Discours
Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics

18 | 2016
Varia

Accessibility and Referential Choice: Personal Pronouns and D-pronouns in Written German

Yvonne Portele and Markus Bader

Electronic version
URL: http://journals.openedition.org/discours/9188
DOI: 10.4000/discours.9188
ISSN: 1963-1723

Publisher:
Laboratoire LATTICE, Presses universitaires de Caen

Electronic reference
Yvonne Portele and Markus Bader, « Accessibility and Referential Choice: Personal Pronouns and D-pronouns in Written German », *Discours* [Online], 18 | 2016, Online since 16 September 2016, connection on 19 April 2019. URL : http://journals.openedition.org/discours/9188 ; DOI : 10.4000/discours.9188

discours
Revue de linguistique, psycholinguistique et informatique
http://discours.revues.org/

Accessibility and Referential Choice: Personal Pronouns and D-pronouns in Written German

Yvonne Portele
Goethe-University Frankfurt

Markus Bader
Goethe-University Frankfurt

Yvonne Portele, Markus Bader, « Accessibility and Referential Choice: Personal Pronouns and D-pronouns in Written German », *Discours* [En ligne], 18 | 2016, mis en ligne le 16 septembre 2016.

URL : http://discours.revues.org/9188

Titre du numéro : *Varia*
Coordination : Saveria Colonna & Sarah Schimke

Accessibility and Referential Choice: Personal Pronouns and D-pronouns in Written German [1]

Yvonne Portele

Goethe-University Frankfurt

Markus Bader

Goethe-University Frankfurt

We present a corpus study and a production experiment that investigated the choice between two types of pronouns in written German – personal pronouns and so-called d-pronouns, which have properties of both personal and demonstrative pronouns. The existing research concerned with these pronouns has focused on language comprehension, in particular with regard to interpretative preferences in the case of referential ambiguity. In contrast to language comprehension, a choice between the two pronominal forms has to be made during language production whether there is an ambiguity or not. The corpus data show that the choice between p(ersonal) pronoun and d-pronoun depends on several factors that have been claimed to determine a referent's accessibility (Ariel, 1990), in particular givenness and syntactic prominence (syntactic function and clausal position). The corpus study is supplemented by a production experiment that required participants to continue a short text passage with a sentence starting with either a p-pronoun or a d-pronoun. The results of the experiment strengthen the conclusion that several factors determine accessibility which in turn governs pronoun choice. We finally discuss several factors besides accessibility that affect the choice of either a p- or a d-pronoun.

Keywords: anaphora, referential choice, language production, corpus linguistics, German, accessibility

Nous présentons une étude de corpus et une expérience de production qui examinent le choix entre deux types de pronoms en allemand écrit, à savoir les pronoms personnels et les « d-pronoms ». Ceux-ci ont non seulement des propriétés de pronoms personnels, mais également des propriétés de pronoms démonstratifs. La recherche déjà existante concernant ces pronoms s'est concentrée sur la compréhension du langage. Cela a été fait en particulier en prenant en considération les préférences interprétatives à travers la notion d'ambiguïté référentielle. Au-delà de la compréhension du langage, il faut faire un choix entre les deux formes pronominales pendant la production du langage, s'il y a une ambiguïté ou non. Les données de corpus montrent que le choix entre « p-pronom » et « d-pronom » dépend de plusieurs facteurs. Ces derniers ont pour but de déterminer l'accessibilité d'un référent (Ariel, 1990), en particulier le « givenness *» et l'importance syntaxique (fonction syntaxique et position propositionnelle). L'étude du corpus est complétée par une expérience de production qui a demandé aux participants de continuer un passage de texte court avec une phrase qui commence soit avec un « p-pronom », soit avec un « d-pronom ». Les*

1. For helpful discussions, we would like to thank Emilia Ellsiepen and Alice Glöckner. We also would like to thank the two reviewers, who provided us with many intriguing comments.

résultats de l'expérience renforcent la conclusion que plusieurs facteurs déterminent l'accessibilité, qui à son tour détermine le choix du pronom. Finalement, nous discutons de facteurs, autres que l'accessibilité, qui affectent le choix du pronom.

Mots clés : *anaphorique, choix référentiel, production du langage, linguistique des corpus, allemand, accessibilité*

1. Introduction

1 Much of language production is concerned with referring back to entities that were introduced at some earlier point in the ongoing discourse. Languages provide a wide variety of referential expressions for this purpose. These expressions can be ordered on a scale of explicitness, ranging from fully reduced pronouns (null pronouns) to full lexical noun phrases (NPs). A simplified version of this referential scale is shown in [1] (see Ariel [2001] for the complete scale).

[1] The referential form scale
null pronoun > pronoun > demonstrative > full noun phrase

2 A large body of cross-linguistic research (for an overview, see Arnold, 2010) has shown that the position of an expression on the referential form scale correlates with discourse properties of its antecedent. As a first approximation, this correlation can be stated as follows: reduced referential forms refer to entities that are highly salient at the current point of the discourse whereas more explicit devices are used when referring to entities that are currently not salient. Intuitively, such a correlation makes sense. After all, when a referent is salient at the current point of discourse, it will be in a highly activated state in working memory and a reduced or even null expression will suffice to refer to this referent. The other way round, a referent that has not been mentioned recently will not be in an active state and more explicit means will be needed to refer to such a referent.

3 Several theoretical approaches exist that are built on the idea of a close relationship between referential form and memory states (Givón, 1992; Gundel et al., 1993; Ariel, 1990; Grosz et al., 1995). In the following, we will concentrate on accessibility theory (Ariel, 1990 and 2001), as this is the most comprehensive theory. The major thrust of accessibility theory has been stated succinctly by Ariel (2001: 29):

> Accessibility theory offers a procedural analysis of referring expressions, as marking varying degrees of mental accessibility. The basic idea is that referring expressions instruct the addressee to retrieve a certain piece of given information from his memory by indicating to him how accessible this piece of information is to him at the current stage of the discourse.

4 Arnold (2010) lists three properties of a referent that contribute to its accessibility: recency, givenness, and syntactic prominence. *Recency* refers to the number of sentences that have been produced since the last mention of the referent. Referents that have been mentioned recently are more accessible than referents that have

not been mentioned recently. *Givenness* captures whether a referent was already mentioned before or not. Given referents are more accessible than new referents. *Syntactic prominence* concerns sentence internal properties of an antecedent. Two different properties can increase a referent's syntactic prominence and thereby make it more accessible. First, a subject is more prominent than an object. Second, a sentence initial NP is more prominent than a sentence final NP.

5 Another line of research has stressed the importance of world knowledge and coherence relations for the interpretation of pronouns (see Kehler & Rohde, 2013, for a review of this research tradition). Kehler and Rohde (2013) propose a Bayesian model of pronoun interpretation that combines the accessibility and the coherence account in a probabilistic way. In line with other researchers (e.g., Fukumura & Van Gompel, 2010), Kehler and Rohde (2013) review evidence suggesting that in contrast to pronoun interpretation, the choice of a referential form depends on accessibility alone. Since we are concerned with production in this paper, we concentrate on accessibility, but we come back to the issue of coherence relations when presenting our experimental results.

6 Of special interest for the notion of accessibility are referential expressions that differ in form but not in lexical content. In this paper, we focus on the difference between two types of pronouns in German – personal pronouns like *er* ("he") and so-called d-pronouns like *der* (lit. "the"). D-pronouns are often analyzed as a kind of demonstrative pronoun. Alternatively, they have been claimed to be a variant of the personal pronoun (for further discussion, see Ahrenholz, 2007) or definite determiner phrases with an empty NP (Wiltschko, 1998).

7 This paper presents a corpus study and an experiment that have investigated the factors that determine the choice between personal pronoun (p-pronoun for short) and d-pronoun during written language production. The organization of this paper is as follows. The next section gives an overview of prior research concerned with p- and d-pronouns in German. The corpus study is presented in Section 3. Section 4 presents the production experiment. Section 5 will discuss the results presented in Sections 3 and 4 with regard to factors going beyond accessibility. The paper concludes with a final discussion in Section 6.

2. P-pronouns and d-pronouns in German

8 A prototypical example for the use of personal pronouns is given in [2]. In this example, the pronoun's antecedent is highly accessible in the sense discussed above: it is already given in the discourse, it occurs in sentence-initial position, and it is a subject[2].

2. The example texts in [2] and [3] are from the "DeWaC" corpus as described below. All corpus examples will be cited in the following way: we first give the number of the "DeWaC" file, then the number of the text within the "DeWaC" file, and finally the URL provided in the "DeWaC" file.

[2] [C -2] Gerade als sich der Schatten eines Baumes mit großen, breiten Blättern verflüchtigte, tauchte aus diesem **Senragor Allagan** auf und grinste seinen Halbonkel unheimlich an. [C -1] **Der Junge** hatte schwarzes, wirres, schulterlanges Haar, eine bleiche Haut und dunkle, aber durchdringende Augen. [T] **Er**[p-pronoun] war gerade erst neun, aber trotzdem schon erstaunlich groß und weit entwickelt für sein Alter.

'[C -2] Just when the shadow of a tree with large, wide leaves vanished, **Senragor Allagan** appeared and grinned scarily at his half-uncle. [C -1] **The boy** had black, fuzzy, shoulder-length hair, a pale skin and dark but shrewd eyes. [T] **He**[p-pronoun] was just nine, but surprisingly large and well developed for his age.'

(corpus = "DeWaC-9" text = "788291" id = "http://www.drachental.de")

9 A prototypical example for the use of d-pronouns is given in [3]. In this case, the pronoun's antecedent is discourse new, occurs sentence finally, and is an object.

