ , which is also further defined in the subspecifications.

### 3.2.2 Subspecification $ER_X$

The subspecification  $ER_X$  is given at (27).

<sup>20</sup>The specification cannot determine whether or not a locative adjunct is contributed by *er* in a given context: the factors that govern native speakers' intuitions about whether a location is contributed by *er* in a given context are left for future research.

$$\begin{aligned}
(27) \quad ER_X &\equiv (\%ER_i \text{ DEF}) \neq + \\
&(\%ER_i \text{ ERTYPE}) = X \\
&\neg(\%ER_j \in E). \%ER_j \neq \%ER_i \wedge (\%ER_j \text{ ERTYPE})=X \\
&ERPATH_i = \text{SUBJ} \\
&\neg *_{n} . *_{n} < \hat{*} \wedge \pi^{-1}(*_{n}) = er
\end{aligned}$$

The first line prevents an instance of  $ER_X$  from contributing to an f-structure from a definite DP or NP. The second line sets the value of the instance's ERTYPE feature to be X, and the third line uses the ERTYPE feature to ensure that there is only one f-structure specified by  $ER_X$  in set  $E$ . The fourth line constrains the f-structure to share structure with the SUBJ of the clause. This licenses  $er$  to occupy Spec-IP. The fifth line is a c-structure precedence constraint relating the terminal c-structure node for this instance of  $er$  ( $\hat{*}$ ) to other nodes in c-structure. It says that there is no other node  $*_{n}$  that precedes this instance of  $er$ , for which the associated word in the string,  $\pi^{-1}(*_{n})$ , is  $er$ . The effect of this is that any f-structure specified by  $ER_X$  is constrained to be contributed by the leftmost instance of  $er$  in the string.

### 3.2.3 Subspecification $ER_L$

The subspecification  $ER_L$  is given at (28).

$$\begin{aligned}
(28) \quad ER_L &\equiv (\%ER_i \text{ PRED}) = \text{'pro'}$$

$$\begin{aligned}
&(\%ER_i \text{ ERTYPE}) = L \\
&\neg(f \in ((\text{DIS } \uparrow) \text{ ADJ})). f \neq \%ER_i \wedge (f \text{ ERTYPE})=L \\
&ERPATH_i = \{\text{OBL}_{Loc} \mid \text{ADJ} \in\} \\
&\neg *_{n} . *_{n} < \hat{*} \wedge \pi^{-1}(*_{n}) = er
\end{aligned}$$

The first line contributes the value PRED = 'pro' to an f-structure which is an instance of  $ER_L$ . The second line sets the value of that f-structure's ERTYPE feature to be L. The third line uses the ERTYPE feature to ensure that there is only one f-structure specified by  $ER_L$  within the adjunct set of the clause. The fourth line constrains the f-structure to share structure with either the clause's  $\text{OBL}_{Loc}$  grammatical function or a member of the clause's adjunct set. And the fifth line again constrains any f-structure specified by  $ER_L$  to be contributed by the leftmost instance of  $er$  in the string.

### 3.2.4 Subspecification $ER_P$

The subspecification  $ER_P$  is given at (29).

$$\begin{aligned}
(29) \quad ER_P &\equiv (\%ER_i \text{ PRED}) = \text{'pro'}$$

$$\begin{aligned}
&ERPATH_i = \{\text{OBL}_{\theta} \mid \text{ADJ} \in\} \text{ OBJ} \\
&\neg *_{n} . *_{n} < \hat{*} \wedge \pi^{-1}(*_{n}) = er
\end{aligned}$$

The first line again contributes the value PRED = 'pro' to an f-structure that is an instance of  $ER_P$ . The second line constrains the f-structure to share structure with

the object of either an oblique grammatical function or a member of the clause’s adjunct set. The presence of an  $ER_P$  f-structure requires there to be an available OBJ, and so it is not necessary to further constrain the number of f-structures with  $ERTYPE = P$ . The third line again constrains any f-structure specified by  $ER_P$  to be contributed by the leftmost instance of  $er$  in the string.

