A southern German use of prefield-\textit{es}: Evidence from the corpus and an experimental study

Abstract: There is a use of the German third person neuter pronoun \textit{es} in the prefield, known as prefield-\textit{es}, which is characterized by being neither referential, nor an argument of the verb. According to Speyer’s (2008, 2009) optimality theoretic prefield ranking, this should only occur if a sentence contains no alternative element eligible to be moved to the prefield. This paper investigates a so far unnoticed use of \textit{es} in the prefield in combination with a demonstrative pronoun \textit{dies} and a copula verb \textit{ist}, which will be referred to as \textit{Es ist dies}-sentence. This construction is an instance of prefield-\textit{es}, but contravenes the expectations about the use of prefield-\textit{es} postulated by Speyer, since \textit{Es ist dies}-sentences do contain a suitable candidate to fill the prefield, the demonstrative pronoun \textit{dies}. In a corpus study, \textit{Es ist dies}-sentences are compared to a sample of \textit{Dies ist}-sentences. According to the corpus data, \textit{Es ist dies} occurs predominantly in southern dialects. Significant differences between the two samples concern 1) the distance to the antecedent of \textit{dies} and 2) the type of content of the sentence. An online rating study, that compared acceptability judgments of \textit{Es ist dies}-sentences between speakers from different regions, confirmed that \textit{Es ist dies}-sentences are a phenomenon of southern dialects. In the light of these results, a modification of Speyer’s (2008, 2009) Stochastic OT model is proposed.

Keywords: Prefield-\textit{es}, southern German dialects, Stochastic Optimality Theory, corpus study, rating study

1 Introduction

Language use can be expected to be economic, since unnecessary material adds to the costs of articulation and processing (e. g. Moser 1971). Some uses of the pronoun \textit{es} in the prefield seem to contravene this expectation. Compare the use of \textit{es} in (1) and (2). In both cases, \textit{es} is not referential. But unlike in (1), where \textit{es} is used
as a formal argument and may therefore not be considered unnecessary material, 
the occurrence of *es* in (2) is not required by the verb.

(1)  *Es regnet.*  
      it rains  
      ‘It is raining.’

(2)  *Es stehen vierzehn verschiedene Eissorten zur Verfügung.*  
      it stand fourteen different ice-cream.flavors at disposal  
      ‘Fourteen different ice-cream flavors are available.’

The fact that *es* disappears when the subject is located in the prefield, shown in 
(3), provides further support for the impression that *es* is redundant in constructions like (2).

(3)  *Vierzehn verschiedene Eissorten stehen zur Auswahl.*  
      fourteen different ice-cream.flavors stand at disposal  
      ‘Fourteen different ice-cream flavors are available.’

Speyer (2009) explains the occurrence of this use of *es*, known as prefield-*es*, 
within the framework of Optimality Theory. According to his approach, prefield-*es* 
is used as a “last resort” to fill the prefield if a sentence contains no eligible elemen-
t to be moved to this position (Speyer 2009: 334). Speyer argues that this is 
the case when no discourse referent can be considered topic because no referent 
is relevant in the previous or subsequent discourse. This correctly predicts that (2) 
may be preferred over (3) in certain contexts.

There is, however, a use of *es* in the prefield, exemplified in (4), which consti-
tutes an even more severe violation in terms of language economy.

(4)  *Es ist dies der schwerste Fall von Marktmanipulation, den wir je*  
      it is this the worst case of market.manipulation that we ever 
      seen have  
      ‘This is the worst case of market manipulation that we have ever seen.’  
      (https://www.tagesanzeiger.ch/wirtschaft/unternehmen-und-konjunktur/Es-ist-der-schwerste-Fall-den-wir-je-gesehen-haben-/story/31098247)

This occurrence of *es* in the prefield is not only intuitively peculiar because the 
combination of *es* and *dies* at the beginning of the sentence appears cluttered. 
It is also unexpected with regard to Speyer’s theory, since the sentence features 
a subject that seems perfectly suitable to be placed in the prefield: the demon-
strative pronoun *dies*. This phoric expression clearly connects to the previous dis-
course.
Nevertheless, sentences like (4), albeit being not very frequent, are indeed used in written contemporary German and have already been used by Goethe. They have, however, not been studied before in the linguistic literature. The aim of this paper is to contribute to closing this research gap.

The construction will be referred to as *Es ist dies*-sentence, and I will use *Dies ist*-sentence as a label to refer to its unmarked counterpart. The paper presents a corpus study, which compared *Es ist dies*- and *Dies ist*-sentences. Furthermore, an online rating study investigated regional differences in the acceptability of this construction. Based on the results of these studies, a modification of Speyer’s prefield ranking is proposed.

The paper is structured as follows. Section 2 shows that the *es* in question really is a case of prefield-*es* by briefly reviewing the characteristics of different types of *es*. Section 3 presents Speyer’s analysis of prefield-*es* and shows that the occurrence of *Es ist dies* cannot be explained by his theory. Section 4 presents two empirical studies, a corpus study and a rating study, and proposes a modification of the prefield ranking. Finally, Section 5 summarizes the results and gives an outlook.

### 2 The type of *es* in *Es ist dies*-sentences

In order to establish that *es* in *Es ist dies*-sentences is an instance of prefield-*es*, this section reviews characteristics of different types of *es* and compares their properties to those of *Es ist dies*-sentences.

Apart from the referential pronoun *es*, there are three types of *es* that can occur in the prefield: *es* as a formal argument (5), correlate-*es* (6), and

---

1 For example, *Es ist dies* (Christus auf dem Meere wandelnd) eine der schönsten Legenden, die ich vor allen lieb habe. (Johann Wolfgang von Goethe, in: Goethes Selbstzeugnisse über seine Stellung zur Religion und zu religiös-kirchlichen Fragen, ed. by Theodor Vogel, Leipzig 1888, p. 151).

2 It should be noted that *Es ist dies* is not an idiom. The construction occurs with a range of variation: There is a variant featuring *das* instead of *dies* and the construction may surface with different inflected verb forms and with or without a relative clause. In this paper, the construction will be investigated based on the variant *Es ist dies* followed by a definite or indefinite noun phrase.

3 It is obvious that *es* in *Es ist dies* is not referential. In theory, both *es* and *dies* could occur alone in the sentence and serve as a referential expression. However, as there are the two of them in *Es ist dies*-sentences, only one can fulfill this function. Demonstratives are unmistakably referential expressions that (re)orient the reader’s or hearer’s attention to discourse referents. In contrast, *es* is usable as a semantically vacuous ‘dummy’ element.
prefield-\textit{es} (7).\textsuperscript{4}

(5) \textit{Es schneit.} \\
\textit{it snows} \\
‘It is snowing.’

(6) \textit{Es stört Fred, dass Erika in der Wohnung raucht.} \\
\textit{it bothers} Fred\textit{ that} Erika in the apartment\textit{ smokes} \\
‘It bothers Fred that Erika smokes in the apartment.’

(7) \textit{Es hängen drei Gemälde an der Wand.} \\
\textit{it hang} three\textit{ paintings} on the wall \\
‘Three paintings are hanging on the wall.’

They have in common that they do not contribute semantically to the proposition expressed by the sentences that contain them. Their characteristic properties are listed in Table 1 (more detailed accounts are given in Paranhos Zitterbart 2002; Pittner and Berman 2004; Sternefeld 2008; Tomaselli 1986, and Zifonun 1995).

\textbf{Table 1: Three types of semantically vacuous \textit{es}.}

| \textit{es} as a formal argument | \textit{correlate-\textit{es}} | \textit{prefield-\textit{es}} |
|----------------------------------|-------------------------------|-----------------------------|
| Formally fulfils function of a subject or object | Correlate to a dependent subject or object clause in the final field | Not an argument, no obvious function except filling of the prefield |
| Agrees with verb | Agreement with verb unclear, presumably \textit{es} and the clause together form a discontinuous constituent | No agreement with the verb |
| Can occur in the middle field if it fulfils the function of a subject | Can occur in the middle field if it serves as a correlate to a subject clause in the final field | Cannot occur in the middle field |
| Demanded by certain verbs that subcategorize an argument without assigning a theta role | Occurs in most cases optionally with verbs that select a subject or object clause | Not limited to certain verbs |

\textsuperscript{4} With regard to prefield-\textit{es} and \textit{es} as a formal argument, terminology differs: Zifonun (1995) labels prefield-\textit{es} as expletive. Pittner and Berman (2004), like Speyer (2009), use the same label for \textit{es} as a formal argument. The terminology used here was chosen in a way that most transparently reflects the role of the respective type of \textit{es}. 
Example (8-a) demonstrates that *es* in *Es ist dies*-sentences can be omitted if another element is moved to the prefield, which shows that it is not an argument. Also, *sein* ‘to be’ does not belong to the group of verbs that subcategorize formal arguments. Moreover, there is no indication that it is associated with another constituent in the sentence unlike correlate-*es*. The ungrammatical sentence (8-b) demonstrates that *es* in *Es ist dies*-sentences cannot occur in the middle field, which is another defining feature of prefield-*es* that sets it apart from the other two types of *es*.