[3] [C -3] Die Deutsche Bahn AG lädt zur Feierstunde. [C -2] Die AG wird zehn Jahre alt, und für die Gratulationstour ist der Ballsaal des Ritz-Carlton gerade gut genug. [C -1] Als Festredner hat Bahnchef Hartmut Mehdorn **den ihm freundschaftlich verbundenen Bundeskanzler** engagiert. [T] **Der** wird, da muss man kein Prophet sein, die Weisheit des Gesetzgebers rühmen, der zum 1. Januar 1994 die Behörden-Bahn abschaffte und das Wirtschaftsunternehmen Bahn schuf.

'[C -3] The Deutsche Bahn AG (German Railway Company) is inviting to celebrate. [C -2] The company will be 10 years old and the ball room of the Ritz-Carlton is just sufficient for congratulations. [C -1] The head of the company Hartmut Mehdorn employed **the Federal Chancellor that he is on cordial terms with**. [T] **He**[d-pronoun] will – you do not have to be a prophet – praise the wisdom of the legislature, who abolished the state-run enterprise and created the business corporation Bahn on 1st January 1994.'

(corpus = "DeWaC-8" text = "678968" id = "http://www.zeit.de")

10 With regard to lexical content, the p-pronoun *er* in [2] and the d-pronoun *der* in [3] do not differ from each other. Both are specified for the features "masculine" and "singular"[3]. In the linguistic literature (Abraham, 2002; Wiemer, 1996; Zifonun et al., 1997), several of the properties discussed above – givenness, syntactic function and linear position – have been considered as candidates for differentiating between p- and d-pronouns. A certain consensus emerging from this literature is that the main functions of p- and d-pronouns must be stated in information-structural terms. Broadly speaking, p-pronouns serve the function of topic continuation whereas d-pronouns signal a topic shift. More recently, the interpretation of p- and

3. Note that these features constraint the potential antecedent expressions and not the referent as such. Since gender is a grammatical category in German, masculine NPs, including pronouns, can refer to male persons but also to things.

d-pronouns has been the subject of several experimental investigations, all concerned with the question of how pronouns are interpreted when the context contains more than a single potential antecedent (Bosch & Umbach, 2007; Bouma & Hopp, 2007; Colonna et al., 2012; Ellert, 2013). As the overview in Ellert (2013: 6) shows, most studies have found that a p-pronoun preferentially takes a subject NP as antecedent, independently of its discourse status or clausal position, whereas the preferred antecedent of a d-pronoun is a discourse-new NP in clause final position, independently of the NP's syntactic function. For purposes of illustration, consider the following examples from Ellert (2013).

[4] a. Der Schrank ist schwerer als der Tisch. / Schwerer als der Tisch ist der Schrank.
The cupboard is heavier than the table. / Heavier than the table is the cupboard.

b. Er/Der stammt aus einem Möbelgeschäft in Belgien.
{P-pronoun/D-pronoun} comes from a furniture store in Belgium.

'The cupboard is heavier than the table. It comes from a furniture store in Belgium.'

11 The context sentence in [4a] is either a subject-initial or a subject-final sentence. The context sentence is followed by the target sentence [4b], which starts either with a p-pronoun or a d-pronoun. In two experiments, Ellert (2013) measured eye-movements while participants simultaneously listened to sentence pairs as in [4] and looked at pictures of the two referents mentioned in the context sentence. The results show a preference for the subject referent when hearing the p-pronoun *er*. When hearing the d-pronoun *der*, in contrast, a preference for the clause-final referent was observed.

12 Due to the lack of a preceding context, the discourse status of the two NPs in [4] was not explicitly specified. Ellert (2013) nevertheless interprets her results in terms of information structure and not in terms of surface properties like syntactic function or linear position within the sentence. This interpretation hinges on default associations between syntactic structure and information structure according to which the sentence topic is preferentially realized by the clause-initial NP whereas the sentence focus typically occurs clause finally. Given these default associations, the results can be rephrased as follows. For canonical subject-initial context sentences, the p-pronoun preferentially refers to the topic and the d-pronoun to the focus. For non-canonical subject-final sentences, in contrast, both p- and d-pronoun prefer the focus NP as antecedent.

13 Based on a broad survey of the existing evidence, Bosch (2013: 42) arrives at the generalization in [5], where "DPro" stands for d-pronoun and "PPro" for p-pronoun (see also Hinterwimmer, 2014).

[5] a. In contexts that provide only one grammatically suitable referent for the pronoun, DPro and PPro occur in free variation, and without any semantic difference.

b. Whenever a DPro must choose among several grammatically suitable referents, it avoids the current topic.

14 This generalization seems to capture the interpretation of d-pronouns quite accurately, but some problems remain. Most experiments on German are similar to the experiments of Ellert (2013) in that they do not provide enough context to unambiguously determine the discourse status of the potential antecedents provided in the context sentence. It is therefore necessary to assume that participants compute a particular information structure for the context sentences by default. Even if this assumption were correct, the interpretation of the observed preferences in terms of information structure is by no means obligatory. As information structure and linear position are confounded, the preferred interpretation of the d-pronoun could as well be stated in linear terms – the d-pronoun preferentially refers to the clause-final NP.

15 A study that provided enough context for controlling the discourse status of the potential antecedent NPs is the study of Finnish p- and d-pronouns by Kaiser and Trueswell (2008). However, even in this study information structure and linear position were confounded. Kaiser and Trueswell (2008) obtained the same pattern that was found for German – a subject preference for the Finnish p-pronoun and a preference for the final, new NP for the Finnish d-pronoun. Since given NPs always occurred sentence-initially and new NPs sentence-finally in their experiment, it is not possible to decide whether the preference observed for the d-pronoun is an effect of givenness or an effect of position.

16 In order to provide new evidence on this issue, Bader and Portele (2015) closely followed the experimental design of Kaiser and Trueswell (2008), but went beyond this study by also varying the linear position of the given and the new referent. In the configuration used by all prior studies (clause-initial given referent, clause-final new referent), Bader and Portele (2015) found again that p-pronouns prefer subjects as antecedents whereas d-pronouns preferentially refer back to referents that are new and occur in clause-final position. In order to disentangle givenness and position, Bader and Portele (2015) investigated short texts as in [6]. Here, the discourse-given NP *den Clown* ("the clown") occurs clause-finally whereas the discourse-new NP *ein Mann* ("a man") occurs clause-initially.

[6] Maria war am Sonntag im Zirkus. Vor der Aufführung sah sie schon **einen Clown** herumlaufen. Ein Mann umarmte **den Clown**. **Er** hat.../**Der** hat...

'Maria visited a circus on Sunday. Before the show, she saw a clown walking around. A man hugged the clown. He (p-pronoun/d-pronoun) has...'

17 Sentence fragments starting with a p-pronoun showed again a subject preference. Sentence fragments with a d-pronoun were most of the time completed in such a way that the d-pronoun was co-referential with the NP *den Clown*. Since the referent of this NP has already been introduced in the sentence before by the indefinite NP *einen Clown*, this means that the d-pronoun prefers a given NP as antecedent. Under the assumption that the referent of the given NP *den Clown* ("the clown") in the final context sentence in [6] is the sentence topic, this

constitutes an exception to the hypothesis that a d-pronoun preferentially refers to a non-topic. We will not discuss this issue at this point, but come back to it in the general discussion.

18 To sum up so far, experimental investigations of the interpretation of p- and d-pronouns converge on two conclusions. First, p-pronouns preferentially refer to an antecedent in subject position, independently of the givenness and the clausal position of the subject. Second, by itself, neither givenness, nor syntactic function, nor clausal position can account for all of the interpretative preferences found for d-pronouns. From the standpoint of accessibility theory, this is no surprise because a central assumption of this theory is that accessibility is a complex property that cannot be defined in terms of a single feature. In accordance with this assumption, Bader and Portele (2015) propose that d-pronouns prefer as antecedent the NP which is least accessible, where accessibility is defined at least in terms of the givenness, the syntactic function and the clausal position of the competing antecedent NPs. For example, in a sentence with subject-object (SO) order, the subject is always more accessible than the object because it is favored by two of the three defining properties – syntactic function and clausal position. Thus, even when the object is given, as in [6], it is still less accessible than the subject, and thus the d-pronoun prefers a given antecedent in this case.

19 One additional factor that has to be taken into account is the referential form of the competing antecedents. As has been pointed out in the literature (Bosch & Umbach, 2007), when a clause-final object is a pronoun itself, the d-pronoun prefers to refer to the clause-initial subject referent, contrary to its usual preference. An authentic example of this kind is provided in [7].

[7] [C -2] *Klaus* ging weiter. [C -1] Und wie er eine Strecke gegangen war, kam **ein Kerl** auf *ihn* zu. [T] **Der** sah nicht nur aus wie der Teufel, sondern **er** war es auch.

'[C -2] *Klaus* went ahead. [C -1] After walking a while, some **guy** approached *him*. [T] **He**[d-pronoun] did not only look like the devil, **he**[p-pronoun] was the devil.'

(corpus = "DeWaC-6" text = "494949" id = "http://www.zzzebra.de")

20 The pronoun *ihn* in the last context sentence is given (increasing its accessibility), an object (decreasing its accessibility) and in final position (also decreasing its accessibility). This pronoun should therefore be less accessible than the indefinite NP *ein Kerl*, which has only one feature that decreases accessibility (it is new), but two features that increase accessibility (it is a subject in clause-initial position). If d-pronouns always referred to the less accessible of two potential antecedents, it should not have been used here. The fact that a d-pronoun is nevertheless used to refer back to *ein Kerl* thus indicates that pronouns are inherently more accessible than lexical NPs [4].