### 3.2.5 Subspecification $ER_Q$

The subspecification  $ER_Q$  is given at (30).

$$\begin{aligned}
 (30) \quad ER_Q &\equiv (\%ER_i \text{ PRED}) = \text{‘pro’} \\
 &(\%ER_i \text{ DEF}) = - \\
 &(\%ER_i \{COMP|NUM\}) \\
 &ERPATH_i = \{SUBJ \mid OBJ \mid OBJ_\theta\} \\
 &\neg *_{n_i} .\hat{*} < *_{n_i} \wedge \pi^{-1}(*_{n_i}) = er \\
 &(\uparrow_{\sigma_i} \text{ PROM}) = -
 \end{aligned}$$

Similarly to the subspecifications  $ER_L$  and  $ER_P$ , the first line contributes the value  $PRED = \text{‘pro’}$  to an f-structure that is an instance of  $ER_P$ . The second line constrains the  $DEF$  feature of that f-structure to be negative. The third line requires the f-structure to have either a  $NUM$  or a  $COMP$  attribute, in line with the requirement discussed in Section 2.1.4 that the nominal antecedent of  $er_Q$  is restricted in some way. The fourth line constrains the f-structure to share structure with one of the term grammatical functions of the clause. In the fifth line, the precedence constraint is reversed so that there is no other terminal node projected by an instance of  $er$  that is preceded by this instance of  $er$ . Thus any f-structure specified by  $ER_Q$  to be contributed by the rightmost instance of  $er$  in the string. The sixth line specifies the information structure feature  $PROM$  to be negative.<sup>21</sup> This clashes with constraints on Spec-IP, discussed below, and thus prevents an f-structure specified by  $ER_Q$  from being contributed by  $er$  in Spec-IP.

## 4 Analysis

The analysis follows the phrase-structure assumptions in Section 1.1. For phrases where  $er$  may appear, the constraint  $(\uparrow \text{SUBJ}) = \downarrow$  on dependents of the verb is replaced by the disjunction  $\{(\uparrow \text{SUBJ}) = \downarrow \mid (\uparrow \text{DIS}) = \downarrow\}$ .<sup>22</sup> Further constraints apply to Spec-IP, shown in (31).<sup>23</sup>

<sup>21</sup>I follow the treatment of information structure in Chapter 10 of Dalrymple et al. (2019). In summary,  $\uparrow_{\sigma_i}$  and  $\downarrow_{\sigma_i}$  represent the i-structures projected by the f-structures  $\uparrow$  and  $\downarrow$  respectively. Within i-structure,  $PROM$  is a feature representing the notion of prominence.

<sup>22</sup>The detail of constraints on  $er$  within the midfield is left for future work. This paper makes the simplifying assumption that  $er$  occurs either in the prefield or the start of the midfield.

<sup>23</sup>DF is an i-structure feature representing discourse functions, allowing sentential content to be associated with TOPIC or FOCUS.

$$(31) \quad \left\{ \begin{array}{l} \text{XP} \\ \{(\uparrow \text{GF}) = \downarrow \mid (\uparrow \text{DIS}) = \downarrow\} \\ \uparrow_{\sigma\iota} = \downarrow_{\sigma\iota} \\ (\downarrow_{\sigma\iota} \text{PROM}) \neq - \\ (\downarrow_{\sigma\iota} \text{DF}) = \text{TOPIC} \mid \\ (\downarrow_{\sigma\iota} \text{DF}) = \text{FOCUS} \mid \\ (\uparrow \text{SUBJ}) = \downarrow \mid \\ \%ER \in \downarrow \wedge \%ER = (\uparrow \text{SUBJ}) \end{array} \right\}$$

The constraint  $(\downarrow_{\sigma\iota} \text{PROM}) \neq -$  means that the constituent occupying Spec-IP must not be intrinsically non-prominent (a characteristic assumed for  $er_Q$ ). The disjunction means that the constituent must provide either topic or focus of the sentence (represented by the value TOPIC or FOCUS for the clause's DF feature at i-structure), or the subject. The final line of the constraint covers the case where  $er$  occupies Spec-IP. In this case, there must be an f-structure in the set which is equal to  $(\uparrow \text{SUBJ})$ . This constraint is the set equivalent of the previous element of the disjunction  $(\uparrow \text{SUBJ}) = \downarrow$ ; it can be satisfied by the presence in the set of  $ER_X$ , by equation (27), or of  $ER_Q$ , by equation (30). However,  $ER_Q$  is incompatible with Spec-IP because of the prominence constraint mentioned above.