(8) *Es ist dies die erste Inszenierung des 34-jährigen Josua Rösing, der*  
   *it is this the first production of the 34-year-old Josua Rösing who*  
   *seit knapp drei Jahren als Regieassistent am Deutschen Theater*  
   *since almost three years as assistant director at the German theater*  
   *arbeitet.*  
   *works*  
   ‘This is the first production of the 34-year-old Josua Rösing, who has been working as assistant director at the German Theater for almost three years.’  
   (https://www.deutschestheater.de/programm/a-z/das-feuerschiff/)

a. *Dies ist die erste Inszenierung des 34-jährigen Josua Rösing,...*  
   *this is the first production of the 34-year-old Josua Rösing*  

b. *Dies ist es die erste Inszenierung des 34-jährigen Josua Rösing,...*  
   *this is it the first production of the 34-year-old Josua Rösing*

Thus, *es* in *Es ist dies*-sentences clearly conforms to the behavior of prefield-*es*.

3 Speyer’s (2008, 2009) analysis of prefield-*es* and *Es ist dies*

In a first step, this section outlines Speyer’s prefield ranking and his account of prefield-*es*. In a second step, it will be shown why Speyer’s ranking has difficulties explaining the occurrence of *Es ist dies*-sentences.

3.1 Speyer’s (2008, 2009) approach

In the previous section, it became apparent that the notion prefield-*es* is defined negatively by what it is not and by the position of *es* in the clause, which evokes the question why prefield-*es* exists at all. The reason for using non-phoric *es* in the prefield instead of the subject is presumed to lie in the information structure.
According to Pittner and Berman (2004: 130) and Zifonun (1995: 55–56) the use of prefield-es enables the positioning of new, focused information towards the end of the clause. Speyer (2009) explains the occurrence of prefield-es within the framework of Optimality Theory. In a corpus study, Speyer (2008) identified three types of elements as preferred prefield fillers. In around 90% of the 405 sentences that were analyzed, a scene-setting element, a poset, i.e., a member of a “partially ordered set” (Hirschberg 1985: 122; Prince 1999; cf. Speyer 2008: 279), or a topic occurred in the prefield (Speyer 2008: 273). In order to find out which of these is most likely to be located in the prefield, he analyzed sentences that featured more than one type of these elements. Speyer (2008: 287) modeled his findings in the following optimality theoretic ranking (9).

(9) 1-VF >> SceneSetting-VF >> Poset-VF >> Topic-VF

The highest ranked constraint, 1-VF, specifies that only one constituent can occur in the prefield. The constraint that is ranked the second highest, SceneSetting-VF, requires that so-called scene-setting elements are moved to the prefield. They are usually adverbials of place or time, and they state a “crucial restriction on the situation in which the proposition is true” (Speyer 2008: 280). The second type of elements that preferably occur in the prefield are elements that stand in a poset relation. A poset relation is a special type of contrast (Speyer 2007: 104‒109, Speyer 2008: 277‒280). Example (10) illustrates this: In (10), Eine Kiste mit häßlichem Geschirr and Ihre alte Lederjacke stand in a poset relation as they are both members of the set “things that got sorted out by Erika”.

(10) Erika hat am Wochenende ihre Wohnung entrümpelt. Eine Kiste mit häßlichem Geschirr hat sie an die Straße gestellt. Ihre alte Lederjacke will sie auf dem Flohmarkt verkaufen.

Erika has on the weekend her apartment tidied out. A box with ugly tableware has she on the street put out. Her old leather jacket wants she on the flea market sell. ‘On the weekend, Erika tidied out her apartment. A box with ugly tableware, she put out on the street. Her old leather jacket, she wants to sell on the flea market.’

The third constraint, Topic-VF, specifies that the topic should be located in the prefield. Speyer (2008: 276) defines the notion of topic based on three criteria: 1) it represents “discourse-old information”, 2) it “functions as heading to which
the sentence in question adds information”, 7 and 3) it “conforms to the definition of backward-looking center”.

It can be argued that the first two criteria are contained in the third. Also, it is the topic conception of Centering Theory (Grosz et al. 1995; Walker et al. 1998) that Speyer primarily refers to (see also Speyer 2007). The backward looking center of Centering Theory – which is “similar to what is elsewhere called topic” (Walker et al. 1998: 3) – is the semantic entity the utterance is about. It creates coherence, as it has already been evoked in the preceding utterance. Therefore, it is salient and often realized as a pronoun (Walker et al. 1998: 4). Speyer (2009: 339) relativizes the claim that a topic has to be discourse-old in order to be allowed in the prefield by stating that a topic has to be relevant with regard to the macrostructure of the text. This means that it must be discourse-old unless it constitutes the topic in the following discourse. The three constraints of the prefield ranking can be assumed to overlap to some extent in terms of Stochastic Optimality Theory (OT),8 since

7 This criterion relates to the notion of aboutness topic (Reinhart 1981).
8 Stochastic Optimality Theory (Boersma and Hayes 2011) is an advancement of Optimality Theory (Prince and Smolensky, 2004 [1993]). In Stochastic OT, it is postulated that constraints are located on a continuous scale of strictness (Boersma and Hayes 2011: 47–49): The higher the constraint is ranked, the higher the value it is located at. A constraint is assumed to be associated not only with one value, but with a range of values, some of which have a higher probability of being selected than others. The ranges of constraints are thought of as probability distributions in the form of Gaussian curves with each constraint’s range having the same standard deviation. As all values of the normal distribution are above zero, constraints overlap. Figure 1 exemplifies this. Here, C1 would outrank C2 in the majority of cases, i.e. when selection points in the white areas are chosen. However, a reverse ranking is possible if the selection points lie in the area of the overlap, as shown in Figure 2.
they do not describe absolute rules but model preferences for the filling of the prefield (Speyer 2009: 333).

According to the ranking, prefield-es should only occur as a “last resort” in order to yield a verb-second clause if none of the preferred candidates for filling the prefield are available (Speyer 2009: 334). Speyer conducted a corpus study, in which he analyzed 190 instances of prefield-es to test his prediction. He reports that all the sentences featuring prefield-es contained relatively few constituents, often only the subject or the subject and an adverbial adjunct. The reason for this is that the verbs of the sentences were intransitive, reflexive or passivized so that they did not allow for an object constituent (Speyer 2009: 334–335).

Speyer’s analysis furthermore confirmed that prefield-es sentences, in fact, did not contain any of the elements that he lists as being attracted to the prefield. The subjects of prefield-es sentences were found to be ineligible for topic status as they represented information that was discourse-new and not prominent in the following discourse; in many cases they were not even mentioned again. These sentences often related additional information (Speyer 2009: 335).

3.2 *Es ist dies* and the prefield ranking

Turning to the question whether *Es ist dies*-sentences are compatible with Speyer’s theory, it has to be stated that the ranking neither considers *Es ist dies*-sentences, nor can it account for their occurrence. *Es ist dies*-sentences deviate from Speyer’s characterization of prefield-es in several respects. They feature more constituents than attested for typical prefield-es sentences. Since they are copular sentences, there is always a further element apart from the subject. Moreover, Speyer’s account that the verbs of prefield-es sentences are intransitive, reflexive or passivized is a characterization that is not appropriate for *Es ist dies*-sentences. Crucially, it is not obvious how *Es ist dies*-sentences conform to the principle that prefield-es only occurs if there is no suitable element that could fill the prefield. According to Speyer, prefield-es is used if none of the sentence’s discourse referents is connected either to the previous or to the subsequent discourse and therefore none of them conforms to the criteria that are decisive for being considered the topic. However, in taking up a referent of the previous discourse, the demonstrative pronoun *dies*, that is the subject in

The advantage of Stochastic OT over “regular” OT is that it explains not only absolute differences in grammaticality but also those cases where two or more forms are grammatical, but one is usually preferred over the other. This is represented by constraints that overlap to a large degree.
Es ist dies-sentences, seems to be perfectly suitable for being placed in the prefield, which is considered a position where coherence-creating expressions are located (Filippova and Strube 2007: 482; Speyer 2007: 111). As the prefield ranking in its current form cannot explain Es ist dies-sentences, the theory needs to be modified.

4 Empirical studies

4.1 Corpus study

In order to further investigate the use of the phenomenon and in search of an explanation of this use of prefield-es, a corpus study that considers Es ist dies-sentences in context was conducted. The study is presented in the following. Based on its results, two modifications to the prefield ranking are proposed.