4. As will be discussed later, a p-pronoun can still act as antecedent for a d-pronoun under certain conditions.

21 From the perspective of language production, the findings from language interpretation raise a range of interesting questions. First of all, a speaker has to make a choice concerning the linguistic form of a referential expression whether there is an ambiguity or not. Most of the existing literature has been concerned with the interpretation of pronouns and has therefore investigated examples that contain two potential antecedents matching the pronoun in morpho-syntactic features. In this case, a relative decision rule can be used. The accessibility of each potential antecedent is determined, and the one with the highest or lowest accessibility value is chosen as antecedent, depending on the particular pronoun. When only a single potential antecedent is available, reference is not ambiguous and therefore no choice must be made.

22 During language production, however, a choice between p- and d-pronoun is necessary whether a competing antecedent is available or not. Like for language interpretation, relative accessibility may be decisive when the context contains a competing referent, but relative accessibility will be of no help when there is no competing referent. According to Bosch's generalization given in [5], in this case p- and d-pronouns are in free variation, and which of the two is used makes no semantic difference. Unless there is a random choice of pronoun form in contexts lacking competing referents, the speaker needs an absolute decision rule, that is, a decision rule that only considers the properties of the single referent under consideration. Such a decision rule could specify some kind of accessibility threshold. When the accessibility of the referent is above this threshold, the p-pronoun is used, when it is below this threshold, the d-pronoun is used. Given that in certain examples both p- and d-pronouns seem to be acceptable, the threshold must be variable or probabilistic in some way. In the current context, the major question is whether a single notion of accessibility can be found that accounts for the choice between p- and d-pronoun in the absence as well as in the presence of a competing referent.

23 A further question relates to the finding that during language comprehension p-pronouns and d-pronouns seem to be differentially sensitive to the various dimensions defining accessibility. As discussed above, p-pronouns prefer antecedents that have the syntactic function of subject whereas d-pronouns prefer the least accessible referent as antecedent. When comprehending language, the hearer or reader knows which pronoun to interpret. Because the pronoun is provided explicitly as part of the input, item-specific preferences of different pronouns can be retrieved from the mental lexicon and be applied to the task of interpretation. When producing language, in contrast, the speaker or writer must choose a pronoun based on the given state of the referent in the current discourse. Taking item-specific preferences into account when making this choice is thus not as straightforward as it is for language comprehension, because the preferences of several possible referential expressions have to be considered simultaneously. Furthermore, the particular preferences found for interpretation can easily lead to a tie. For example, in order to help the hearer, a speaker who wants to refer back to the subject in an

object-before-subject sentence could either use a p-pronoun (because p-pronouns prefer a subject antecedent) or a d-pronoun (because d-pronouns prefer a sentence final antecedent). The question then is how the speaker nevertheless comes to a decision.

24 Because of its focus on pronoun interpretation, the existing literature does not provide much information on these questions. We know of only two studies that have addressed the choice between p-pronoun and d-pronoun during language production. Bosch et al. (2003) present a corpus study based on the "Negra" corpus, a corpus of German newspaper texts. They found 1,436 p-pronouns and 180 d-pronouns. For p-pronouns with an antecedent in the immediately preceding clause, the antecedent was a subject in 86.7% of all cases and a non-subject in the remaining 13.2% cases. This confirms the strong subject orientation of p-pronouns. For d-pronouns, in contrast, an object bias was found. With 76.4% non-subject and 23.6% subject antecedents, the object bias for d-pronouns was somewhat weaker than the subject bias for p-pronouns. Since Bosch et al. (2003) did not look at other features of the antecedent, their study leaves open whether properties other than the antecedent's syntactic function influence the choice between p- and d-pronoun, or even make reference to the antecedent's syntactic function superfluous.

25 Bittner and Dery (2015) had participants narrate short picture stories and found different preferences for German p- and d-pronouns in terms of discourse coherence. In case of situations not involving anaphoric disambiguation, p-pronouns are used to background the referent whereas d-pronouns serve a forward orientation in the discourse. The authors claim that for both devices, the two types of pronouns can be described in different terms based on the salience and/or activation of their referents: whereas p-pronouns are chosen to refer to salient and activated referents, d-pronouns are chosen to refer to referents that need to be strengthened in terms of salience/activation. Bittner and Dery (2015: 67) found that in situations encompassing pronoun use in anaphoric disambiguation, however, the choice between p- or d-pronouns may not be described in terms of information status of the referent in the ongoing discourse.

26 In order to broaden the empirical basis with regard to the choice between p- and d-pronouns during language production, we conducted a corpus study which was backed up by a production experiment testing the influence of givenness and syntactic prominence. Both the corpus study and the production experiment are confined to written language. Possible limitations resulting from this restriction are discussed below.

3. Corpus study

27 The corpus analyzed in this paper is the "DeWaC" corpus made available by the University of Bologna (see Baroni et al., 2009; and http://wacky.sslmit.unibo.it). The "DeWaC" corpus is a huge part-of-speech tagged corpus of written German built by web

crawling. It contains about 1,600,000,000 tokens of text in ca. 92,000,000 sentences. We first discuss the syntactic construction that we chose for analysis. We then describe how the corpus examples were extracted and prepared for later analysis. We finally present various analyses of the extracted examples.

3.1. Choice of syntactic construction

28 In accordance with the experimental literature on this topic, we restrict our analysis to sentences in which the pronoun occurs clause-initially as the subject within a main clause. This was the case for all the examples considered so far. Excluded from the analysis are thus object pronouns in general and subject pronouns occurring in the so-called middlefield of a German sentence [5]. An initial screening of about a sixth of the "DeWaC" corpus revealed about 149,183 hits for the query "*Er* + finite Verb" and 6,518 hits for the query "*Der* + finite Verb". Thus, sentences with an initial p-pronoun occurred about 23 times more often than sentences with a d-pronoun. This ratio is almost three times higher than the ratio found by Bosch et al. (2003). There are two main differences between the study of Bosch et al. and the current study [6]. First, two different corpora were investigated. Second, we only consider sentences in which a subject pronoun occurs clause-initially within a main clause whereas the position of the pronouns was not restricted in the study of Bosch et al. How these differences account for the different ratios between p- and d-pronouns is an open question [7].

29 For the corresponding accusative object pronouns, similar searches revealed 582 hits for *Ihn* ("him/p-pronoun") followed by a finite verb and 943 hits for *Den* ("him/d-pronoun") followed by a finite verb. Thus, in striking contrast to the case of subject pronouns, for pronouns in the function of a direct object the d-pronoun outnumbers the p-pronoun. The reason for this is the well-known fact that personal pronouns in object function are severely restricted with regard to their occurrence in the prefield of a main clause, whereas d-pronouns are not restricted in the same way (see Lenerz, 1992). Since the syntactic constraints that are responsible for the placement of object pronouns are beyond the scope of the current paper, we only consider subject pronouns in the following.

---

5. The middlefield is that part of a German sentence that is demarcated to the left by the finite verb (main clauses)/the complementizer (embedded clauses) and to the right by the clause-final verbs.

6. One frequent use of d-pronouns is for making reference to a proposition, as in example [i]. In such a case, it is hardly possible to replace the d-pronoun *das* by the corresponding p-pronoun *es*.
[i] Maria hat einen wichtigen Preis gewonnen. Das freut mich sehr.
M. has a important price won. That pleases me much.
'Maria won an important price. I'm very pleased by that.'
Since this case was excluded both in Bosch et al. (2003) and in our study, it cannot be responsible for the different ratios between p- and d-pronouns.

7. We also looked at the "Tiger Corpus" (version 2; http://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger.html) and found a ratio between main clause initial *er* vs. *der* of 20.6, which is only slightly less than the ratio we found for the "DeWaC" corpus.

30 In an additional search, we looked for strings of the form "finite Verb + *er*". This search string corresponds to sentences in which the subject pronoun *er* is located within the first position after the finite verb in a verb-second sentence. This position is assumed to be the preferred place for topics in general and personal pronouns in particular (Rambow, 1993; Frey, 2004). There were 174,098 corpus hits for *er* immediately following the finite verb, which contrasts with the 149,183 hits for *er* immediately preceding the finite verb. Thus, the subject pronoun *er* occurred more often after than before the finite verb. However, the frequency difference is only moderate, and in absolute terms, both constructions are of very high frequency.

31 In sum, restricting our analysis to subject pronouns in sentence-initial position allows us to concentrate as far as possible on factors that are immediately relevant for defining accessibility and thus for choosing between a p- and a d-pronoun. Extending this line of research to other cases, for example to sentences with object pronouns, must be left as a task for future research.

3.2. Corpus preparation

32 We first retrieved all sentences beginning either with the p-pronoun *er* or the d-pronoun *der* immediately followed by a word tagged as a finite verb. For each sentence, the preceding context was also retrieved, limited to five sentences. This resulted in a total number of 940,779 corpus hits. Of these, 901,486 or 95.8% contained the p-pronoun and 39,293 or 4.2% the d-pronoun, resulting in a ratio of 23:1 as in the subset analyzed above.