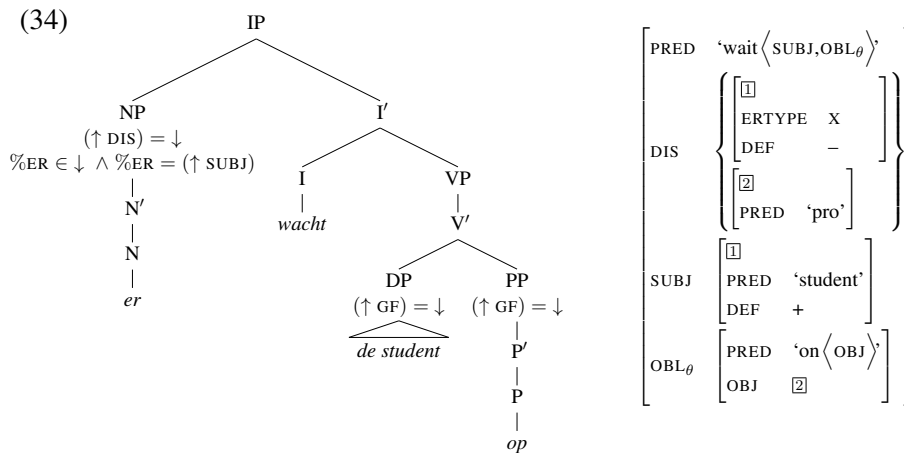
In the c-structure diagrams that follow the constraint  $\uparrow = \downarrow$  is omitted for clarity, dependents of the verb show only the applicable element of the disjunction  $\{(\uparrow \text{GF}) = \downarrow \mid (\uparrow \text{DIS}) = \downarrow\}$ , and only the relevant constraints on Spec-IP from (31) are shown.

#### 4.1 $Er_P$ in the prefield with and without $er_X$

Example (32) is ungrammatical. Only  $er_X$  licenses  $er$  in the prefield, through the c-structure constraint  $(\uparrow \text{SUBJ}) \in \downarrow$ . The set  $E$  contains two f-structures, one specified by  $ER_X$  and one by  $ER_P$ . The resulting lexical specification for  $er$  is given in (33). Because  $er_X$  carries the constraint  $(ER_1 \text{DEF}) \neq +$ , there is a feature clash with the definite subject *de student* 'the student' (34).

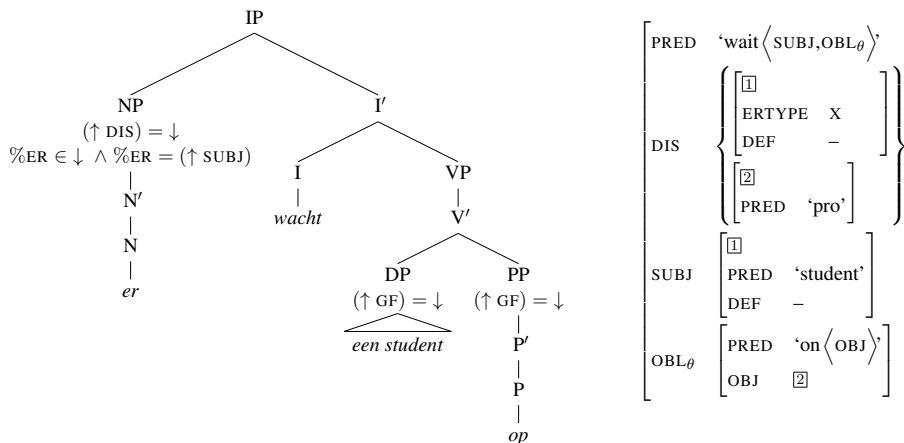
(32) \* *Er wacht de student op*  
 ER waits the student on  
 (intended) "The student is waiting for it."

(33)  $er$  N (DIS  $\uparrow$ )  
 $\neg *_{n} . *_{n} < \hat{*} \wedge \pi^{-1}(*_{n}) = er$   
 $\{E: ER_1, ER_2\}$   
 $(ER_1 \text{DEF}) \neq +$   
 $(ER_1 \text{ERTYPE}) = X$   
 $\neg (\%ER_j \in E) . \%ER_j \neq \%ER_1 \wedge (\%ER_j \text{ERTYPE}) = X$   
 $ER_1 = ((\text{DIS } \uparrow) \text{SUBJ})$   
 $(ER_2 \text{PRED}) = \text{'pro'}$   
 $ER_2 = ((\text{DIS } \uparrow) \{OBL_{\theta} \mid \text{ADJ} \in\} \text{OBJ})$



Replacing the definite subject with the corresponding indefinite *een student* ‘a student’ removes the feature clash and the sentence becomes grammatical (35).