4.1.1 Introduction

When examining Es ist dies-sentences in context, several cases were observed in which, like in (11), the antecedent of dies does not occur in the directly preceding sentence but is located at a larger distance.

(11) (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 sitting majesty should plant his or her own tree.)

Es ist dies nun der 17. Baum, der seinen Platz im Park findet.

‘This is now the 17th tree that finds its place in the park.’

(Braunschweiger Zeitung, 22.04.2013)

In (11), dies refers back to the tree that is mentioned in the first sentence of the passage, not to the one directly preceding the Es ist dies-sentence. In the example, the first sentence reports the event of the champion planting his tree and the sec-

9 Parts of the corpus study have been presented in Fricke and Tönnis (2018).
ond sentence gives background information on this ritual. The *Es ist dies*-sentence redirects attention to the tree in question. The example displays a ubiquitous feature of *Es ist dies*-sentences, namely a construction of the form ‘It is this the n^th...’, which indicates a summarizing strategy.

The use of this construction is also exemplified in (12). In addition, *dies* in (12) does not refer back to a noun phrase but rather to the content of several sentences, which describe an accident and its consequences. This is an instance of discourse deixis.

(12) (The flags of the nations participating in the bob-world championship in Lake Placid are flying at half-mast. During the training session for the title races, the two-man bobsled of the Belgians Max Houven and Jaques Mouvét ran out of control in a bend and both athletes fell out of their bobsled. The 42-year-old Belgian top athlete Max Houven was severely hurt and died after a few minutes. His fellow countryman Mouvét is in a critical condition.)

Es ist dies seit den Olympischen Winterspielen 1932 der erste
tödliche Unfall auf der Mount Hövenbergbahn.

‘Since the Winter Olympics in 1932, this has been the first fatal accident on the Mount Hövenbergbahn.’

(Neues Deutschland, 12.02.1949, p. 4)

Moreover, there were a number of cases featuring evaluative comments like (13). (13) also exemplifies discourse deixis. Interestingly, here *dies* appears to refer back to the content of the heading, which relates the introduction of a new regulation. The preceding text does not mention the regulation explicitly. Instead, it discusses its disadvantageous repercussions.

(13) (Lately, drivers have been obliged to drive with the lights on during daytime.
In the “good old times” when I was learning to drive, you would switch on the lights only when it seemed necessary: when it darkened or in misty conditions. That, by the way, also saved a bit of energy. I don’t know how the current practice came into existence. Is the union of the driving instructors responsible, some experts or someone else? Like I said, I don’t know. What I do know, though, is that what is now obligatory has blatant disadvantages, which clearly outweigh the alleged advantages: Depending on the situation, approaching drivers are dazzled; the police is kept
from focusing on other offenses, such as pulling out of a roundabout without turning on the indicator or ignoring cross-walks. I do not see any advantages of lights that are permanently switched on.)

*Es ist dies ein klassischer Fall von Regulierungswut.*

‘This is a classic case of regulation mania.’

(A liberal system trusts the prudence of the individual and only interferes where it is inevitable.)

(http://www.weltwoche.ch/weiche/hinweisgesperrt.html?hidID=550303, 31.03.2017)

Furthermore, in many examples, superlatives are used, as shown in (14). In this example, the *Es ist dies*-sentence classifies the severity of the fire, which constituted the subject matter throughout the four preceding sentences.

(14) (So far, hardly more than 40 percent of the devastating brushfire in the US-state New Mexico could be brought under control. For the time being, the greatest danger for the nuclear power plant is thus averted. The fire approached to one and a half kilometers to the center where the first nuclear bombs had been built. Numerous ancient Native American cult sites and settling areas in the Bandelier national park have already fallen victim to the flames.)

*Es ist dies der schlimmste Brand in der Region seit 20 Jahren.*

‘This has been the worst fire in the region for 20 years.’

(Tiroler Tageszeitung, 02.05.1996)

Sentential antecedents for *dies* are indeed quite common (e.g. Diessel 1999: 100–103; Gundel et al. 2004). Moreover, demonstratives are characterized as devices that point to antecedents that are not very salient, or that even establish discourse referents themselves (Webber 1991: 111–112). Compared to personal pronouns, demonstratives serve to reorient attention since they have the tendency to not refer to topics (Bosch and Umbach 2007)\textsuperscript{10} or to perspectival centers (Bosch and Hinterwimmer 2016).

However, it is not clear how distant antecedents of demonstrative pronouns are usually located and how large the discourse portions are that can be referred to via *dies*. Graefen (1997: 220–223) states that in the texts she examined in a corpus

\textsuperscript{10} I thank an anonymous reviewer for bringing this paper to my attention.
study *dies* usually takes up the content of only one sentence; another study by Koeppel (1993: 258–268) attests that *dies* expresses “proximity” and claims that it is therefore located close to its antecedent.

The examples for the use of *Es ist dies*, in which *dies* refers to the content of several sentences and to antecedents that cannot be found in the immediately preceding sentence could indicate that the construction is used if the entity that is referred to by *dies* is hard to access. A pronominal expression, like *dies*, in the prefield creates coherence because it establishes a connection to a previously evoked referent. Inserting *es* in the prefield instead of *dies* can be seen as a disruptive signal that the antecedent of *dies* is not easily accessible. An *Es ist dies*-sentence for which the antecedent of *dies* is located far away, constitutes the only exception that Speyer’s ranking could explain, since in his definition the notion of topic is defined as being discourse-old and the topic must occur in the directly preceding sentence.\(^1\)

With regard to the content of the examined text excerpts, it is noticeable that in (11) and in (12) the *Es ist dies*-sentences literally take stock of the total number of trees in (11) and of accidents in (12). Moreover, the statements “This is the 17\(^{th}\) tree” (11), “This is the first fatal accident” (12) and “This is the worst fire in the region” (14) are subsuming comments that classify the subject matter of the previous sentences. The effect of the *Es ist dies*-construction here seems to be that it indicates a break in discourse, i.e. it marks the transition from an elaboration or a description to a classifying statement. *Es ist dies*-sentences might, thus, be a means of making the organization of discourse transparent.

### 4.1.2 Method

A total of 300 *Es ist dies*-sentences, randomly taken from the DeReKo\(^1\) corpus of written German (corpus “W-öffentlich”), were manually annotated by the author of this paper. They were the results of a corpus search for the combinations shown in (15).

\[
(15) \quad \text{Es ist dies der/ die/ das/ ein/ eine} \\
\text{it is this the.masc the.fem the.neut a.masc/neut a.fem}
\]

These sentences were compared to a sample of 300 *Dies ist*-sentences, which can be regarded as the unmarked counterpart of the *Es ist dies*-construction.

---

\(^{11}\) *Dies* can be considered the topic, as it refers to the entity the sentence is about.

\(^{12}\) *Das Deutsche Referenzkorpus DeReKo*, http://www.ids-mannheim.de/kl/projekte/korpora/, am Leibniz-Institut für Deutsche Sprache, Mannheim.
They were taken in the same proportion from the same sources as the *Es ist dies*-sentences to mirror their distribution in the sample. That is, if there were 20 instances of *Es ist dies*-sentences in the *Süddeutsche Zeitung*, 20 instances of *Dies ist*-sentences from the *Süddeutsche Zeitung* were annotated as well. Sentences that are part of a quotation were not annotated, because in those cases *dies* might refer to an antecedent not included in the quoted portion. Furthermore, excluded were *Wikipedia* articles, as those are potentially written by several authors, which could affect the coherence of the text. Moreover, it is not possible to collect metadata for *Wikipedia* articles. Minutes of plenary sessions, however, were included in the annotation. Even though the discourse they report is oral, they were estimated to display properties of conceptually written discourse (Koch and Oesterreicher 1994). The following categories were annotated:

**Metadata**

The medium that the sentence occurred in as well as its geographical origin were annotated to examine whether *Es ist dies*-sentences could be a dialectal phenomenon. Even if a direct relation between the medium’s region of origin and the writer’s dialect cannot be taken for granted, especially for nation-wide newspapers like *Süddeutsche Zeitung*, it is assumed that should *Es ist dies*-sentences be a dialectal phenomenon, they would rather be used in texts written in and for regions where that dialect is spoken.

**Properties related to the antecedent**

As several instances of *Es ist dies*-sentences with a large distance to the antecedent were observed, the distance to the antecedent of *dies* was measured in terms of the number of both finite and matrix verbs. In addition, the type of the antecedent was annotated, i.e. whether it is, for example, a definite or an indefinite noun phrase or the content of a sentence. It was expected that *Es ist dies*-sentences subsume the content of larger parts of text than *Dies ist*-sentences. Finally, the grammatical function of nominal antecedents was annotated, too.