33 Because the set of corpus hits was too large to be analyzed completely, we drew a random selection of 500 examples for each pronoun. All examples were checked and erroneous examples were removed from the sample. Most of these were false hits because a word that was not a verb had been tagged as verb in the "DeWaC" corpus. In addition, in some cases the preceding context and the target sentence did not form a coherent discourse, reflecting problems with automatically deriving texts from internet sites. Finally, in five examples the d-pronoun *der* was feminine, thus acting as a dative object. The final sample contained 465 instances for the p-pronoun and 436 instances for the d-pronoun. This means that the proportions of p- and d-pronoun examples in our sample does not match the proportion in the complete "DeWaC" corpus. We will therefore always report separate percentages for the two pronouns in the following analyses.

34 All instances were inspected and all NPs co-referent with the pronoun were coded by hand. The last co-referential NP will be called the antecedent NP, with one exception as explained below. For ease of exposition, we will use the term antecedent both for the antecedent NP and for the referent of the antecedent NP in the following, unless the context requires the more specific term. In each of the prior examples [2] and [3], the antecedent is the co-referential NP in the immediately preceding sentence. There are two cases where identifying the antecedent is not straightforward. In the first one, a reflexive pronoun is the last co-referential element, as in [8].

[8] [C -1] Nachdem **der Vogel** soweit mit dem Nest zufrieden war, drehte **er** **sich** um und hüpfte näher ans Fenster, um David anzusehen. [T] **Er** hüpfte auf und ab und zwitscherte laut und erst dann merkte David das krumme Bein!

'[C -1] After **the bird** was satisfied with the nest, **he**[p-pronoun] turned (**himself**) around and hopped closer to the window to look at David. [T] **He**[p-pronoun] leaped up and down twittering loudly and it was not until then that David noticed the injured leg.'

(corpus = "DeWaC-6" text = "500797" id = "http://rsw.beck.de")

35 There were 19 cases of this kind, 15 with a following p-pronoun and 4 with a following d-pronoun. Because in most cases the reflexive was an inherent reflexive, we do not take the reflexive as the antecedent of the following pronoun but the reflexive's antecedent, *er* ("he") in the example above.

36 A related case is illustrated in [9]. Here, the last co-referential NP is the possessive pronoun *seine* which itself is co-referent with the subject NP *Döring*[8].

[9] [C -1] **Döring** begann **seine** berufliche Laufbahn in der Verwaltung. [T] **Er** war unter anderem im Planungsstab des Präsidenten der Universität Hamburg und in der Folgezeit im Bildungsministerium Schleswig-Holsteins tätig.

'[C -1] **Döring** started **his** professional career in administration. [T] **He**[p-pronoun] among others things took part in the planning staff of the president of the Hamburg University and afterwards worked in the Ministry of Education of Schleswig-Holstein.'

(corpus = "DeWaC-9" text = "801409" id = "http://www.todesstrafe-usa.de")

37 Here, one may again wonder whether the antecedent of *er* is the proper name *Döring* or the intervening possessive pronoun *seine*. The issue is somewhat more complicated than in the case of reflexives because the antecedent of the possessive pronoun does not necessarily occur within the same sentence. This was the case in 8 out of the 54 corpus texts where a possessive pronoun was the last expression co-referential with the upcoming sentence-initial pronoun. Of these 54 corpus texts, 48 contained the p-pronoun *er* and 6 the d-pronoun *der*. We removed these 54 corpus texts from our sample and present a separate analysis for them after we have presented the analysis of the main corpus. This way, we can let our data decide which NP is the antecedent of the p- or d-pronoun – the possessive pronoun or its antecedent. The main corpus thus contains 417 instances of the p-pronoun and 430 instances of the d-pronoun.

38 The following properties of the antecedent were coded partly by hand, partly automatically in order to uncover the factors that govern the choice between p- and d-pronoun.

8. This point holds whether the possessive pronoun is analyzed as an NP located in a specifier position of its containing NP or as the head of this NP.

– *Givenness – number of mentions*: the number of NPs referring to the antecedent's referent, including the antecedent NP itself.
– *Givenness – given or new*: the antecedent was classified as given if the preceding context contained at least one additional reference to it, that is, when the number of mentions was two or greater. Otherwise the antecedent was classified as new.
– *Syntactic function*: all antecedents were classified as either subject or non-subject.
– *Position within clause*: when no further referential NP occurred after the antecedent, it was classified as final. Otherwise it was classified as non-final. Non-referential NPs that were not counted when determining the clausal position were predicative NPs in copula constructions and NPs that are non-referential because they are part of an idiomatic expression.
– *Recency*: the number of context sentences intervening between the pronoun and the antecedent. When the antecedent was contained in the context sentence immediately preceding the pronoun sentence, this number was zero.
– *Animacy*: according to prescriptive grammars of German, it is impolite to refer to a person by a d-pronoun unless one wants to put special emphasis on the pronoun (Dudenredaktion, 2011). In order to test for such an influence, we coded all antecedents as human if they referred to persons or collections of humans like institutions or companies. All other antecedents were coded as non-human.
– *Definiteness*: all antecedent NPs were classified into the six definiteness categories shown in [10]. The definitions for proper names, definite NPs and indefinite NPs follow the corpus study of Van Bergen & de Swart (2010).

[10] a. p-pronoun: personal pronouns including possessive pronouns;
b. d-pronoun: demonstrative pronouns used without a following noun;
c. proper name: personal names, place names and names of companies;
d. definite NP: nouns preceded by a definite article, a demonstrative article, a possessive determiner or a strong quantifier;
e. indefinite NP: bare nouns, generic nouns and nouns preceded by a weak quantifier or an indefinite article;
f. w-word: the w-word *wer* ("who"), either in a question or, more often, in a free relative clause, as in example [11].

[11] [C -1] **Wer** dieses Ziel nicht mehr hat, braucht auch keine Wege und Teilschritte zu entwickeln. [T] **Der** verwaltet das Bestehende.
'[C -1] **Who** does not have this goal anymore, does not need to develop ways and substeps. [T] **He**[d-pronoun] maintains what is established.'
(corpus = "DeWaC-6" text = "496753" id = "http://www.brsd.de")

– *Ambiguity*: all masculine singular NPs not co-referent with the pronoun were marked as competitors.

3.3. Descriptive results

39 This section presents the results of the individual properties that were defined above. An analysis that takes all properties into account simultaneously is presented in the next section. All statistical analyses reported here and later were computed using the statistics software R, version 3.2.3 (R Development Core Team, 2015).

3.3.1. Recency, givenness and syntactic prominence

40 We start by considering three properties that are identified by Arnold (2010) as crucial for defining accessibility: the givenness, the syntactic prominence and the recency of the pronoun's antecedent.

41 In order to determine the influence of recency, Table 1 shows the distance between pronoun and antecedent in terms of the number of sentences intervening between pronoun and antecedent. Table 1 reveals that in the vast majority of sentences, no sentence intervenes between pronoun and antecedent. In other words, with few exceptions the antecedent occurs in the sentence immediately preceding the sentence containing the pronoun. This holds for d-pronouns slightly stronger than for p-pronouns. Although the difference is small, it is significant (Fisher's exact test, p = 0.006).

	0	1	2	3	4
P-pronoun	91.1	6.7	2.6	0.0	0.0
D-pronoun	95.8	3.5	0.5	0.2	0.0

Table 1. Percentages of number of sentences intervening between pronoun and antecedent, depending on pronoun type

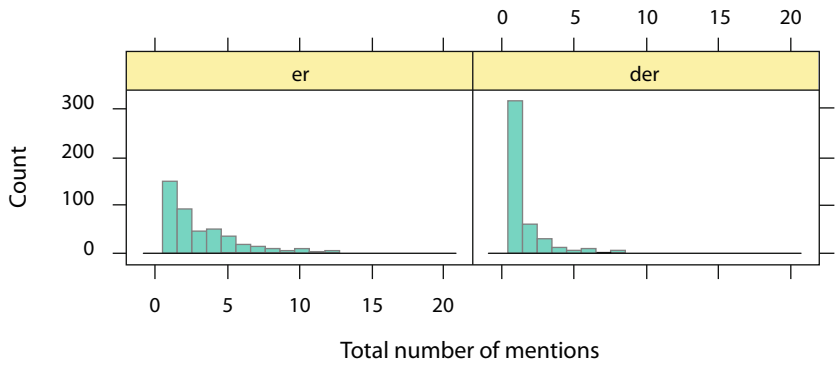


Figure 1. Number of mentions of the pronoun's referent in the preceding context for *er* (p-pronoun) and *der* (d-pronoun)

42 With regard to givenness in terms of number of mentions, consider first Figure 1, which shows how often the referent of the pronoun was referred to in the preceding context. A single mention means that the antecedent NP was the only referential expression co-referent with the pronoun. When the number of mentions was higher than one, the antecedent was preceded by further mentions of the pronoun's referent. Figure 1 reveals a clear difference between the p-pronoun *er* and the d-pronoun *der* (Fisher's exact test, $p < 0.001$). In the majority of all cases (73.0%), the d-pronoun refers to a referent that has been mentioned only once in the preceding context. The number of cases in which the referent of the d-pronoun was mentioned more than once declines rapidly. For the p-pronoun, single-mention referents are the most frequent category too, but with 35.7% of all cases, they occur less often than referents that are mentioned more than once. Among the 64.3% cases where a referent is mentioned more than once, the highest value is found for examples in which the referent of the p-pronoun is mentioned twice, but cases in which the referent of the p-pronoun is mentioned three times or more also occur with some regularity.