- (35) *Er wacht een student op*  
 ER waits a student on  
 “A student is waiting for it.”



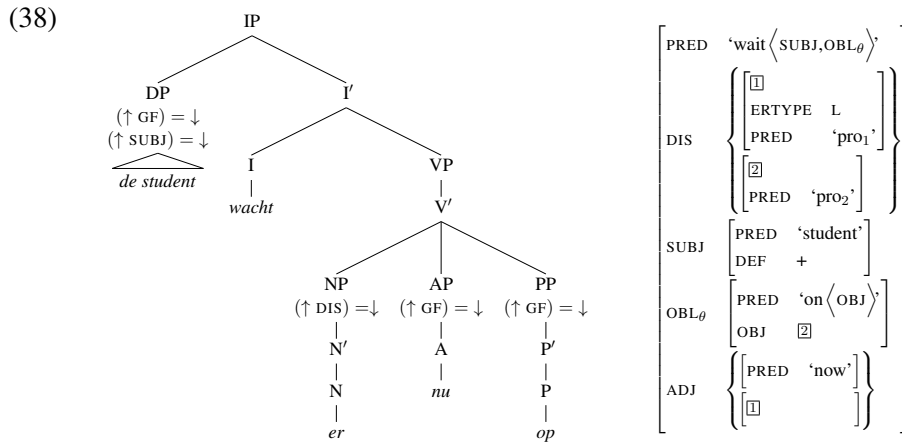
## 4.2 $Er_L$ and $er_P$ in the midfield: one $er$ or two?

Example (36), repeated from (22a), is grammatical. Spec-IP is occupied by the subject *de student* ‘the student’, with *er* at the start of the midfield. The sentence contains  $er_L$  and  $er_P$ , and as a result the set  $E$  has two f-structures, one specified by  $ER_L$  and one by  $ER_P$ . The lexical specification generated for *er* in this case is given at (37), and the c- and f-structure pair is shown at (38).

- (36) *De student wacht er<sub>LP</sub> nu op*  
 The student waits ER now on  
 “The student is waiting there for it now.”



$$\begin{aligned}
(37) \quad & er \quad N \quad (DIS \uparrow) \\
& \neg *_{n} . *_{n} < \hat{*} \wedge \pi^{-1}(*_{n}) = er \\
& \{E: ER_1, ER_2\} \\
& (ER_1 \text{ PRED}) = \text{'pro'} \\
& (ER_1 \text{ ERTYPE}) = L \\
& \neg(f \in ((DIS \uparrow) \text{ ADJ})).f \neq \%ER_1 \wedge (f \text{ ERTYPE})=L \\
& ER_1 = ((DIS \uparrow) \{OBL_{Loc} \mid \text{ADJ} \in\}) \\
& (ER_2 \text{ PRED}) = \text{'pro'} \\
& ER_2 = ((DIS \uparrow) \{OBL_{\theta} \mid \text{ADJ} \in\} \text{ OBJ})
\end{aligned}$$



Attempting to add a second instance of *er* in the sentence, so that each of  $er_L$  and  $er_P$  has a separate word contributing a PRED value, results in ungrammaticality. The clause again requires two functions of *er* to be present,  $er_L$  and  $er_P$ , but this time two lexical specifications for *er* are generated, one for each instance. The lexical specification from (37) is still valid, because the subspecifications  $ER_L$  and  $ER_P$  must both be part of the specification for the leftmost instance of *er*. The attempt to generate a specification for the second instance of *er* fails (39): there are no other functions of *er* required by the clause and so the constraint that  $er_2$  projects a non-empty set cannot be satisfied.