**Content-related properties**

The content-related properties that had frequently been observed in *Es ist dies*-sentences, namely evaluative comments, superlatives, and ‘It is this n\textsuperscript{th} ...’-constructions (‘This is the n\textsuperscript{th} ...’ in the case of *Dies ist*) were annotated as well. Whereas the latter two properties are easy to identify, evaluative comments were defined as subjective comments that express a positive or negative evaluation.
For the inferential statistical analysis the software R (R Core Team 2017) was used.

4.1.3 Results

Metadata
The vast majority of Es ist dies-sentences was found in newspaper articles (96 %). However, it should be noted that the percentage of newspaper articles in the DeReKo is 98.73 % according to a publication by Bubenhofer et al. (2014: 66). The remaining instances occurred in minutes of parliamentary sessions (4 %). Es ist dies-sentences occurred predominantly in texts from southern regions: 38 % of the total amount were found in texts from Switzerland, 32 % in texts from Austria and 18 % in texts from Bavaria. The remaining 12 % were scattered occurrences in texts from various regions. In the corpus, the percentage of texts that are classified as being from a southern region (including Austria and Switzerland) is 32.33 %. 69.61 % of the texts in the DeReKo are classified as being from Germany, 21.89 % are from Austria and 7.56 % are from Switzerland (Bubenhofer et al. 2014: 69–70).

Properties related to the antecedent
The different types of antecedents are listed in Figure 3.

Figure 3: Antecedent types.
It is notable that the *Es ist dies*-sentences more often have a nominal antecedent, like in (16), than the *Dies ist*-sentences. The $\chi^2$-test yielded that this difference is significant ($\chi^2(1) = 20.22$, $p < 0.001$). Sentential antecedents, as in (17), however, are more frequently attested for *Dies ist*-sentences. This difference is significant, too ($\chi^2(1) = 17.35$, $p < 0.001$). The further differences concerning the antecedent type are not statistically significant.\footnote{Fisher’s exact test yielded $p = 0.34$ with regard to the deictic cases, $p = 0.14$ for the category “postcedent” and $p = 0.14$ for the category “abstract or implicit entity.”}

(16) (After the SVP Graubünden had to start over three years ago, they are delighted about the great outcome of last Sunday’s election.)

*Es ist dies das Resultat einer konsequenten Sachpolitik ohne personal Angriffe.*

‘This is the result of a strictly objective policy without personal attacks.’

(Die Südostschweiz, 26.10.2011, p. 22)

(17) (Former Bundesliga player Walter Gründahl, initially general secretary, and, after the president retired due to health problems, between 1991 and 1994, the leader of the eighth largest German sports association with 700 000 members, is supposed to replace his successor Hans Giesecke.)

*Dies ist der Vorschlag eines vor drei Monaten mit der Kandidaten-Suche beauftragten Gremiums, dem die Vorsitzenden der fünf größten Landesverbände aus Bayern, Hessen, Niedersachsen, Westdeutschland und Württemberg angehören.*

‘This is what was proposed by a committee commissioned with the search for a candidate three months ago. The committee is composed of the chairmen of the five largest state associations from Bavaria, Hesse, Lower Saxony, Western Germany and Württemberg.’

(Frankfurter Rundschau, 13.09.1999, p. 35)
The sentential antecedents can be split up further into ‘content of one sentence’ and ‘content of two or more sentences’. Here, relative to the total numbers of sentential antecedents, _dies_ of _Es ist dies_-sentences refers a little more often to larger units than _dies_ of _Dies ist_-sentences, whereas with regard to the content of one sentence _Dies ist_-sentences were slightly more frequent. Statistically, these differences are not significant ($\chi^2(1) = 2.37, p = 0.12$). The percentages are shown in Figure 4.

![Figure 4: Sentential antecedents in percent (absolute numbers).](image)

Figure 5 displays the grammatical functions of the nominal antecedents in percent. With regard to this category, it stands out that objects are relatively more frequently the antecedents of _Es ist dies_-sentences than of _Dies ist_-sentences. This difference is significant ($\chi^2(1) = 7.69, p < 0.01$). Another difference was found concerning the percentage of adjuncts. Here, slightly more cases are attested for _Dies ist_, but this difference is not statistically significant ($\chi^2(1) = 2.25, p = 0.13$).

Figure 6 shows the distance to the antecedent measured in terms of the number of both finite and matrix verbs for the whole sample. _Dies ist_-sentences occur more often at a distance of zero finite or matrix verbs than _Es ist dies_-sentences. _Es ist dies_-sentences, however, occur more often than _Dies ist_-sentences at distances of one and two or more finite or matrix verbs.

The $\chi^2$-test yielded significant differences between the two samples concerning the distance to the antecedent, both for the measurement in terms of finite verbs ($\chi^2(2) = 15.76, p < 0.001$) and in terms of matrix verbs ($\chi^2(2) = 8.27, p < 0.05$). This analysis considers all types of antecedents together. Ideally, an analysis with a generalized mixed model should be performed taking different types of antecedents into account. However, due to the structure of the data set, this is not possible. There are too many categories with only few observations, so that there is not enough variance to use such a model – even when the data set is re-
Figure 5: Grammatical function of the nominal antecedents in percent (absolute numbers).

Figure 6: Distance to antecedents of all types in percent (absolute numbers).
duced to sentences of the antecedent categories “noun phrase” and “sentential antecedent”.

When only sentential antecedents are considered, the same pattern can be observed: *Es ist dies*-sentences more often have a long distance to the antecedent than *Dies ist*-sentences (see Figure 7). An example of a sentential antecedent located at a distance of more than two verbs is shown below in (18). Fisher’s exact test yields that the differences are significant both for the distance measured in terms of finite verbs ($p < 0.01$), and for the distance measured in terms of matrix verbs ($p < 0.001$).

(18) (It is a risky scenario: The troop extends the mandate customary for blue helmets to interpose themselves between the war parties – violently if necessary – and becomes a war party itself. Martin Kobler from Germany, head of the UNO-mission (Monusco), seems to confirm this. Last Sunday, he visited a military hospital in Goma and praised injured Congolese soldiers: “You are heros, we are proud of you.” In an interview with the French radio station RFI, Kobler was even more explicit: “We fought in a very determined manner alongside the Congolese forces.”)

*Es ist dies eine Eskalation im wohl komplexesten Krisengebiet Afrikas.*

it is this an escalation in the presumably most complex

*Krisengebiet Afrikas.*

conflict area of Africa
‘This is an escalation in the presumably most complex conflict area of Africa.’
(St. Galler Tageblatt, 29.08.2013, p. 9)

Concerning the nominal antecedents, however, relative differences are smaller when measured in terms of finite verbs. With regard to the distance measured in terms of matrix verbs, the results deviate from the pattern observed before, as there is no difference between the samples for the distance “1” (see Figure 8). Furthermore, Es ist dies-sentences are used slightly more often to refer to antecedents located at a distance of 0 matrix verbs, whereas there are somewhat more Dies ist-sentences that refer to an antecedent found at a distance of 2 or more matrix verbs. The differences for finite verbs ($\chi^2(2) = 0.66, p = 0.72$) and for matrix verbs ($\chi^2(2) = 0.24, p = 0.86$) are not significant.

![Figure 8](image-url)  
**Figure 8:** Distance to nominal antecedents in percent (absolute numbers).

**Content-related properties**

With respect to the content-related properties (see Figure 9) of the Es ist dies- and Dies ist-sentences, there is a large difference between the two samples with regard to the category ‘It is this/This is the n\textsuperscript{th}…’. This type is much more common among Es ist dies-sentences. The $\chi^2$-test yielded a significant result for this difference ($\chi^2(1) = 41.35, p < 0.001$). In contrast, the numbers of superlatives and evaluative comments\(^{14}\) are almost identical.

\(^{14}\) The category “evaluative comment” turned out to be difficult to annotate objectively, as it required a large amount of interpretive activity.
4.1.4 Discussion

Beginning with the metadata, the fact that *Es ist dies*-sentences were almost exclusively found in newspaper articles is no sound evidence that they are preferably used in press texts given the high percentage of newspaper articles in the corpus. However, the metadata provide a strong indication that *Es ist dies* is a dialectal phenomenon of southern German varieties. Taking the composition of the corpus into account, the differences between the regions are striking – even though, as stated above, a one-to-one relation between the medium's region of origin and the writer's dialect cannot be taken for granted.

Of particular interest was the aspect of “distance to the antecedent”. Concerning this category, a significant difference between the two samples was found when all types of antecedents were considered together. With regard to sentential antecedents tested in isolation, there was a significant difference, too. However, nominal antecedents tested in isolation showed no significant difference concerning this aspect. The use of *Es ist dies*-sentences for the purpose of referring to antecedents located at a greater distance would be in accord with Speyer’s ranking as shown in Table 2. However, *dies* in *Es ist dies*-sentences mostly refers to an antecedent in the immediately preceding sentence.