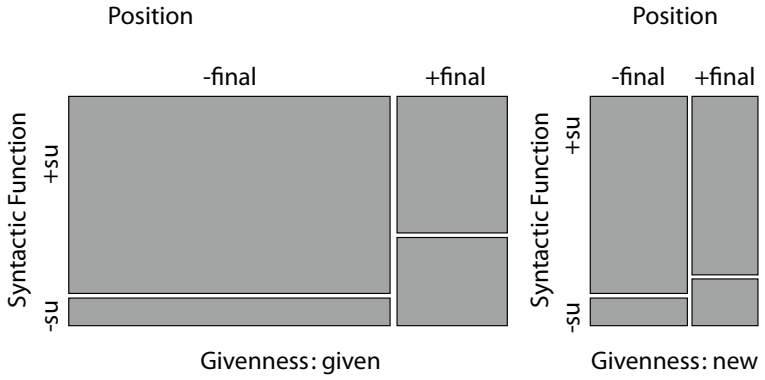
43 As explained above, the numeric variable "number of mentions" was converted into a categorical variable "givenness" with the two values "given" (number of mentions > 1) and "new" (number of mentions = 1). The results for givenness defined in this way and the two syntactic prominence properties of syntactic function and clausal position are shown in Table 2. For each property, this table shows separate percentages for p-pronouns and d-pronouns.

44 For each property shown in Table 2, p-pronouns and d-pronouns behave in opposite ways (givenness: $\chi^2 = 119$; syntactic function: $\chi^2 = 260$; clausal position: $\chi^2 = 137$; all p-values < 0.001). For the p-pronoun, the value that increases accessibility always occurs much more frequently than the value that decreases accessibility. The asymmetry is strongest for the syntactic function of the antecedent (84.7% vs. 15.3%) and weakest for the givenness of the antecedent (64.3% vs. 35.7%). The reverse pattern is found for the d-pronoun: in each case, the accessibility-decreasing value occurs more often than the accessibility-increasing one. Here, all three properties show about the same ratio of ca. 70:30. With regard to the effect of syntactic function, the results in Table 2 are close to those found by Bosch et al. (2003). When the pronoun's antecedent was contained in the immediately preceding clause, Bosch et al. found that the antecedent for a p-pronoun was a subject in 86.7% of all cases whereas the antecedent for a d-pronoun was an object in 76.4% of all cases. The values found in our study are 84.7% subject antecedents for the p-pronoun and 70.2% object antecedents for d-pronoun. Both corpus studies thus find that the subject bias for p-pronouns is stronger than the object bias for d-pronouns.

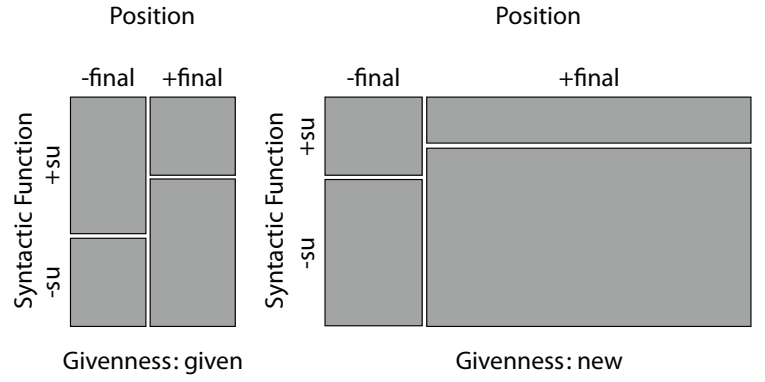
45 The joint distribution of the three properties included in Table 2 is shown in Figure 2 for both the p-pronoun *er* and the d-pronoun *der*. What is most striking is that the two graphs are approximately mirror images of each other. In particular, by far the largest area for the p-pronoun in Figure 2 corresponds to the feature

	Givenness		Syntactic function		Clausal position	
	given	new	subject	non-subject	non-final	final
P-pronoun	64.3	35.7	84.7	15.3	68.8	31.2
D-pronoun	27.0	73.0	29.8	70.2	28.6	71.4

Table 2. Percentages of given vs. new, subject vs. non-subject, and non-final vs. final antecedent NPs, depending on pronoun type



P-pronoun *er*



D-pronoun *der*

Figure 2. Joint distribution of the three properties "syntactic function", "position" and "givenness" of the antecedent of the p-pronoun *er* and the d-pronoun *der*

combination "+subject, -final and given", taking up 43.2% of all cases. For the d-pronoun in Figure 2, in contrast, the opposite feature combination "-subject, +final and new" takes up 45.4% and is thus as dominant as its counterpart for the p-pronoun. A further noteworthy finding revealed by Figure 2 is that for both pronouns, our corpus sample contains examples for all eight feature combinations. Thus, the three major properties defining accessibility discussed so far do not provide a categorical distinction between p-pronoun and d-pronoun, neither alone nor in combination.

46 In sum, of the core properties defining salience according to Arnold (2010), givenness and syntactic prominence (syntactic function and clausal position) strongly differ between p-pronouns and d-pronouns. The antecedent of a p-pronoun is typically a given NP that occurs as subject in a non-final clausal position. The antecedent of a d-pronoun, in contrast, is typically a new NP that occurs as non-subject in clause-final position. Recency, in contrast, only showed a minimal difference between p-pronouns and d-pronouns. Both pronouns seem to require an antecedent that occurred recently and is therefore in an activated state in working memory. In this respect, pronouns differ from lexical NPs, which often take their antecedent over a longer distance (see Arnold, 2010 for discussion). The similar behavior of p- and d-pronouns with regard to recency can be attributed to the fact that both have the same impoverished lexical content ("masculine singular").

3.3.2. Definiteness

47 Before discussing how the definiteness of the antecedent NP influences the choice between a p- and a d-pronoun, we first consider the relationship between the antecedent's definiteness and its discourse status as given or new. Definiteness and givenness are expected to correlate. For example, when the antecedent is itself a pronoun, it must be given, that is, it must be preceded by some non-pronominal NP in the prior discourse. On the other hand, when the antecedent is an indefinite NP, it has likely been newly introduced to the discourse. In order to assess how strong the correlation between definiteness and givenness of the antecedent is, Table 3 shows the percentages of given and new uses for each definiteness category of the antecedent NP. When the antecedent was a p- or d-pronoun itself, it was almost always given. The fact that p-pronoun antecedents were given only 97% of the time is due to restricting the prior context to five sentences. The opposite behavior is found for indefinite NPs and w-words, which are new in the overwhelming majority of cases. Proper names and definite NPs are in between, with a slight preference for given antecedents for proper names and a moderate preference for new antecedents for definite NPs. The finding of 69.2% new uses for antecedents that are definite NPs may seem surprising if one considers the anaphoric use of definite NPs as basic. However, prior research has demonstrated that the use of definite NPs without a textual antecedent is quite common. For example, Fraurud (1990) found that 60.9% of the definite NPs were new in his corpus study. With 69.2%, the value that we found is only slightly higher. This slight increase is possibly again due to restricting the context to five sentences.

	P-pronoun	D-pronoun	Proper name	Def	Indef	W-word
N	165	8	192	335	136	7
Percentage given	97.0	100.0	51.6	30.8	7.4	0.0
Percentage new	3.0	0.0	48.4	69.2	92.7	100.0

Table 3. Relationship between definiteness and givenness of the antecedent NP

	P-pronoun	D-pronoun	Proper name	Def	Indef	W-word
P-pronoun (417)	32.1 (134)	0.2 (1)	26.6 (111)	31.9 (133)	8.9 (37)	0.2 (1)
D-pronoun (430)	7.3 (31)	1.6 (7)	19.0 (81)	47.4 (202)	23.2 (99)	1.4 (6)

Table 4. Percentages (n) of definiteness categories of antecedents depending on pronoun type

	Proper name antecedent		Definite antecedent	
	given	new	given	new
P-pronoun	65.8 (73)	34.2 (38)	44.4 (59)	55.6 (74)
D-pronoun	32.1 (26)	67.9 (55)	21.8 (44)	78.2 (158)

Table 5. Percentages (n) of given vs. new antecedent NPs depending on pronoun type, for proper name antecedents and definite antecedents

48 Table 4 shows that the p- and the d-pronoun are associated with different distributions of the antecedent's definiteness (Fisher's exact test, $p < 0.001$). For the p-pronoun, the two most frequent definiteness categories are pronoun and definite NP, closely followed by proper name. Taken together, these three categories account for about 90% of all antecedents for the p-pronoun. For the d-pronoun, the three most frequent categories are proper name, definite NP and indefinite NP. These three also account for about 90% of all antecedents for the d-pronoun. There are two major differences between the p- and the d-pronoun. First, the p-pronoun's antecedent is a p-pronoun itself in a substantial number of cases, whereas p-pronouns are rarely the antecedent for the d-pronoun, although some cases still occur. The second major difference is that indefinite antecedents occur much more often with the d-pronoun than with the p-pronoun.

49 These differences notwithstanding, Table 4 shows a large overlap between the p- and d-pronoun. For both pronouns, definite NPs and proper names together account for the majority of all antecedents (58.5% for the p-pronoun and 66.4%

for the d-pronoun). For these two antecedent types, Table 5 shows how often the antecedent was already given in the prior discourse and how often it was newly introduced.

50 For proper name antecedents, we see a distribution that is close to the distribution found in the whole corpus (see Table 2). Proper name antecedents of p-pronouns are more often given than new whereas proper name antecedents of d-pronouns are more often new than given. For definite NP antecedents, both pronouns show a majority of new uses, which is much stronger for d- than for p-pronouns. For the syntactic prominence features, proper name and definite NP antecedents vary even stronger depending on the following pronoun. For example, 87% of the definite NP antecedents of the p-pronoun are subjects, whereas 80% of the definite NP antecedents of the d-pronoun are non-subjects.

