

Lea M. Fricke, Swantje Tönnis

***Es ist dies* – A Special Use of German Prefield-*es*¹**

Abstract We present a corpus study on a hitherto unstudied use of the German prefield-*es* in combination with a demonstrative subject *dies* and a copula verb *ist*, which we call *Es ist dies*-sentences. In such constructions, the prefield-*es* appears redundant as they contain a suitable and mostly preferred candidate to fill the prefield, the demonstrative pronoun *dies*. According to our corpus data, this construction is predominantly used in southern varieties of German (Swiss, Austrian and Bavarian German). In order to better understand the distribution of these constructions, we compared *Es ist dies*-sentences to a sample of unmarked *Dies ist*-sentences that mirrored the distribution of the prefield-*es* cases. We found two significant differences between the two samples with regard to a) the distance to the antecedent of *dies* and b) the content of the sentence. Based on our findings, we propose a modification of Speyer's (2008, 2009) stochastic Optimality Theoretic (OT) model of prefield ranking.

Keywords Prefield-*es*, information structure, stochastic OT, southern varieties of German

1 Introduction

The use of prefield-*es* in sentences like (1) seems redundant and therefore marked, as the sentence ostensibly offers a better candidate to fill the prefield position: the demonstrative pronoun *dies* ('this'). Thus, the version presented in (2) appears more natural.

- 1 Many thanks go to Edgar Onea for his valuable suggestions as well as his comments on earlier versions of this paper, and to Alexander Schreiber for advising us on the statistical analysis. We also want to thank the organizers of *Grammar and Corpora 2016* for providing the opportunity for such an inspiring exchange. We are grateful for the helpful comments we received from the audience at the poster session, we thank in particular Erik Fuß, Carlo Geraci, Marek Konopka, and Helmut Weiß. Finally, we thank two anonymous reviewers for their comments.

- (1) **Es ist dies** der schwerste Fall von Marktmanipulation, den wir
 It is this the most severe case of market manipulation that we
 je gesehen haben.²
 ever seen have
 ‘This is the worst case of market manipulation that we have ever seen.’
- (2) **Dies ist** der schwerste Fall von Marktmanipulation, den wir
 this is the most severe case of market manipulation that we
 je gesehen haben.
 ever seen have
 ‘This is the most severe case of market manipulation that we have ever
 seen.’

This construction is not an idiom, since it allows a significant range of variation. There is also a variant featuring *das*, ‘that’, instead of *dies*. Moreover, the construction may surface with different inflected forms, with and without a relative clause, and the NP can be preceded by further elements. In this paper, we focus on the variant presented in (1). After explaining in more detail why this construction violates expectations about the use of prefield-*es*, we present our corpus study, which is an investigation into the conditions of its use. Based on the results we present our tentative analysis, a modification of Speyer’s (2008, 2009) prefield ranking.

2 Background

In a standard German declarative matrix clause, the finite verb occurs in the second position. This means that the prefield, the position in front of the finite verb, needs to be filled by one constituent. In some cases, exemplified by sentence (3a.), this is brought about by the non-phoric use of the third person neuter pronoun *es*, which does not contribute to the truth-conditions of the sentence. Unlike the also non-phoric subject-*es* which functions as a formal subject for verbs that do not assign thematic roles (e.g. *es regnet*, ‘it is raining’), this so-called prefield-*es* is not an argument (see Pütz 1986, Tomaselli 1986, Cardinaletti 1990, Zifonun 1995, Paranhos Zitterbart 2002, and Pittner & Bermann 2004 for the different uses of *es*). It only serves to fill the prefield in order to have a verb second clause. The prefield is the only position this type of *es* can occur in, as shown by the ungrammatical example (3b.).

2 <http://www.tagesanzeiger.ch/wirtschaft/unternehmen-und-konjunktur/Es-ist-der-schwerste-Fall-den-wir-je-gesehen-haben/story/31098247>.

- (3) a. Es kommen viele internationale Gäste.
 it come many international guests
 ‘Many international guests are coming.’
- b. *Hoffentlich kommen es viele internationale Gäste.
 hopefully come it many international guests
 ‘Hopefully, many international guests are coming.’