There is yet another problem with the prefield ranking: It predicts that *Es ist dies* is the optimal candidate if *dies* refers to an antecedent that is not located in the preceding sentence. However, even in texts from southern regions both *Es ist dies* and *Dies ist* are used in these cases. Moreover, a search in the DeReKo for *Es ist dies*-sentences shows the low frequency of *Es ist dies*-sentences, as it
yields 4870 hits as opposed to 91726 hits for the corresponding search for Dies ist-sentences. Thus, Es ist dies is a rare structure. This is not reflected in the ranking (see Table 2).

Therefore, the ranking needs to be modified in order to account for the fact that the Dies ist-version is generally more frequent and that Es ist dies is a marked construction. To this end, I propose to replace the constraint Topic-VF, which requires macrostructural relevance of the topic, by the constraint ABOUTNESSTopic-VF (19). This has the effect that the ranking no longer states that Es ist dies-sentences are optimal in cases of a long-distance antecedent. The reason is that the notion of topic is understood as an aboutness topic, which merely stands for the entity the sentence is about.

(19) ABOUTNESSTopic-VF: Place aboutness topic in the prefield

In the new version of the prefield ranking, shown in Table 3, the use of es in the prefield constitutes a violation when dies is available irrespective of the location of its antecedent because the discourse referent dies refers to is the aboutness topic. Thus, now the ranking adequately predicts that Es ist dies is marked and less frequently used than Dies ist.

What is still missing in the modified ranking is an explanation as to why Es ist dies occurs at all. It was hypothesized that inserting es in the position where normally a discourse connecting demonstrative pronoun would occur, is a disruptive signal that indicates that the antecedent is not easily accessible. The corpus

---

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

| Candidates | 1-VF | SCENE-SETTING-VF | POSET-VF | TOPIC-VF |
|------------|------|------------------|----------|----------|
| Dies ist | Preced.Sent | | | |
| Es ist dies | Preced.Sent | * | | |
| Dies ist | NotPreced.Sent | | | |
| Es ist dies | NotPreced.Sent | * | | |

**Table 3:** Prefield ranking featuring a new topic constraint.

| Candidates | 1-VF | SCENE-SETTING-VF | POSET-VF | ABOUTNESSTopic-VF |
|------------|------|------------------|----------|-------------------|
| Dies ist | Preced.Sent | | | |
| Es ist dies | Preced.Sent | * | | |
| Dies ist | NotPreced.Sent | | | |
| Es ist dies | NotPreced.Sent | * | | |
data, indeed, showed a difference between *Es ist dies*- and *Dies ist*-sentences with regard to the distance to the antecedent. Furthermore, on the level of the content it was observed that *Es ist dies*-sentences often indicate a summarizing strategy and in the corpus sample they occur significantly more frequently than *Dies ist*-sentences in the corresponding content category ‘It is this/This is the n\textsuperscript{th} ...’. Both observations involve a discontinuity in discourse. Referring back to a long distance antecedent constitutes an unexpected discourse move and in the case of ‘It is this/This is the n\textsuperscript{th} ...’, there is a shift from a descriptive or elaborative passage to a subsuming comment on the content of this passage. Considering that the pre-field is normally a site for coherence creating expressions, it appears reasonable that those breaks in discourse continuity should be marked in this position to make the discourse structure more transparent.

Therefore, another constraint is introduced to the ranking: \textsc{markbreak-vf} ([20], see also Table 4).

\begin{equation}
\text{(20) MarkBreak-VF: Mark breaks in discourse}
\end{equation}

It has to be assumed that MarkBreak-VF slightly overlaps with \textsc{aboutnessTopic-vf}. As a result, MarkBreak-VF may outrank AboutnessTopic-VF. This accounts for the occurrence of *Es ist dies* when *dies* refers to an unexpected discourse referent, which could be a long distance antecedent.\textsuperscript{15} It also accounts for the occurrence of *Es ist dies* when it makes sense to mark a cesura in discourse in order to facilitate comprehension as in ‘It is this the n\textsuperscript{th} ...’-constructions. In all these cases, prefield-*es* marks the break in discourse continuity. Similarly, it has often been shown that shifts of topics tend to be marked (see Givón 1983; Bestgen and Vonk 2000; Breindl, 2008; 2011). Thus, the constraint represents a general principle of discourse organization. However, as desired, unless in the case of a reverse ranking, the new version of the ranking still predicts that *Dies ist* remains the optimal candidate, since MarkBreak-VF is ranked lower than AboutnessTopic-VF (see Table 4).

The new constraints also explain the standard occurrences of prefield-*es*, as characterized by Speyer. In Section 3, prefield-*es* was characterized to occur in sentences that lack an element that can be considered topic in terms of Centering Theory. For example, in (21) the subjects of the prefield-*es* sentences are neither of relevance in the preceding nor in the subsequent text.

\textsuperscript{15} The tendency – which was however not statistically significant (\( p = 0.12 \)) – that *Es ist dies*-sentences are used to refer to larger antecedents is also in line with this analysis, as it seems plausible to indicate a break when a larger portion of text is subsumed.
Table 4: Modified prefield ranking.

| Candidates                          | 1-VF | SCENESETTING-VF | ABOUTNESSTOPIC-VF | MARKBREAK-VF |
|-------------------------------------|------|-----------------|-------------------|--------------|
| pre Dies ist PRECED.Sent            |      |                 |                   |              |
| Es ist dies PRECED.Sent             |      |                 |                   | *            |
| post Dies ist NOT PRECED.Sent       |      |                 |                   | *            |
| Es ist dies NOT PRECED.Sent         |      |                 |                   | *            |
| Dies ist ‘This is the nth...’       |      |                 |                   | *            |
| Es ist dies ‘It is this the nth...’ |      |                 |                   | *            |

(21) (According to a study, 60 percent of the population in Germany are on a search for meaning.)

*Es fehlen Vorbilder, es wächst das Bedürfnis nach*

it lack role.models it grows the need for

Selbstverwirklichung.

self-fulfillment

‘Role models are lacking, the need for self-fulfillment is growing.’

(Families are falling apart, people are turning away from the churches. More often answers are searched for in esotericism.)

(St. Galler Tagblatt, 24.04.1997)

Such thetic sentences describe a general situation and do not feature a regular topic; instead the situation has to be considered the topic (Krifka 2007: 43). In these cases, prefield-es marks this lack of a constant topic and the shift to a description of a situation. This is in accordance with MARKBREAK-VF.

The modified ranking, moreover, accounts for the regional differences. It has to be assumed that in non-southern varieties of German, the constraints overlap only to a small extent, so that ABOUTNESSTOPIC-VF standardly outranks MARKBREAK-VF. This explains that Es ist dies-sentences almost never occur in non-southern dialects of German. In contrast, in southern varieties the two constraints overlap more extensively, which leads to a higher probability that MARKBREAK-VF outranks ABOUTNESSTOPIC-VF effecting that Es ist dies-sentences are used more frequently.

Still, for a large number of the corpus findings it is not yet possible to make out the reason for the use of es in the prefield. The comparison between the Es ist dies and the Dies ist sample yielded a further significant difference with regard to the grammatical function of the antecedent: Objects were more often attested for Es ist dies-sentences than for Dies ist-sentences. Indeed, objects are syntactically less prominent than subjects, but demonstrative pronouns are generally
used to refer to referents that are not in the focus of attention (Gundel et al. 2004; Hegarty et al. 2001; Hedberg et al. 2007; Bosch and Hinterwimmer 2016). The use of prefield-es could be an optional way of further highlighting that the accessibility of the discourse referent is relatively low. After all, combinations like *Es ist er + NP* or *Es ist sie + NP* where grammatical gender already limits the number of possible antecedents are not attested.

Nevertheless, further research concerning the usage conditions of *Es ist dies*-sentences is necessary. Apart from the need for more evidence for the postulation that *Es ist dies* marks a break in discourse, it should be investigated why *Es ist dies*-sentences refer to nominal antecedents more frequently than *Dies ist*-sentences, which more often feature sentential antecedents.

### 4.2 Rating study

This chapter presents an online rating study, which compared the acceptability of *Es ist dies*- and *Dies ist*-sentences between speakers of non-southern German dialects\(^\text{16}\) and speakers of southern German dialects.

#### 4.2.1 Introduction

The results of the corpus study suggest that *Es ist dies*-sentences are a phenomenon of southern varieties of German, since 88% of the instances of *Es ist dies*-sentences occurred in texts from Switzerland, Austria, and Bavaria. Thus, it can be expected that speakers from these regions consider the construction more acceptable than speakers of non-southern varieties of German. An online rating study tested this prediction by comparing acceptability judgments from the two groups of speakers. In addition, assuming that there are contextual features or other general features of *Es ist dies*-sentences which prompt their use, speakers of southern varieties of German can be expected to give higher ratings to original *Es ist dies*-sentences than to ones derived from *Dies ist*-sentences. To test this, both original and derived sentences were used in the experiment. A difference between the two types could provide insights into aspects of the usage conditions of *Es ist dies*-sentences, that have, so far, remained undetected.