51 In sum, the antecedent's definiteness has some predictive value with regard to the choice between p- and d-pronoun when the antecedent is a pronoun or, to a lesser extent, when it is an indefinite NP. However, in these cases definiteness is strongly correlated with givenness. When the antecedent is a proper name or a definite NP, in contrast, definiteness is of no help in deciding between p- and d-pronoun. The results of this section therefore suggest that the choice between p- and d-pronoun is a matter of the actual discourse status and the syntactic properties of the antecedent, not of the formal realization in terms of definiteness.

3.3.3. Ambiguity

52 As discussed in the introduction, the prior literature looked at the distinction between p- and d-pronouns mainly from the perspective of language comprehension. This implies a focus on ambiguity, which is present when the preceding context contains at least one competitor to the actual antecedent. During language production, in contrast, a choice concerning the referential form used for referring to an entity given in the context has to be made even when the context does not contain a competitor. An important question is therefore whether ambiguity affects the choice between p-pronoun and d-pronoun.

53 Given the almost inviolable morpho-syntactic constraint that pronoun and antecedent must match with regard to gender and number, an ambiguity arises in our corpus sample when there is at least one additional NP specified for the features "masculine singular". When this NP occurs within the last context sentence, that is, in the sentence immediately preceding the sentence containing the pronoun, we will call it a local competitor. Table 6 shows the frequency of a local competitor as well as the frequency of a competitor anywhere within the preceding context.

54 When the competitor can occur anywhere, the large majority of corpus examples contains at least one competitor to the actual antecedent, with no difference between p-pronoun and d-pronoun ($\chi^2 = 0.04$, $p = 0.9$). When only looking at examples with a local competitor, a difference between p-pronoun and d-pronoun emerges ($\chi^2 = 10$, $p < 0.01$). For the p-pronoun, a local competitor is present in 35.7% of all cases.

	Competitor anywhere		Local competitor	
	present	absent	present	absent
P-pronoun	89.9	10.1	35.7	64.3
D-pronoun	89.5	10.5	46.3	53.7

Table 6. Percentages of sentences with a competitor to the actual antecedent anywhere or locally in the prior context, depending on pronoun type

This number increases to 46.3% for the d-pronoun. Thus, the d-pronoun is used somewhat more often in situations of ambiguity than the p-pronoun. Importantly, however, in slightly more than half of the cases where a d-pronoun is used, no local competitor is present. This shows that the d-pronoun is not primarily used for purposes of ambiguity avoidance.

3.3.4. Animacy

55 The final property that we consider is the animacy of the referent that the pronoun refers to. Because of the prescriptive rule mentioned above, we consider only the distinction between humans and non-humans in our study.

56 The proportions of human and non-human referents depending on pronoun type are shown in Table 7. Both pronouns show a clear bias towards human referents. This bias is stronger for the p-pronoun than for the d-pronoun ($\chi^2 = 20$, $p < 0.001$), although the difference is not a large one. An inspection of the examples did not provide evidence that the d-pronoun generally goes along with special emphasis, although – as discussed below – in some cases emphasis does indeed seem to play a role. The general bias toward human referents thus suggests that the prescriptive advice against the use of d-pronouns for human referents does not have a strong effect if it has an effect at all. A minor effect of this prescriptive advice cannot be excluded, as evidenced by the higher rate of non-human referents for the d-pronoun than for the p-pronoun. However, this finding is open to alternative explanations. In particular, it could follow from the fact that p-pronouns are subject-oriented whereas d-pronouns are object-oriented. Since most subjects are human whereas most objects are non-human, d-pronouns may more often refer to a non-human referent than p-pronouns. The two reasons for the observed difference between p- and d-pronouns with regard to animacy do of course not exclude each other. Both may account for some part of the observed difference.

	Animacy	
	human	non-human
P-pronoun	81.1	18.9
D-pronoun	67.4	32.6

Table 7. Percentages of human and non-human referents depending on pronoun type

3.4. Predicting referential choice

57 As pointed out above, the sample sizes for the p-pronoun and the d-pronoun do not reflect the ratio that is found in the "DeWaC" corpus. While we analyzed samples of about equal sizes, in the "DeWaC" corpus the p-pronoun *er* occurs about 23 times more often than the d-pronoun *der* at the beginning of a main clause. Computing a logistic regression analysis for our complete sample would thus give invalid estimates. In order to approximate the original p-pronoun to d-pronoun ratio, we formed a sample consisting of the complete p-pronoun sample and 18 randomly drawn examples from the d-pronoun sample. This new sample was analyzed by means of logistic regression. This process was repeated 1,000 times.

58 Table 8 shows the mean percentages of d-pronoun sentences in the 1,000 corpus samples reflecting the ratio of d-pronoun and p-pronoun in the complete "DeWaC" corpus. In all conditions, a minority of the sentences contains a d-pronoun. As expected, by far the highest value is found in the condition where the use of a d-pronoun is favored most strongly, but even here a mean value of 38.3% means that the p-pronoun is still favored.

	Given		New	
	+subject	-subject	+subject	-subject
-Final	0.7	3.9	1.2	16.6
+Final	1.8	6.8	4.1	38.3

Table 8. Mean percentages of d-pronoun sentences in 1,000 corpus samples reflecting the ratio of d-pronoun and p-pronoun in the complete "DeWaC" corpus

Factor	Mean estimate	Mean *t*-value	Percentage significant
Syntactic function: -subject	0.13600	5.5577	99.7
Givenness: new	0.04577	2.1109	56.6
Position: +final	0.03616	1.7587	42.9
Local competitor: true	0.02182	1.1146	18.8
Competitor: true	-0.02323	-0.7546	11.1
Animacy: human	0.00246	0.0979	6.5
Antecedent: pronoun	-0.01167	-0.5204	0.3

Table 9. Results of the logistic-regression analysis for the corpus data

59 For each factor, Table 9 shows the mean estimate, the mean *t*-value and the percentage of significant *t*-values, obtained in 1,000 runs. The factor "syntactic function" was significant in almost every run. All other factors were significant much less often, with "givenness" and "position" still being significant in a substantial number of times. Somers' D reached a mean value of 0.87 (range: 0.73-0.97). When the fitted probability values were converted into binary choices between p- and d-pronoun, the mean success rate was 96.0 (range: 94.9%-97.5%). This value corresponds almost exactly to the 95.9% p-pronoun sentences in the analyzed samples. And in fact, the accuracy was almost perfect for the p-pronoun *er* (mean: 99.7%; range: 98.1%-100%) but rather low for the d-pronoun *der* (mean: 8.3%; range: 0%-61.1%). The strong dominance of the p-pronoun thus causes the models to overgeneralize the use of the p-pronoun, making it rather difficult to correctly predict the use of a d-pronoun.

3.5. The status of possessive pronouns

60 As discussed above, when the last co-referential item of the pronoun is a possessive pronoun, the question arises whether the pronoun's antecedent is the possessive pronoun itself or the possessive pronoun's antecedent. For example, is the antecedent of the p-pronoun *er* in example [9] the possessive pronoun *seine* or the proper name *Döring*? Of the 54 examples with a possessive pronoun as potential antecedent, 48 were followed by the p-pronoun *er* and 6 by the d-pronoun *der*. Because the low number of *der* examples makes it difficult to arrive at firm conclusions, we only consider the 48 instances with *er* in the following. Table 10 shows the distribution of the possessive pronoun and the antecedent of the possessive pronoun depending on the three major factors "givenness", "syntactic function", and "clausal position". For the possessive pronoun, the two factors "syntactic function" and "clausal position" are defined with regard to the phrase containing the possessive pronoun.

61 Given the main results presented above, the antecedent of the p-pronoun *er* typically has the following properties: it is a subject NP that occurs in clause initial position and it is already given in the prior discourse. Table 10 shows that the antecedent of the possessive pronoun has all three properties in the majority of cases whereas for the possessive pronoun itself this is only true for the property of givenness. The clearest deviation from the pattern typical of the antecedent of a p-pronoun concerns the antecedent's syntactic function. The possessive pronoun is part of a non-subject in the large majority of cases. The possessive pronoun's antecedent, in contrast, is almost always a subject. Given the data in Table 10, we conclude that the antecedent of the possessive pronoun is the antecedent of the p-pronoun, not the possessive pronoun itself.

3.6. Discussion

62 The corpus data presented in this section confirm that p- and d-pronouns differ in a range of dimensions as expected given the prior literature. The strongest differences were found for two core properties defining accessibility – givenness and syntactic prominence, for which both syntactic function and clausal position had strong

	Givenness		Syntactic function		Clausal position	
	given	new	subject	non-subject	non-final	final
Possessive pronoun	100	0	19	81	48	52
Antecedent of possessive pronoun	71	29	92	8	96	4

Table 10. Percentages of given vs. new, subject vs. non-subject, and non-final vs. final antecedent NPs, for possessive pronouns as antecedents and for antecedents of possessive pronouns

effects. For recency, in contrast, we found only a minimal difference between the two pronoun types. In addition, we found minor differences between the two pronoun types with regard to the definiteness and animacy of the antecedent and with regard to the presence of a local competitor. The question of which of the observed differences are ultimately decisive for choosing a p- or a d-pronoun could only be answered tentatively. A major obstacle in this regard was the fact that the p-pronoun *er* occurred with a much higher rate than the d-pronoun *der*. As a consequence, the logistic regression model overpredicted the use of the p-pronoun, which resulted in a low success rate for the d-pronoun. Before we elaborate on this point, we first present an experimental study that further explores some of the findings obtained so far.