The reason for using non-phoric *es* in the prefield instead of the subject has been argued to lie in the information structure. It has been suggested that *es* may be located in the prefield as a placeholder when the subject carries the informational load (Zifonun et al. 1997) and represents new information (Pittner & Bermann 2004) and therefore tends to be located towards the end of the sentence. Speyer (2009) further investigates the conditions that allow the occurrence of *es* in the prefield. He characterizes the use of prefield-*es* as a “last resort” to fill the prefield in order to have a V2-sentence if there is no better candidate available. A better candidate according to his stochastic OT based prefield ranking would be the topic of the sentence which Speyer (2009: 339) defines in terms of Centering Theory (Grosz et al. 1995; Walker et al. 1998) as a ‘macrostructurally relevant’ entity. This means a topic either needs to be discourse-old (i.e. it occurs in the directly preceding sentence, Speyer 2009: 336), or relevant in the further course of the text in order to be allowed in the prefield. In addition to lacking a more appropriate prefield filler, sentences featuring prefield-*es* were observed to contain few constituents, often only the subject (Speyer 2009: 334).

Clearly, in the last respect our *Es ist dies*-sentences differ from the classic cases of prefield-*es*, as they are copular sentences, which always contain at least two arguments. Furthermore, the fact that the phoric *dies*³ refers back to an antecedent in the text indicates that it does not represent new information and seems to point to its macrostructural relevance. However, in a first explorative examination of *Es ist dies*-sentences in context, we observed many cases in which the antecedent is not located in the preceding sentence but found at a greater distance. The text passage in (4) exemplifies this.

- (4) (Last Sunday, his majesty, Jens-Peter I, the current champion marksman, planted **his royal tree** in the castle garden in Warberg, accompanied by the former majesties of the Warberg shooting association. In the context of the re-development of the castle garden, the shooters had decided that every reigning majesty should plant his or her own tree.)

3 It is a characteristic of German demonstratives in copular sentences that they can remain uninflected. Diessel (1999) uses the label demonstrative identifier for this use of demonstrative pronouns.

Es ist dies nun der 17. Baum der seinen Platz im Park findet.⁴
 it is this now the 17th tree that its place in the park finds
 ‘This is now the 17th tree that finds its place in the park.’

Here, *dies* refers back to the tree that is mentioned in the beginning of the short passage, not to the one directly preceding the *Es ist dies*-sentence, as *his or her own tree* is within the scope of a quantifier and therefore cannot be the antecedent of *dies*. Mentioning that the tree just planted is the 17th tree in (4) is an instance of taking stock. Such sentences of the form *It is this the nth...* were found quite often. We also found a number of occurrences of *Es ist dies*-sentences that express evaluative comments, like example (1). Frequently, they included superlatives, also like (1). Our findings led us to formulate the tentative hypothesis that an *Es ist dies*-sentence is used if an antecedent is not easily accessible, and therefore the pronoun *dies* does not constitute an optimal candidate to be located in the prefield. The construction possibly serves to mark this circumstance pragmatically. Its use can create an effect of distance to the preceding discourse or indicate a break in the text, potentially used to take stock. These hypotheses were the starting point of our empirical investigation, which is presented in the next section.

3 Corpus Study

3.1 Method

To investigate the use of this construction, we annotated 300 *Es ist dies*-sentences randomly taken from the DeReKo corpus of written German with regard to the following categories: a) metadata, i.e. the source the sentence occurred in and the region of the source, as well as b) properties of the antecedent of *dies*, such as the distance to the antecedent measured in finite and in matrix verbs. For example, in (4) above, the distance to the antecedent of *dies* measured in finite verbs is two and one if measured in matrix verbs. Moreover, we annotated c) semantic properties of the sentences in order to account for the impression that *Es ist dies*-sentences often express evaluations or are used to take stock. There were three categories regarding the semantic properties: *It is this the nth-constructions* (5), superlatives (6) and evaluative comments (7).

4 Braunschweiger Zeitung, 22.04.2013.

- (5) *Es ist dies das 23. Turnier seit 1994.*⁵
it is this the 23rd tournament since 1994
'This is the 23rd tournament since 1994.'
- (6) *Es ist dies der früheste Reisebericht über Afrika und Indien.*⁶
it is this the earliest travel report about Africa and India
'This is the earliest travel report about Africa and India.'
- (7) *Es ist dies eine heikle und bedauerliche Entscheidung.*⁷
it is this a precarious and regrettable decision
'This is a precarious and regrettable decision.'