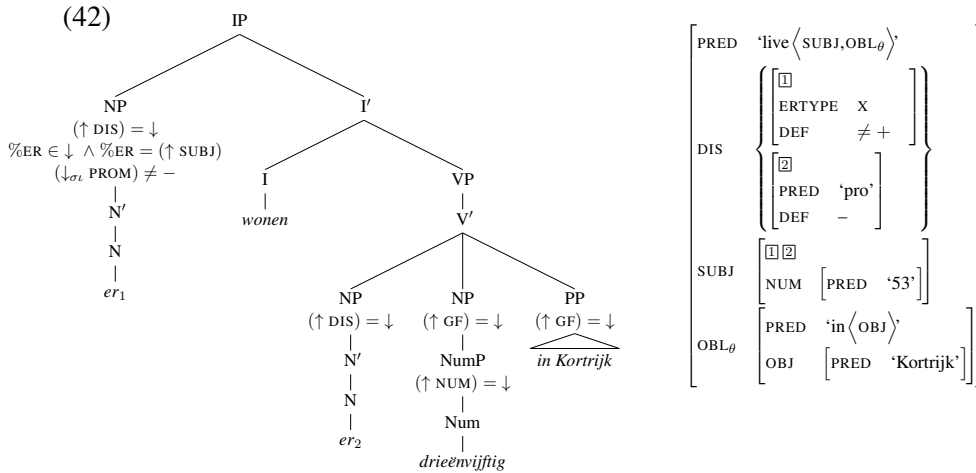
$$\begin{aligned}
(39) \quad & er_2 \quad N \quad (DIS \uparrow) \\
& \{E: \%ER_1, \dots, \%ER_n\}, |E| \geq 1 \\
& \%ER_i = \{ER_X \mid ER_L \mid ER_P \mid ER_Q\} \\
& \%ER_i = ((DIS \uparrow) \text{ ERPATH}_i)
\end{aligned}$$

### 4.3 Sentences with $er_Q$

Example (40), repeated from (21b), shows the case where two instances of *er* in a clause are grammatical. Each instance of *er* generates a lexical specification. The specification for the first instance (41a) holds the constraints for  $er_X$  and that for the second instance (41b) holds the constraints for  $er_Q$ .

- (40)  $Er_X$  *wonen*  $er_Q$  *drieënvijftig* *in Kortrijk*  
 ER live ER fifty-three in Kortrijk  
 “There are fifty-three living in Kortrijk.”

- (41) a.  $er_1$  N (DIS  $\uparrow$ )  
 $\neg *n . *n < \hat{*} \wedge \pi^{-1}(*n) = er$   
 $\{E: ER_1\}$   
 $(ER_1 \text{ DEF}) \neq +$   
 $(ER_1 \text{ ERTYPE}) = X$   
 $\neg (\%ER_j \in E) . \%ER_j \neq \%ER_1 \wedge (\%ER_j \text{ ERTYPE}) = X$   
 $ER_1 = ((\text{DIS } \uparrow) \text{ SUBJ})$
- b.  $er_2$  N (DIS  $\uparrow$ )  
 $\neg *n . \hat{*} < *n \wedge \pi^{-1}(*n) = er$   
 $(\uparrow_{\sigma_l} \text{ PROM}) = -$   
 $\{E: ER_2\}$   
 $(ER_2 \text{ PRED}) = \text{'pro'}$   
 $(ER_2 \text{ DEF}) = -$   
 $(ER_2 \{ \text{COMP} | \text{NUM} \})$   
 $ER_2 = ((\text{DIS } \uparrow) \{ \text{SUBJ} | \text{OBJ} | \text{OBJ}_\theta \})$