4. Experiment: choice of referential pronouns in subject-object structures

63 To broaden the picture of pronominal referential choice in sentence production, we conducted a sentence completion experiment investigating the preferred choice of p- and d-pronouns in German. We were especially interested in the question whether the corpus finding that a d-pronoun is chosen to refer to a given NP in sentence final position also holds for completions written by participants.

4.1. Method

4.1.1. Participants

64 Twenty-four German native speakers completed a written questionnaire. They were students at the Goethe University Frankfurt and received course credit for participation in the experiment.

4.1.2. Materials

65 We constructed 16 experimental items each consisting of three context sentences, as illustrated by the original experimental item in [12] [9].

9. In the original questionnaire, the referents that are underlined here were surrounded by a frame.

[12] [C -3] Maria war am Sonntag im Zirkus. [C -2] Bevor die Aufführung begann, hatte sie schon einen Clown herumlaufen sehen.

'Maria visited a circus on Sunday. Before the show, she saw a clown walking around.'

[C -1] A. *Given NP in initial position*
Der Clown umarmte einen Mann, der ganz wirre Haare hatte.
The clown hugged a man who had very shaggy hair.

B. *Given NP in final position*
Ein Mann, der ganz wirre Haare hatte, umarmte den Clown.
A man who had very shaggy hair hugged the clown.

[T] CONTINUATION: _____

Instruction: write down a sentence starting with a pronoun referring to the underlined NP.

66 In the first sentence, an individual was introduced using a proper name or indefinite NP. In the second sentence, a second entity, which was always masculine, was introduced using an indefinite NP. In all cases, the final context sentence [C -1] is a sentence with subject-before-object word order. What is varied is whether the given NP precedes the new NP or whether it is the other way round. We decided to hold word order constant and to vary the position of the given and new NP for two reasons. First, when the subject is given and the object is new and participants are required to refer to the object, the use of a d-pronoun is strongly favored, as this is the prototypical situation for using a d-pronoun. Our corpus results nevertheless suggest that even here the p-pronoun may be used more often than the d-pronoun. This condition thus serves as a baseline for the use of d-pronouns.

67 The second reason to concentrate on SO sentences concerns the condition where the subject referent is new and the object referent is given and the given referent thus occurs in sentence-final position. This condition has been neglected so far, because prior research has typically investigated sentences where the first NP was given and the second NP new. When the subject is new and the object is given and participants have to refer to the object, d-pronouns should be avoided according to the generalization that d-pronouns refer to non-topical antecedents, at least when the given object arguably is the sentence topic. On the other hand, if a d-pronoun is used to refer to the less accessible antecedent, a d-pronoun should be used in this case. The second question addressed in the following experiment thus is how the rate of choosing a d-pronoun differs depending on whether an object in clause-final position is new or given.

68 The critical third context sentence differed depending on the experimental condition. The givenness of these two constituents was also reflected in terms of the definite or indefinite article of the NP. The new and indefinite entity of the last sentence was accompanied by a relative clause in order to give it more content and thereby increase the naturalness of the sentences. Crossing the givenness of the

underlined antecedent (given or new) and its clausal position (first or last) resulted in four experimental conditions shown in [12] (for reasons of space, both referents are underlined in [12]).

69 Four versions of the questionnaire were created, so that each experimental item appeared in a different condition in each version of the questionnaire. The experimental items were supplemented with 32 filler items that – other than the experimental material – also contained non-male and inanimate entities in the third context sentence. Target items were separated from each other by at least two filler items.

4.1.3. Procedure

70 Participants were asked to read the short context passages and to write down a continuation sentence. They were instructed to start their sentence with a pronoun (*er/sie/es* or *der/die/das*) referring to the underlined constituent of the last context sentence. In the questionnaire, only one of the referents in context sentence [C -1] was underlined. Participants' continuations were coded according to which type of pronoun (p- or d-pronoun) they chose to refer to the underlined entity of the previous sentence. 25 continuations had to be excluded from our analysis: they either started with a definite full lexical NP (n = 5), the alternative demonstrative pronoun *dieser/diese/dieses* (n = 1), an inanimate referent mentioned in the context instead of the intended referent (n = 5), or a neuter pronoun (*das* [*that*]/*es* [*it*]; n = 7). In addition, in five cases the d-pronoun was followed by a verb-final structure and two continuations were incomplete or missing. In total, 359 pronominal continuations were analyzed with about an equal number of cases in each of the four experimental conditions.

4.2. Results

71 The percentages of the two pronominal continuation types in the different conditions are shown in Table 11. Responses were analyzed by means of linear mixed-effects logistic regression using the R-package "lme4" (Bates et al., 2015). As the dependent variable, we took proportions of the choice of the d-pronoun. For the fixed factors, effect coding was used. Because a logistic regression cannot be computed when one or more cells contain a value of 0% or 100%, we set the mean value in the cell "given antecedent in first position" to 2% by random resampling. We included participants and items as crossed random effects. Following the advice given in Barr et al. (2013), we first computed a model containing the full factorial design in the random slopes, but this model did not converge. When the interaction terms were dropped from the random factors, a converging model resulted. For each contrast, Table 12 shows the estimate, the standard error, the resulting *z*-value and the corresponding *p*-value.

72 As can be seen in Table 11, there is a general preference for the personal pronoun *er*. This preference is strongest when the antecedent is given and appearing as a subject in the first position of the sentence. Being new in the discourse minimally reduces

	NP1/Subj. given, NP2/Obj. new		NP1/Subj. new, NP2/Obj. given	
Chosen pronoun	reference to NP1	reference to NP2	reference to NP1	reference to NP2
P-pronoun *er*	100 (94)	80.2 (69)	96.7 (87)	65.2 (58)
D-pronoun *der*	0 (0)	19.8 (17)	3.3 (3)	34.8 (31)

Table 11. Percentages (n) of chosen pronoun type in the continuations depending on the order of the given and new NP and position (syntactic function) of the antecedent NP

	Estimate	Standard error	*Z*-value	Pr(> |z|)
Intercept	-3.202	0.476	-6.73	$p < 0.001$
Givenness	1.372	0.946	1.45	0.147
Position	-2.974	0.962	-3.09	0.002
Givenness by position	1.779	1.704	1.04	0.296

Table 12. Mixed-effects model for the results of the completion experiment

this preference. A stronger influence on the choice of pronominal form can be seen when looking at the position of the antecedent in the sentence and its syntactic function. Being an object appearing in last position raises the choice of the d-pronoun *der* up to 19.8% for new antecedents and to 34.8% for given antecedents. Despite the numeric difference between given and new antecedents in clause-final position, the interaction between givenness and position did not reach significance.

73 This difference could possibly be due to some form of priming caused by the determiner of the antecedent. The determiner of a new NP was an indefinite article whereas the determiner of a given NP was a definite article. Since the definite article and the d-pronoun share forms (or are even the same lexical item, as proposed by Wiltschko, 1998), a definite determiner may increase the probability of using the d-pronoun.

4.3. A note on coherence relations

74 As mentioned in the introduction, coherence relations play a major part in the interpretation of pronouns (see Kehler & Rohde, 2013 for a recent review). During language production, coherence relations are also crucially involved with regard to the decision about which referent of a given sentence to take up in the next sentence. However, when a referent has been selected and a choice has to be made as to the referential form for referring to it, coherence relations have been found to have no effect (e.g., Fukumura & Van Gompel, 2010; Kehler & Rohde, 2013). Since we are ultimately interested in both interpretation and production of pronominal

forms, we report the coherence relations observed in our experimental data in this subsection. We analyzed the continuations that participants wrote in our experiment according to the four coherence relations shown in [13]-[16]: *elaboration*, *explanation*, *narration*, *result*.

[13] Elaboration:

Der Mitschüler ärgerte einen Jungen, der eine viel zu große Mütze trug. Er schubste ihn dabei mehrmals.

'The classmate teased a boy, who was wearing an oversized cap. In doing so, he pushed him several times.'

[14] Explanation:

Der Koch schubste einen Bäcker, der am Ende der Schlange stand. Er war schlecht gelaunt.

'The cook pushed a baker, who was standing at the end of the line. He was in a bad mood.'

[15] Narration:

Der Wanderer stützte einen Kameraden, der schon sehr alt war. Er trug ihn so bis ins Tal.

'The hiker physically supported a comrade, who was very old. In this manner, he carried him into the valley.'

[16] Result:

Ein Besucher, der grüngefärbte Haare hatte, provozierte den Sicherheitsmann. Er schmiss ihn daraufhin hinaus.

'A visitor, who had green hair, provoked the security guy. He subsequently expelled him.'

75 An elaboration provides more information of the eventuality described in the preceding sentence. In [13], the continuation specifies the action of teasing somebody to pushing somebody. An explanation may be postulated when the continuation describes a cause or reason of the event pictured in the context sentence. In [14], the continuation brings us to notice that the cook performed the action of pushing because he was in a bad mood. We coded narration for sentences that (temporally) continued the discourse by responding to the event of the preceding context sentence. The second sentence in [15] takes up the action introduced in the preceding sentence and adds the information that this condition lasted until they reached the valley. Finally, a result relation holds when the context sentence provides a cause to the event or state of the continuation written by the participants. In [16], the provocation described in the context sentence causes the expulsion that is pictured in the continuation.