The *Es ist dies*-sentences were compared to *Dies ist*-sentences, which can be regarded as the unmarked counterpart to the *Es ist dies*-construction and which were taken in the same proportion from the same sources as the *Es ist dies*-sentences to achieve a maximally exact mirroring.

3.2 Results

In our sample, *Es ist dies*-sentences occurred almost exclusively in texts⁸ from southern regions: 38% of the instances were found in texts from Switzerland, 32% were from Austria and 18% from Bavaria. The remaining 12% were singular occurrences in texts from various regions. The results of the measurement of the distance to the antecedent are presented in Table 1. In the majority of cases, both constructions feature an antecedent located at a distance of zero finite or matrix verbs. However, *Es ist dies*-sentences refer to antecedents located at larger distances more frequently than *Dies ist*-sentences. The χ^2 -test yielded significant differences between the two samples with regard to the feature 'distance to the antecedent' ($p < .01$ both for the measurement in finite verbs and for the measurement in matrix verbs).

Concerning the semantic properties of the sentences, we found 86 instances of *It is this the nth*-constructions in the *Es ist dies*-sentences opposed to 26 instances among the *Dies ist*-counterparts. The χ^2 -test yielded a significant result for this difference ($p < .001$), too. The other semantic properties that were annotated, superlatives and evaluative comments, did not yield significant differences.

5 Nordkurier, 09.09.2008.

6 Tiroler Tageszeitung, 12.11.1997.

7 Süddeutsche Zeitung, 05.04.2008.

8 96% of these texts were newspaper articles.

Table 1: Distance to the antecedent (absolute values)

Distance to the antecedent	Sentence Type	
	<i>Es ist dies</i>	<i>Dies ist</i>
Number of finite verbs		
0	198	239
1	43	20
2 or more	36	17
Number of matrix verbs		
0	219	246
1	39	20
2 or more	19	10

3.3 Discussion

Speyer's ranking (2008, 2009; Table 2) predicts that prefield-*es* is only used if a sentence contains none of the preferred prefield fillers which are scene-setting elements, poset elements and topics.⁹ As *dies* refers to the discourse referent the sentence is about, one could consider it to be the topic. In Speyer's prefield ranking, the notion 'topic' is defined as discourse-old and occurring in the directly preceding sentence. In the majority of cases, the *dies* of our *Es ist dies*-sentences does actually refer to an antecedent in the immediately preceding sentence. Hence, Speyer's ranking incorrectly predicts *dies*, instead of *es*, to occur in the prefield for those cases. His ranking only predicts *Es ist dies* to be the optimal candidate in cases in which *dies* refers to an antecedent that is not located in the preceding sentence.

However, *Es ist dies*-sentences are used rather rarely.¹⁰ Hence, the fact that the *Dies ist*-version is generally more frequent and that *Es ist dies* is a marked construction should be represented in the ranking. We therefore suggest replacing

9 The first constraint 1-VF specifies that only one constituent can occur in the prefield. SCENE-SETTING-VF requires elements such as adverbials of time to be moved to the prefield. 'POSET' stands for 'partially ordered set' and a poset relation is a type of contrast. In (5) *Fresh vegetables* and *pasta* stand in a poset relation as they are both members of the set 'food Peter buys'.

(5) Frisches Gemüse kauft Peter auf dem Markt.
 fresh vegetables buys Peter at the market
 'Fresh vegetables Peter buys at the supermarket.
 Nudeln besorgt er immer im Supermarkt.
 pasta gets he always at the supermarket
 Pasta he always gets at the supermarket.'

10 To illustrate this: Our search request for *Es ist dies*-sentences yielded 4,870 hits from the DeReKo as opposed to 91,726 hits for the corresponding request for *Dies ist*-sentences.

Table 2: Speyer’s (2008, 2009) prefield ranking

Candidates	1-VF	SCENE- SETTING-VF	POSET-VF	TOPIC-VF
☞ <i>Dies ist</i> _{Preced.Sent}				
<i>Es ist dies</i> _{Preced.Sent}				*
<i>Dies ist</i> _{NotPreced.Sent.}				*
☞ <i>Es ist dies</i> _{NotPreced.Sent}				

Table 3: Modified prefield ranking.