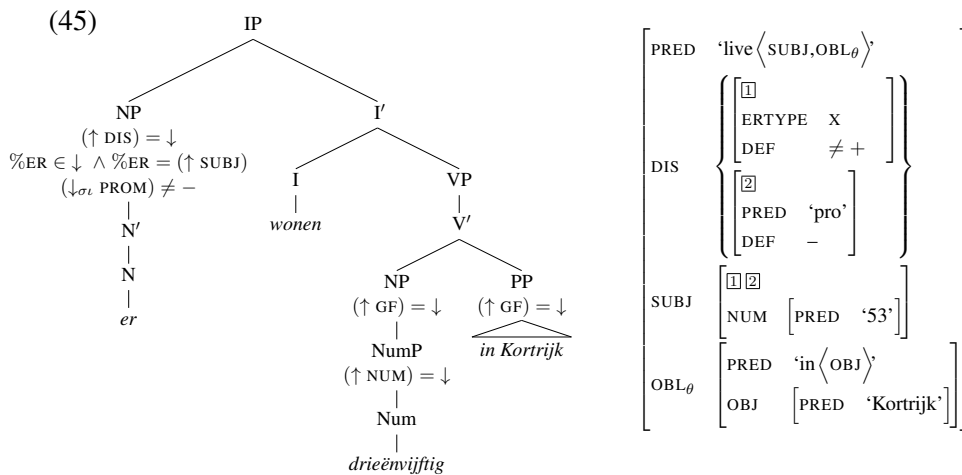


If the second instance of *er* is deleted, the sentence becomes ungrammatical (43).

- (43) \* $Er_{XQ}$  *wonen* *drieënvijftig* *in Kortrijk*  
 ER live fifty-three in Kortrijk  
 (intended) “There are fifty-three living in Kortrijk.”

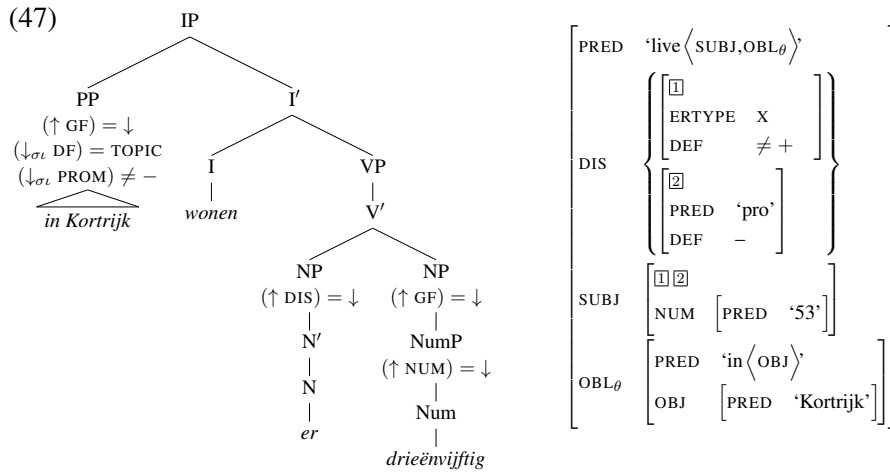
The lexical specification for the single instance of *er* must now hold the constraints for both  $er_X$  and  $er_Q$  (44). Although the f-structure is well-formed, ungrammaticality arises at information structure because the constraint  $(\uparrow_{\sigma_l} \text{ PROM}) = -$  contributed by  $ER_Q$  is incompatible with the constraint  $(\downarrow_{\sigma_l} \text{ PROM}) \neq -$  associated with Spec-IP (45).

- (44) *er* N (DIS ↑)  
 $(\uparrow_{\sigma_i} \text{PROM}) = -$   
 $\neg *_{*n} . *_{*n} < \hat{*} \wedge \pi^{-1}(*_{*n}) = er$   
 $\neg *_{*n} . \hat{*} < *_{*n} \wedge \pi^{-1}(*_{*n}) = er$
- $\{E: ER_1, ER_2\}$   
 $(ER_1 \text{ DEF}) \neq +$   
 $(ER_1 \text{ ERTYPE}) = X$   
 $\neg(\%ER_j \in E) . \%ER_j \neq \%ER_1 \wedge (\%ER_j \text{ ERTYPE}) = X$   
 $ER_1 = ((\text{DIS } \uparrow) \text{SUBJ})$
- $(ER_2 \text{ PRED}) = \text{'pro'}$   
 $(ER_2 \text{ DEF}) = -$   
 $(ER_2 \{ \text{COMP} | \text{NUM} \})$   
 $ER_2 = ((\text{DIS } \uparrow) \{ \text{SUBJ} | \text{OBJ} | \text{OBJ}_{\theta} \})$



However, if the first instance of *er* is deleted and the PP *in Kortrijk* 'in Kortrijk' occupies Spec-IP as the topic, the resulting sentence is grammatical (46). The single instance of *er* no longer occupies a position that has an information structure constraint (47).