The test items were short excerpts from newspaper articles taken from the results of the corpus search. The set of test items consisted of a) original *Es ist dies*-sentences as found in a corpus search, b) original *Dies ist*-sentences, c) *Es ist*

---
\(^{16}\) The label *non-southern* encompasses dialects from various regions of Germany including, for example, North Rhine-Westphalia as well as Saxony.
dies-sentences derived from original Dies ist-sentences and d) Dies ist-sentences derived from original Es ist dies-sentences. Each item occurred in two versions: 1) featuring an Es ist dies-sentence and 2) featuring a Dies ist-sentence. In (22), the Es ist dies-variant is in the original form as found in the corpus whereas Dies ist is the derived form; in (23) Dies ist is the original.

(22) In Barcelona wurde der Fernsehjournalist Fernando Delgado mit dem spanischen Literaturpreis „Planeta“ für sein Buch „Der Blick des Anderen“ (La mirada del otro) ausgezeichnet. Es ist dies/Dies ist der mit 4,96 Millionen Euro höchstdotierte Preis des Landes.

(In Barcelona, the TV journalist Fernando Delgado was awarded the Spanish literature award “Planeta” for his book “The gaze of the other” [La Mirada del otro]. Including prize money of 4.96 million Euros, this is the most highly remunerated award of the country.)

(23) In Israel könnte bald die Beschimpfung von Gegnern als „Nazis“ und die Verwendung von Holocaust-Symbolen bei Protesten per Gesetz verboten sein. Dies ist/Es ist dies eine unmittelbare Reaktion auf die jüngsten Demonstrationen von Ultra-Orthodoxen, bei denen Männer und sogar Kinder verkleidet als KZ-Häftlinge auf ihre vermeintliche Verfolgung durch den Staat aufmerksam machen wollten.

(In Israel, calling opponents “nazis”, and the use of holocaust symbol at protests could soon be forbidden by law. This is an immediate reaction to the latest protests by Ultra-Orthodox Jews, during which men and even children, dressed up as concentration camp prisoners, wanted to bring their alleged persecution by the state to attention.)

It is predicted that a) speakers of non-southern varieties of German judge Dies ist-sentences to be more acceptable than Es ist dies-sentences and b) that speakers of southern varieties of German judge Es ist dies-sentences to be more acceptable than speakers of non-southern varieties. Thus, an interaction of the factors PARTICIPANT’S DIALECT (southern vs. non-southern) and ITEM TYPE (Es ist dies vs. Dies ist) is predicted (hypothesis one). Since Es ist dies-sentences are generally marked, they are expected to receive lower ratings than Dies ist-sentences from both groups of participants. Thus, a main effect of ITEM TYPE is predicted, too.

Concerning the source of the test items, it can be expected that a) speakers of southern varieties of German give higher ratings to the original Es ist dies-sentences than to the derived ones. Moreover, it can be expected that b) the derived Dies ist-sentences receive higher ratings by those participants than the derived Es ist dies-sentences, as the former type, being the unmarked choice in
general, is assumed to be unaffected by the factor source. Accordingly, an interaction of the factors source (derived vs. original) and item type (Es ist dies vs. Dies ist) is predicted (hypothesis two). Again, in both conditions, Es ist dies-sentences should receive slightly lower ratings than Dies ist-sentences. Thus, a main effect of item type is expected, too.

4.2.2 Method

Participants
80 native speakers of southern varieties of German and 80 native speakers of non-southern varieties of German took part in the experiment. Prior to the experiment, the participants had to answer questions about their dialectal background: They were asked about their residence, since when they have lived there, about their hometown, and about the hometown of both parents. By collecting this information, it was checked whether a participant clearly fitted one of the two categories, speaker of southern German and speaker of non-southern German. Participants with a mixed dialectal background – e.g. a person who currently lives in northern Germany but grew up in Bavaria – were excluded prior to a further analysis of their ratings. The number of participants reported here refers to the number of participants that passed the dialect check. Table 5 shows the states/countries of origin for both groups of participants.

The group of speakers of southern German varieties consisted of 6 males and 74 females who were between 18 and 54 years old ($M = 27.84$ years, $SD = 9.14$). 32 of them had a higher education entrance qualification, 24 had completed a Bachelor’s degree, 23 had a Master’s degree or higher, and one participant stated “other” as their degree. The group of speakers of non-southern German varieties consisted of 22 males, and 58 females. Their age ranged between 18 and 68 years ($M = 26.74$ years, $SD = 9.78$). 49 of them had a higher education entrance qualification, 21 had a Bachelor’s degree, 5 had obtained a Master’s degree or higher, and 5 stated “other”. The participants were acquired via posts on university-related internet platforms or university-related mailing lists. They took part in the experiment on a voluntary basis but had the chance to win one of five 10€ Amazon gift cards, which were raffled off among the participants who wanted to take part in the lottery.

17 Moreover, participants could choose to answer the question whether they read a newspaper on a regular basis and if so, which. This was asked to check whether the speakers of non-southern variants of German read newspapers from regions where a southern dialect is spoken. If they do, they might have encountered more Es ist dies-sentences than participants who do not read newspapers from the south. As only three participants indicated that they regularly read a southern newspaper, the data concerning the newspaper reading habits were not considered further.
| State/country of origin | Number of participants |
|------------------------|------------------------|
| Austria                | 43                     |
| Switzerland            | 29                     |
| Bavaria                | 8                      |

| State of origin         | Number of participants |
|-------------------------|------------------------|
| Lower Saxony            | 43                     |
| North Rhine-Westphalia  | 14                     |
| Hesse                   | 5                      |
| Bremen                  | 4                      |
| Hamburg                 | 4                      |
| Schleswig-Holstein      | 3                      |
| Rhineland-Palatinate    | 2                      |
| Thuringia               | 2                      |
| Berlin                  | 1                      |
| Mecklenburg-Western Pomerania | 1 |
| Saxony                  | 1                      |

Table 5: Participants’ states/countries of origin.

Material
The test items were excerpts from newspaper articles taken from the corpus search for *Es ist dies* and *Dies ist*-sentences. As shown in (22) and (23), each test item consisted of a context of two to four lines followed by the target sentence, which was highlighted in bold face. There were 32 test items, 16 featured an *Es ist dies*-sentence and 16 featured a *Dies ist*-sentence. Using original excerpts from newspapers makes it difficult to yield a uniform set of items. Therefore, some parameters were set to ensure uniformity at least to some extent. Both sentence types occurred in two versions: First, there were sentences, like in (22), with a definite noun phrase as the complement of the copula. *Dies* in these sentences referred back to the referent of a nominal antecedent and the sentence contained either a superlative or was an ‘It is this the n\(^{th}\)…’-construction. The complement of the second type was an indefinite NP and *dies* referred back to the content of the preceding sentence as in (23). These two versions were chosen, first, because they exemplify typical instances of *Es ist dies*- and *Dies ist*-sentences according to the corpus study and, second, because in those cases the length and amount of the

---

18 Further experimental material is attached as Appendix, available in the online version of this article.
required context sentences is appropriate for an experiment. In contrast, larger sentential antecedents or long distance antecedents would have disturbed the uniformity of the item set. From each of these original items a counterpart was derived by either changing *Es ist dies* to *Dies ist* or vice versa. In addition, minor changes were made to the items in order to avoid confusion for the participants due to outdated orthography, unusual names or the like. The items were allocated to four experimental lists so that each list featured eight items, one item of each type. In addition to the test items, each list contained 20 fillers. The fillers consisted also of a context of two to four lines followed by a target sentence in bold face. There were five categories of fillers: *weil*-verb-second sentences, cleft sentences, scrambling sentences, i.e. sentences with non-default word order, well-formed sentences, and ungrammatical sentences with violations of word order or incongruences. The latter two categories served as controls. Apart from the cleft sentences, which were taken from the results of a corpus search for clefts, all fillers were also excerpts from newspaper articles taken from the same sources as the test items. They were modified to fit their category, i.e. the word order was changed, incongruences were added, etc.

**Procedure**

Participants took part via a web-experiment created with the software OnExp 1.3.1 (Onea and Syring 2011). They were informed that they would be presented with newspaper excerpts and instructed to first read the excerpt and to then rate the sentence in bold print. In addition, the participants were instructed to rely on their linguistic intuition when making their judgments and not on prescriptive grammar rules. It was pointed out that the judgments should not be made based on political views, opinions on certain historical events or an estimation of the truth of the content. This was thought necessary, as some of the newspaper excerpts covered controversial topics like holocaust, war, or nationalism. Moreover, participants were asked to take the task seriously and to avoid distraction. As shown in Figure 10, there was a 7-point Likert scale below each item. Only the endpoints were labeled as “very bad” and “very good”. The order of the items was randomized. Before the actual experiment started, participants had to rate three trial excerpts containing a well-formed sentence, a *weil*-verb-second sentence and a scrambling sentence in order to get accustomed to the task.