Candidates	1-VF	SCENE- SETTING-VF	POSET-VF	ABOUTNESS TOPIC-VF	MARK- SHIFT-VF
☞ <i>Dies ist</i> _{Preced.Sent}					
<i>Es ist dies</i> _{Preced.Sent}				*	
☞ <i>Dies ist</i> _{NotPreced.Sent}					*
<i>Es ist dies</i> _{NotPrecedSent}				*	
☞ <i>Dies ist</i> _{Count}					*
<i>Es ist dies</i> _{Count}				*	

TOPIC-VF with two different constraints, ABOUTNESSTOPIC-VF and MARKSHIFT-VF (Table 3). This is a first tentative approach to explain the observed phenomenon.

ABOUTNESSTOPIC-VF specifies that the topic, understood as an ‘aboutness’ topic following Reinhart (1981), should be located in the prefield. Unlike TOPIC-VF, ABOUTNESSTOPIC-VF does not require the topic to occur in the immediately preceding sentence which accounts for the higher frequency of *Dies ist*-sentences in general. The modified ranking specifies that, unless there is a reverse ranking, *Dies ist* is always the preferred candidate.¹¹ MARKSHIFT-VF reflects the

11 In stochastic OT (Boersma & Hayes 2001) constraints are not discrete but ordered on a continuous scale of strictness. A constraint is assumed to be associated not only with one value, but with a range of values which is thought of as a probability distribution in the form of a Gaussian curve. Thus, some values have a higher probability of being selected than others. Depending on how close to each other the constraints are located on the scale, the extent to which they overlap varies. A ranking in which two constraints overlap to a large degree accounts for cases where two forms are grammatical, but one is preferred over the other. In these cases, the probability that values are selected which result in a reverse ranking is relatively high.

discourse connecting function of the prefield (see e.g. Fillipova & Strube 2007). In the default case, the prefield is expected to be filled by an element that adds to the coherence of the text. The new constraint requires a marking of breaks or unexpected moves in discourse. It has often been observed that shifts of topics tend to be marked (see Givón 1983, Bestgen/Vonk 2000, and Breindl 2008, 2011).¹² Similarly, we argue that *Es ist dies* can mark a cesura in discourse, e.g. when *dies* refers to an antecedent that is not easily accessible. We assume that MARKSHIFT-VF slightly overlaps with ABOUTNESSTOPIC-VF, which has the effect that, at times, ABOUTNESSTOPIC-VF is outranked by MARKSHIFT-VF. This accounts for the difference between *Es ist dies*-sentences and *Dies ist*-sentences with regard to the distance to the antecedent since referring back to an antecedent that is located at a greater distance is an unexpected discourse move. The significant difference in the content category *It is this the nth* is also in line with our approach as it makes sense to indicate a break in discourse when taking stock (i.e., no violation of MARKSHIFT-VF).

However, for a large number of cases, we are not yet able to pinpoint the reason for using prefield-*es*. It might be related to the often mentioned observation that anaphorically used demonstrative pronouns tend to refer to an antecedent that is harder to access (e.g. Diessel 1999: 96, Gundel et al. 2003). Using prefield-*es* could be an optional way of further highlighting this. After all, we did not find constructions such as *Es ist er* + NP oder *Es ist sie* + NP where grammatical gender already limits the number of possible antecedents.

Furthermore, an interesting question is how our modified OT ranking relates to the regional differences. We suggest that in more northern varieties of standard German the two constraints ABOUTNESSTOPIC-VF and MARKSHIFT-VF are located far apart from each other on the scale of constraints and therefore overlap to a very small extent. This has the effect that ABOUTNESSTOPIC-VF outranks MARKSHIFT-VF more regularly, which is why speakers of northern varieties of German find *Es ist dies*-sentence odd but not ungrammatical.¹³ In contrast, in more southern varieties the stochastic overlap of the two constraints is stronger, which leads to a larger probability that MARKSHIFT-VF outranks ABOUTNESSTOPIC-VF. We can therefore understand regional variation as a purely stochastic difference without any modification of our proposed ranking.

12 For example, in German, one way to mark a shift of topic is inserting an adverbial in the so-called “Nacherstposition” in front of the finite verb (Breindl 2008, 2011).