- (46) *In Kortrijk wonen er<sub>XQ</sub> drieënvijftig*  
 In Kortrijk live ER fifty-three  
 "There are 53 living in Kortrijk"



## 5 Conclusion

The above account demonstrates how the LFG architecture can account for the complex distribution of *er*, including its ability to refer to multiple distinct antecedents. Rather than assume unexpressed or deleted elements of c-structure, the account assumes that optional resources can be added to meet the constraints introduced by other elements of the string.

The role of sets in f-structure is long established. The innovation in this paper is the ability for a set to be generated by a single lexical item. The choice of a set rather than a disjunction is motivated by the assumption that there is no upper syntactic constraint on the number of antecedents to *er*, but that pragmatic or processing constraints may introduce an effective upper limit to acceptability: compare the syntactically correct English sentence *It<sub>i</sub> put it<sub>j</sub> next to it<sub>k</sub> on it<sub>l</sub> using it<sub>l</sub>*. Work to investigate this assumption is ongoing.

## References

- Asudeh, Ash. 2009. Adjacency and Locality: A Constraint-Based Analysis of Complementizer-Adjacent Extraction. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG09 Conference*, pages 106–126, Stanford, CA.: CSLI Publications.
- Bech, Gunnar. 1952. Über das niederländische Adverbialpronomen *er*. *Travaux du Cercle Linguistique de Copenhague* 8, 5–32.
- Bennis, Hans. 1986. *Gaps and dummies*. Dordrecht: Foris.
- Booth, Hannah and Schätzle, Christin. 2019. The Syntactic Encoding of Information Structure in the History of Icelandic. In Miriam Butt, Tracy Holloway King and Ida Toivonen (eds.), *Proceedings of the LFG'19 Conference*, pages 69–89, Stanford, CA.: CSLI Publications.

- Bresnan, Joan and Mchombo, Sam A. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language* 63(4), 741–782.
- Dalrymple, Mary, Lowe, John J. and Mycock, Louise. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.
- Donaldson, Bruce. 2008. *Dutch: A Comprehensive Grammar*. Abingdon: Routledge, second edition.
- Grondelaers, Stefan, Speelman, Dirk, Drieghe, Denis, Brysbaert, Marc and Geeraerts, Dirk. 2009. Introducing a new entity into discourse: Comprehension and production evidence for the status of Dutch *er* “there” as a higher-level expectancy monitor. *Acta Psychologica* 130, 153–160.
- Haeseryn, Walter Jan-Maria, Romijn, Kirsten, Geerts, Guido, Rooij, J de and Van den Toorn, Maarten C. 1997. *Algemene Nederlandse spraakkunst [2 volumes]*. Groningen/Deurne: Nijhoff/Wolters Plantyn.
- Klooster, Wim G. 2014. Re *er*. In Freek Van de Velde, Hans Smessaert, Frank Van Eynde and Sarah Verbrugge (eds.), *Patroon en Argument: een dubbelfeestbundel bij het emeritaat van William van Belle en Joop van der Horst*, pages 255–270, Universitaire Pers Leuven.
- Neeleman, Ad and van de Koot, Hans. 2006. Syntactic haplology. In Martin Everaert and Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, volume 1, Chapter 60, pages 687–710, Oxford: Blackwell.
- Odijk, Johannes Engelbertus Josephus Maria. 1993. *Compositionality and syntactic generalizations*. Ph. D.thesis, Tilburg University.
- Rosén, Victoria, De Smedt, Koenraad, Meurer, Paul and Dyvik, Helge. 2012. An open infrastructure for advanced treebanking. In Jan Hajič, Koenraad De Smedt, Marko Tadić and António Branco (eds.), *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*, pages 22–29, Istanbul, Turkey.
- Sells, Peter. 2001. *Structure, Alignment and Optimality in Swedish*. Stanford, CA.: CSLI Publications.
- Sells, Peter. 2005. The peripherality of the Icelandic expletive. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG05 Conference*, pages 408–428, Stanford, CA.: CSLI Publications.
- van Riemsdijk, Henk. 1978. *A case study in syntactic markedness*. Dordrecht: Foris.
- Webelhuth, Gert and Bonami, Olivier. 2019. Syntactic haplology and the Dutch proform “er”. In Stefan Müller and Petya Osenova (eds.), *Proceedings of the 26th International Conference on Head-Driven Phrase Structure Grammar*, pages 100–119, Stanford, CA.: CSLI Publications.