**Factorial design and predictions**

With regard to the first hypothesis, the experimental design was 2 (participant’s dialect; southern vs. non-southern) × 2 (item type; *Es ist dies* vs. *Dies ist*). The
factor PARTICIPANT’S DIALECT was tested between participants, but within items, and the factor ITEM TYPE was tested both within participants and items. It was predicted that there would be an interaction of the two factors: Dies ist-sentences should receive higher ratings than Es ist dies-sentences among speakers of non-southern dialects. Moreover, Es ist dies-sentences should be more acceptable among speakers of southern dialects than among speakers of non-southern dialects. Also, a main effect of ITEM TYPE was predicted: Es ist dies-sentences should generally be rated lower.

With regard to the second hypothesis, which only concerned speakers of southern varieties of German, the design was 2 (SOURCE; original vs. derived) × 2 (ITEM TYPE; Es ist dies vs. Dies ist). Both factors were tested within participants and within items. An interaction between the factors was predicted: Es ist dies-sentences should be rated higher when they are presented in the original version than when appearing in the derived version. Also, derived Dies ist-sentences should be more acceptable than derived Es ist dies-sentences. A main effect of ITEM TYPE was predicted, too: Also among speakers of southern German, the Es ist dies type was expected to be rated lower than the Dies ist type.

4.2.3 Results

Controls
Based on their ratings for the controls, the data of 13 speakers of southern variants of German and six speakers of non-southern variants of German were ex-
cluded from further analysis. Participants were excluded if two of their ratings were deviant. With regard to the well-formed controls, a rating lower than three was counted as deviant and with regard to the ungrammatical controls,\textsuperscript{19} ratings higher than three were considered deviant. In total, the data of 141 participants (62 female and 5 male speakers of southern varieties of German and 54 female and 20 male speakers of non-southern varieties of German) were subjected to statistical analysis.

**Descriptive statistics**

The mean ratings of *Es ist dies*– and *Dies ist*–sentences by PARTICIPANT’S DIALECT are reported in Table 6 and graphically represented in Figure 11. They

| PARTICIPANT’S DIALECT | ITEM TYPE | *
|-----------------------|-----------|
|                       | *Dies ist*| *Es ist dies* |
| southern              | 5.85 (1.36) | 4.18 (1.84) |
| non-southern          | 5.82 (1.35) | 2.79 (1.79) |

\textbf{Table 6:} Means: item type, participant’s dialect (standard deviation).

\textbf{Figure 11:} Mean rating: item type, dialect.

\textsuperscript{19} Only three of the original four ungrammatical fillers were considered controls, since one of them received unexpectedly high rating scores. The target sentence displayed below was ungrammatical because of two incongruences, which were apparently easy to miss:

*Die von einem Industriellen gespendeten Luxuskarossen sollte zur Finanzierung eines Altenheimes in Arles beitragen, das von Mitarbeiterinnen der Ordensschwester geführt werden.*
display the expected pattern: There is a difference of 3.03 between the mean rating of the Dies ist- and the mean rating of the Es ist dies- sentences in the group of speakers of non-southern varieties of German, the Dies ist-sentences being rated higher. Furthermore, there is a difference of 1.39 with regard to the Es ist dies-sentences between mean ratings of the two groups of speakers; the mean of the Es ist dies type is higher among the speakers of southern dialects. Also, Es ist dies-sentences generally received lower ratings than the Dies ist-sentences.

Table 7 shows the mean ratings of Es ist dies- and Dies ist-sentences by the two manifestations of the factor source for the subset of southern German speakers. These results are graphically presented in Figure 12. As predicted, the means of the Es ist dies-sentences were lower than those of the Dies ist-sentences. However, there is no difference in the means between the Es ist dies-sentences in the derived and in the original condition. Furthermore, it can be observed that with a difference of 0.22, Dies ist-sentences receive slightly higher ratings in the original than in the derived version, which was not predicted.

| SOURCE  | ITEM TYPE | Dies ist | Es ist dies |
|---------|-----------|---------|-------------|
|         | Dies ist  | 5.96 (1.19) | 4.18 (1.77) |
| original| Es ist dies | 5.74 (1.51) | 4.18 (1.91) |

**Table 7:** Means: item type, source (standard deviation).

**Figure 12:** Mean rating: ITEM TYPE, SOURCE.
Inferential statistics
For the inferential statistical analysis, the software R (R Core Team 2017) and the packages lme4 (Bates et al. 2015) and lmerTest (Kuznetsova et al. 2017) were used. Pertaining to the two hypotheses, two linear mixed effects analyses and t-tests using Satterthwaite’s method were performed. Regarding the first hypothesis, a model containing PARTICIPANT’S DIALECT and ITEM TYPE as fixed effects was created. As random effects, intercepts for participants and items were added. The results are displayed in Tables 8 and 9.

Table 8: Random effects (hypothesis 1).

| RANDOM EFFECTS | VARIANCE | STD. DEV. |
|----------------|----------|-----------|
| participant    | 0.5      | 0.71      |
| item           | 0.17     | 0.41      |
| residual       | 1.9      | 1.4       |

Table 9: Fixed effects (hypothesis 1).

| FIXED EFFECTS                              | ESTIMATE | STD. ERROR | T VALUE | P-VALUE |
|--------------------------------------------|----------|------------|---------|---------|
| intercept                                  | 5.82     | 0.15       | 37.82   | <0.001  |
| item type (Es ist dies)                    | −3.02    | 0.18       | −16.41  | <0.001  |
| dialect (southern)                         | 0.04     | 0.17       | 0.25    | 0.81    |
| item type (Es ist dies) * dialect (southern)| 1.33     | 0.16       | 8.10    | <0.001  |

As displayed in the table, the significance of the interaction between ITEM TYPE and PARTICIPANT’S DIALECT was confirmed. As predicted, a significant effect of ITEM TYPE was found, too. There was no significant effect of PARTICIPANT’S DIALECT.

Pertaining to the second hypothesis, a model containing ITEM TYPE and SOURCE as fixed effects was created. As random effects, intercepts for participants and items were added. The results are shown in Table 10 and Table 11. Table 11 shows that the predicted effect of item type is significant. However, there is no significant interaction between the factors ITEM TYPE and SOURCE and there is no significant effect of the factor SOURCE.

20 The model was as follows: lmer(rating ~ item type * participant’s dialect + (1 | participant) + (1|item).
21 The model was as follows: lmer(rating ~ item type * source + (1 | participant) + (1|item).
Table 10: Random effects (hypothesis 2).

| RANDOM EFFECTS | VARIANCE | STD. DEV. |
|----------------|----------|-----------|
| participant    | 0.59     | 0.77      |
| item           | 0.18     | 0.43      |
| residual       | 1.87     | 1.37      |

Table 11: Fixed effects (hypothesis 2).

| FIXED EFFECTS                                | ESTIMATE | STD. ERROR | T VALUE | P-VALUE |
|----------------------------------------------|----------|------------|---------|---------|
| intercept                                    | 5.74     | 0.22       | 26.85   | <0.001  |
| item type (Es ist dies)                      | −1.58    | 0.27       | −5.82   | <0.001  |
| source (original)                            | 0.23     | 0.27       | 0.83    | 0.41    |
| item type (Es ist dies) * source (original)  | −0.21    | 0.38       | −0.54   | 0.6     |

### 4.2.4 Discussion

The first hypothesis, repeated for convenience, was that Dies ist-sentences would receive higher ratings than Es ist dies-sentences among speakers of non-southern dialects. Moreover, Es ist dies-sentences were predicted to be more acceptable among speakers of southern dialects than among speakers of non-southern dialects. In addition, Es ist dies-sentences were predicted to be generally rated lower. The fact that the interaction between participant’s dialect and item type was significant makes it safe to conclude that Es ist dies-sentences are definitely a phenomenon of southern varieties of German. Nevertheless, even speakers of southern German dialects rated Dies ist-sentences generally higher than Es ist dies-sentences and this effect of item type was significant, too. This result is in conformity with the modified prefield ranking, according to which the Dies ist-type constitutes the optimal candidate and the Es ist dies-type only occurs in case of a reverse ranking.