13 This claim is based on the judgments from the authors of the paper as well as from other consultants from northern Germany.

4 Conclusion

The modified version of the prefield ranking incorporates a use of prefield-*es* that the old model did not factor in. It reflects that *Es ist dies*-sentences are a rarely occurring phenomenon, but it accounts for the fact that they do occur. The significant differences that were found between *Es ist dies*-sentences and their unmarked counterparts with regard to the distance to the antecedent of *dies* and the frequency of the content type *It is this the nth* were explained by the addition of the constraint MARKSHIFT-VF. What is still needed is an explanation of those occurrences of *Es ist dies*-sentences, for which neither a large distance to the antecedent nor the content type *It is this the nth* was attested. We leave this question for further research.

References

- Bestgen, Yves and Wietske Vonk. 2000. Temporal adverbials as segmentation markers in discourse comprehension. *Journal of Memory and Language* 42: 74–87.
- Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32(1): 45–86.
- Breindl, Eva. 2008. Die Brigitte nun kann der Hans nicht ausstehen. Gebundene Topiks im Deutschen. *Deutsche Sprache* 36: 27–49.
- Breindl, Eva. 2011. Nach Rom freilich führen viele Wege. Zur Interaktion von Informationsstruktur, Diskursstruktur und Prosodie bei der Besetzung der Nacherstposition. In Gisella Ferraresi (ed.), *Konnektoren im Deutschen und im Sprachvergleich. Beschreibung und grammatische Analyse*, 17–56. Tübingen: Narr.
- Cardinaletti, Anna. 1990. Es, pro and sentential arguments in German. *Linguistische Berichte* 126: 135–164.
- Diessel, Holger. 1999. *Demonstratives: Form, function and grammaticalization*. Amsterdam and Philadelphia, PA: Benjamins.
- Filippova, Katja and Michael Strube. 2007. The German Vorfeld and local coherence. *Journal of Logic, Language, and Information* 16(4): 465–485.
- Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In Talmy Givón (ed.), *Topic continuity in discourse: A quantitative cross-language study*, 1–42. Amsterdam et al.: Benjamins.
- Gundel, Jeanette K., Michael Hegarty and Kaja Borthen. 2003. Cognitive status, information structure, and pronominal reference to clausally introduced entities. *Journal of Logic and Information* 12: 281–299.

- Grosz, Barbara J., Aravind K. Joshi, & Scott Weinstein. 1995. Centering: A framework for modelling local coherence of discourse. *Computational Linguistics* 21: 203–225.
- Paranhos Zitterbart, Jussara. 2002. Zur Mittelfeldfähigkeit des Korrelat *es* in Verbindung mit Subjektsätzen. *Sprachwissenschaft* 27: 149–195.
- Pittner, Karin and Judith Bermann. 2004. *Deutsche Syntax: Ein Arbeitsbuch*. Tübingen: Narr.
- Pütz, Herbert. 1986. Über die Syntax der Pronominalform ‚es‘ im modernen Deutsch. Tübingen: Narr.
- Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27(1): 53–94.
- Speyer, Augustin. 2008. German vorfeld-filling as constraint interaction. In Anton Benz & Peter Kühnlein (eds.), *Constraints in discourse*, 267–290. Amsterdam and Philadelphia, PA: Benjamins.
- Speyer, Augustin. 2009. Das Vorfeldranking und das Vorfeld-es. *Linguistische Berichte* 219: 323–353.
- Tomaselli, Alessandra. 1986. Das unpersönliche „es“: Eine Analyse im Rahmen der Generativen Grammatik. *Linguistische Berichte* 102: 171–190.
- Walker, Marylin A., Aravind K. Joshi and Ellen F. Prince. 1998. Centering in naturally occurring discourse: An overview. In Marylin A. Walker, Aravind K. Joshi & Ellen F. Prince (eds.), *Centering Theory in Discourse*, 1–28. Oxford: Oxford University Press.
- Zifonun, Gisela. 1995. Minimalia Grammaticalia: Das nicht-phorische es als Prüfstein grammatischer Theoriebildung. *Deutsche Sprache* 23: 39–60.
- Zifonun, Gisela, Ludger Hoffmann, Bruno Strecker et al. 1997. *Grammatik der deutschen Sprache*, Bd. 2. Berlin: de Gruyter.