The second hypothesis concerned only speakers of southern dialects. It was predicted that original Es ist dies-sentences would be rated higher than derived ones. However, it did not make a difference for speakers of southern variants of German whether an Es ist dies-sentence was an original Es ist dies-sentence or derived from a Dies ist-sentence. Thus, with the set of items used in this experiment it was not possible to discern a factor concerning the textual context or the features of the Es ist dies-sentence that prompts the use of this construction among speakers of southern varieties of German.
Summary and outlook

The starting point of this paper was the observation that a so far unnoticed use of prefield-"es" exists. This occurrence of prefield-"es" cannot be accounted for by Speyer’s (2008, 2009) optimality theoretic prefield ranking, which predicts that prefield-"es" should only occur if there is no better candidate available. A corpus study that compared a sample of *Es ist dies*-sentences to *Dies ist*-sentences revealed significant differences between the two samples concerning the feature “distance to the antecedent”. Furthermore, it was found that *Es ist dies*-sentences featured significantly more often ‘It is this the n-th ...’. This construction indicates a summarizing strategy, which was interpreted as a reason to mark a break in discourse. Crucially, the corpus study indicated that *Es ist dies*-sentences are a phenomenon of southern varieties of German. Based on the findings of the corpus study, a modified version of the prefield ranking was proposed, which replaces the constraint Topic-VF with the constraints AboutnessTopic-VF and MarkBreak-VF. This new version of the ranking accounts for the occurrence of *Es ist dies*-sentences, the regional differences, as well as for some differences between *Es ist dies*- and *Dies ist*-sentences. An open question that remains is how to account for the use of *Es ist dies*-sentences in cases in which so far no reason to mark a break in discourse became apparent. In addition, the finding that the *Es ist dies*-sentences in the sample more often have nominal antecedents than the *Dies ist*-sentences, which more frequently feature sentential antecedents, also lacks an explanation.

The subsequent rating study provided clear evidence for the regional character of the construction. It was shown that speakers of southern German varieties find *Es ist dies*-sentence more acceptable than speakers from non-southern regions. However, the study did not provide further insights into the conditions that promote the use of this construction. More research is necessary in this area.

Another issue that has not been addressed yet is whether *Es ist dies*-sentences are used exclusively in written language or if they also occur in spoken language. This should be further investigated. Future research should also focus on the usage conditions of *Es ist dies*-sentences. By including factors like different types of antecedents in an experimental design, their influence on the acceptability of *Es ist dies*-sentences could be tested in a straightforward way.

Acknowledgment: I wish to thank Edgar Onea, Alexander Schreiber, Swantje Tönnis, Thomas Weskott and the audience at “Grammar and Corpora 2016” for their valuable suggestions and helpful discussion as well as Maya Cortez Espinoza and Patrick Stucke for proofreading. Moreover, I am grateful for the constructive
criticism and helpful comments I received from the editor and three anonymous reviewers. All remaining mistakes are of course my own.

References

Bates, Douglas, Martin Maechler, Ben Bolker and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

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. 2011. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32(1). 45–86.

Bosch, Peter and Stefan Hinterwimmer. 2016. Anaphoric reference by demonstrative pronouns: In search of the relevant parameters. In Anke Holler and Katja Suckow (eds.), *Empirical perspectives on anaphora resolution*, 193–212. Berlin & Boston: de Gruyter.

Bosch, Peter and Carla Umbach. 2007. Reference determination for demonstrative pronouns. *ZAS Papers in Linguistics* 48. 39–51.

Bubenhofer, Noah, Marek Konopka and Roman Schneider. 2014. *Präliminarien einer Korpusgrammatik*. Tübingen: Narr.

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.

Diessel, Holger. 1999. *Demonstratives: Form, function and grammaticalization*. Amsterdam: Benjamins.

Filippova, Katja and Michael Strube. 2007. The German Vorfeld and local coherence. *Journal of Logic, Language, and Information* 16(4). 465–485.

Fricke, Lea M. and Swantje Tönnis. 2018. Es ist dies – A special use of German prefield-es. In Eric Fuß, Marek Konopka, Beata Trawiński and Ulrich H. Waßner (eds.), *Grammar and Corpora 2016*, 221–230. Heidelberg: Heidelberg University Publishing.

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: Benjamins.

Graefen, Gabrielle. 1997. *Der wissenschaftliche Artikel: Textart und Textorganisation*. Frankfurt a. M.: Lang.

Grosz, Barbara J., Aravind K. Joshi and Scott Weinstein. 1995. Centering: A framework for modeling local coherence of discourse. *Computational Linguistics* 21. 203–225.

Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski. 2004. Demonstrative pronouns in natural discourse. In *Proceedings of the Fifth Discourse Anaphora and Anaphora Resolution Colloquium (DAARC-2004)*, 81–86.

Hedberg, Nancy, Jeanette K. Gundel and Ron Zacharski. 2007. Directly and indirectly anaphoric demonstrative and personal pronouns in newspaper articles. In *Proceedings of the Sixth Discourse Anaphora and Anaphora Resolution Colloquium (DAARC-2007)*, 31–36.
Hegarty, Michael, Jeanette K. Gundel and Kaja Borthen. 2001. Information structure and the accessibility of clausally introduced referents. *Theoretical Linguistics* 27. 163–186.

Hirschberg, Julia Bell. 1985. *A theory of scalar implicatures*. Philadelphia, PA: University of Pennsylvania dissertation.

Koch, Peter and Wulf Oesterreicher. 1994. Schriftlichkeit und Sprache. In Hartmut Günther and Otto Ludwig (eds.), *Writing and its use: An interdisciplinary handbook of international research*, Vol. 1, 587–604. Berlin & New York: de Gruyter.

Koeppel, Rolf. 1993. *Satzbezogene Verweisformen: Eine datenbankgestützte Untersuchung zu ihrer Distribution und Funktion in mündlichen Texten, schriftlichen Texten und schriftlichen Fachtexten des Deutschen*. Tübingen: Narr.

Krifka, Manfred. 2007. Basic notions of information structure. In Caroline Féry, Gisbert Fanselow and Manfred Krifka (eds.), *The notion of information structure* (Interdisciplinary Studies on Information Structure, Vol. 6), 13–56. Potsdam: Universitätsverlag.

Kuznetsova, Alexandra, Per B. Brockhoff and Rune H. B. Christensen. 2017. *lmerTest package: Tests in linear mixed effects models*. *Journal of Statistical Software* 82(13). DOI: 10.18637/jss.v082.i13.

Moser, Hugo. 1971. Typen sprachlicher Ökonomie im heutigen Deutsch. In Hugo Moser (ed.), *Sprache und Gesellschaft: Beiträge zur soziallinguistischen Beschreibung der deutschen Gegenwartssprache*, 89–117. Düsseldorf: Schwann.

Onea, Edgar and Alexander Syring. 2011. *OnExp 1.3.1*. https://onexp.textstrukturen.uni-goettingen.de/.

Paranhos Zitterbart, Jussara. 2002. Zur Mittelfeldfähigkeit des Korrelat es in Verbindung mit Subjektstätzen. *Sprachwissenschaft* 27. 149–195.

Pittner, Karin and Judith Berman. 2004. *Deutsche Syntax: Ein Arbeitsbuch*. Tübingen: Narr.

Prince, Alan and Paul Smolensky. 2004 [1993]. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.

Prince, Ellen F. 1999. How to not mark topics: Topicalization in English and Yiddish. In *Texas Linguistics Forum*, 41. Austin, TX: University of Texas Press.

R Core Team 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27(1). 53–94.

Speyer, Augustin. 2007. Die Bedeutung der Centering Theory für Fragen der Vorfeldbesetzung im Deutschen. *Zeitschrift für Sprachwissenschaft* 26(1). 83–115.

Speyer, Augustin. 2008. German Vorfeld-filling as constraint interaction. In Anton Benz and Peter Kühnlein (eds.), *Constraints in discourse*, 267–290. Amsterdam: Benjamins.

Speyer, Augustin. 2009. Das Vorfeldranking und das Vorfeld-es. *Linguistische Berichte* 219. 323–353.

Sternfeld, Wolfgang. 2008. *Syntax. Eine morphologisch motivierte generative Beschreibung des Deutschen, Band 1. 3.*, überarb. Aufl. Tübingen: Stauffenberg.

Tomaselli, Alessandra. 1986. Das unpersönliche „es“ – Eine Analyse im Rahmen der Generativen Grammatik. *Linguistische Berichte* 102. 171–190.

Walker, Marilyn, Aravind Joshi and Ellen Prince. 1998. Centering in naturally occurring discourse: An overview. In Marilyn A. Walker, Aravind K. Joshi and Ellen F. Prince (eds.), *Centering theory in discourse*, 1–28. Oxford: Oxford University Press.
Webber, Bonnie L. 1991. Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes* 6(2). 107–135.

Zifonun, Gisela. 1995. Minimalia Grammaticalia: Das nicht-phorische *es* als Prüfstein grammatischer Theoriebildung, *Deutsche Sprache* 23. 39–60.

**Supplemental Material:** The online version of this article offers supplementary material (https://doi.org/10.1515/zfs-2020-2012).