	Coherence relation			
Chosen pronoun	elaboration	explanation	narration	result
P-pronoun *er*	52 (94)	16 (29)	15 (27)	17 (31)
D-pronoun *der*	100 (3)	0 (0)	0 (0)	0 (0)

Table 13. Percentages (n) of the coherence relation established with the continuation depending on chosen pronoun type when the subject of the last context sentence was the antecedent of the pronoun

	Coherence relation			
Chosen pronoun	elaboration	explanation	narration	result
P-pronoun *er*	40.2 (51)	13.4 (17)	9.4 (12)	37.0 (47)
D-pronoun *der*	20.8 (10)	6.2 (3)	8.3 (4)	64.6 (31)

Table 14. Percentages (n) of the coherence relation established with the continuation depending on chosen pronoun type when the object of the last context sentence was the antecedent of the pronoun

76 As can be seen in Table 13, when participants had to refer to the subject of the last context sentence, about half of the cases of a chosen p-pronoun continue the discourse with an elaboration of the event or participants of the previous sentence. The remaining cases are about evenly distributed across the other three coherence relations. For the d-pronoun, all continuations represent an elaboration of the preceding sentence, but since this number is extremely small (n = 3), we cannot conclude anything from it.

77 A different picture emerges when reference was made to the object of the context sentence, as can be seen in Table 14. For the p-pronoun, we find a similar number of elaborations and results on the one hand and also a similar distribution of explanations and narrations on the other hand. For the d-pronoun, however, we see a difference regarding the four coherence relations. Most of the continuations serve as a result for the event of the preceding context sentence.

78 In sum, the data presented in this section indicate that the referential form that is chosen for referring to a particular referent in a continuation sentence is not independent of the coherence relation that connects the continuation sentence with the preceding context. This contrasts with findings that there is no connection between choice of referential form and coherence relations (e.g., Fukumura & Van Gompel, 2010). Before any conclusions can be drawn from this finding, further experiments are necessary in order to see whether this finding, which was obtained in a situation in which participants were not free with regard to the choice of an antecedent phrase, can be confirmed.

4.4. Discussion

79 The results of our experiment show that the syntactic prominence of the antecedent – which was jointly determined by position and syntactic function – has an influence on the referential choice of p- and d-pronouns in contexts including two possible antecedents. The characteristics of an antecedent being an object in sentence-final position makes it more likely to be taken up again as a d-pronoun. However, the rate of d-pronouns is still low even in cases where it is favored by several properties, as when the referent is an object that is new to the discourse. In these cases it is still the personal pronoun *er* that is chosen for the greater part. This leaves us with the finding that though the p-pronoun *er* is biased to subject antecedents, it is also easily used for referring back to objects.

5. Beyond accessibility

80 The results presented so far have shown that the choice between p- and d-pronoun is strongly influenced by two of the main factors defining accessibility, namely givenness and syntactic prominence, which itself can be further decomposed into syntactic function and clausal position. However, it has also become clear that more than accessibility is needed for predicting whether writers use a p- or a d-pronoun. In order to identify cases that are not captured by accessibility, we consider first how far we get when we use accessibility alone for deciding whether to use a p- or a d-pronoun. Since this decision has to be made even if there is no competitor to the actual antecedent, an absolute decision criterion is needed. To this end, let us define the accessibility value of a potential antecedent as the sum of its accessibility increasing properties, where accessibility increasing properties are given, +subject and -final. The accessibility value thus ranges from 0 to 3. Table 15 shows the distribution of p- and d-pronoun according to the accessibility value just defined.

Accessibility value	Syntactic function	Position	Givenness	*Er*Choice	*Er*Prop	*Der*Choice	*Der*Prop
3	+su	-final	+given	180	43.17	**33**	**7.67**
2	-su	-final	+given	20	4.80	**21**	**4.88**
2	+su	-final	-given	78	18.71	**24**	**5.58**
2	+su	+final	+given	46	11.03	**21**	**4.88**
1	-su	-final	-given	**9**	**2.16**	45	10.47
1	-su	+final	+given	**22**	**5.28**	41	9.53
1	+su	+final	-given	**49**	**11.75**	50	11.63
0	-su	+final	-given	**13**	**3.12**	195	45.35

Table 15. Distribution of corpus examples with p- or d-pronoun depending on the accessibility value of the antecedent NP

81 Given that the accessibility value ranges from 0 to 3, the simplest decision criterion is one that predicts the use of the p-pronoun *er* when accessibility is high (2 or 3) and the use of the d-pronoun *der* when accessibility is low (0 or 1). Thus, for half of the eight combinations of the three features under consideration *er* is predicted, for the other half *der* is predicted. The numbers printed in boldface are the cases that are erroneously classified by this simple decision criterion. It makes correct predictions for 77.7% of the corpus examples, for both the p-pronoun and the d-pronoun. For the p-pronoun, this is lower than what was achieved by the logistic regression model introduced above, but it is higher in the case of the d-pronoun. This difference reflects the fact that the decision criterion does not take into account that p-pronouns are much more frequent than d-pronouns in our corpus sample, something which is taken care of by the logistic regression analysis. More important than the absolute success rates is what we learn by taking a closer look at the cases where the accessibility-based criterion fails.

82 Starting with p-pronouns, consider first antecedents with an accessibility value of 0, that is, final objects that are new to the prior discourse. This is the prototypical context for the use of a d-pronoun, and p-pronouns are therefore least expected. Our corpus sample nevertheless contains a small set of 13 examples where the p-pronoun's antecedent has an accessibility value of 0. A particularly striking example is provided in [17].

[17] [C -4] 1982 übernahm *Haimerl* das Referat für Datenverarbeitung, das *er* fast 18 Jahre leitete. [C -3] 1984 wurde *er* stellvertretender Leiter der Zentralabteilung. [C -2] Chef der Abteilung "Ländlicher Raum, Betriebswirtschaft und Förderung" war er seit 1999. [C -1] Zu *Haimerls* Nachfolger hat der Minister **Ministerialrat Friedrich Mayer** berufen. [T] **Er** war bislang Leiter des Referats "Grundsatzfragen der Agrarförderung und agrarpolitische Sonderaufgaben".

'In 1982 Haimerl took over the department of data-handling, which he managed for almost 18 years. In 1984 he became deputy head of the central department. Since 1999 he was head of the department of "rural area, business administration and support". The minister employed **undersecretary Friedrich Mayer** as Haimerl's successor. **He**[p-pronoun] so far was head of the department of "fundamental issues of agricultural aid and secial tasks in agricultural policy".'

(corpus = "DeWaC-5" text = "438281" id = "http://www.stmlf.bayern.de")

83 Every account of pronoun choice in German would have predicted a d-pronoun in this case, especially given the presence of competing antecedents within the final context sentence [C -1]. Nevertheless, the author of this text decided to use the p-pronoun instead. We suspect that this is an effect of the prescriptive rule to avoid d-pronouns for human referents. Although this rule does not in general prevent writers to use d-pronouns for referring to humans, some effect cannot be excluded. The text in [17] is from the website of a state department where a deliberately formal writing style may be in use. Classifying all web sources of our

corpus sample as to formal style is beyond the scope of this paper. We must leave this as a topic for future research. It is clear, however, that this factor cannot act in a categorical way. For example, our very first example for the use of a d-pronoun, example [3], is from one of the most prestigious German newspapers. As witnessed by the example, this did not prevent the writer from using a d-pronoun for referring to the former German chancellor.

84 We now turn to the data for the d-pronoun. With about 45%, the largest class is formed by examples where the accessibility score is 0. The three classes with an accessibility value of 1 all occur with a much lower frequency of about 10%, and the three classes with an accessibility value of 2 with an even lower frequency of about 5%. Surprisingly, however, the frequency of examples where the antecedent has the highest possible accessibility value of 3 does not go down even further but increases to a value of about 7.7%. Examples of this kind are least expected in the case of the d-pronoun because the d-pronoun's antecedent is a non-final subject that is already given in the discourse. There were 33 examples of this kind, in contrast to the 13 corresponding examples for the p-pronoun. A closer inspection revealed a high proportion of examples were the antecedent was either a p-pronoun (n = 12 or 31.6%) or a d-pronoun (n = 7 or 18.4%).

85 A p-pronoun as antecedent of a d-pronoun is found in the example in [18].

[18] [C -5] Zusammen sind wir drei nach Athen weitergeflogen. [C -4] Apropos **Ludger**: [C -3] Abends vorher habe ich mir noch die Eröffnungsfeier im Fernsehen angeguckt. [C -2] War schon ne phänomenale Show. [C -1] Und **unser Ludi** war echt Klasse, wie **er** die Fahne der gesamten deutschen Mannschaft getragen hat. [T] **Der** ist extra am Freitag nach Athen geflogen, um mit der Mannschaft einzumarschieren.

'Together the three of us flew to Athens. Speaking of **Ludger**: I watched the opening ceremony on television the evening before. It was a phenomenal show. And **our Ludi** was awesome, **he**[p-pronoun] carried the flag of the entire German team. **He**[d-pronoun] specially flew to Athens on Friday so he could march in with the team.'

(corpus = "DeWaC-2" text = "166111" id = "http://www.fn-dokr.de")

86 Examples of this kind have already been discussed in the prior literature (see Bosch, 2013). Here, the d-pronoun seems to be used for reasons of emphasis. The same seems to hold in the examples where the antecedent of a d-pronoun is itself a d-pronoun, as in [19].

[19] [C -5] Und als Jesus gestorben ist, schlagen sie sich an die Brust, ein wortloses Zeichen für Reue und Scham. [C -4] Schweigend gehen sie auseinander. [C -3] Sie können es nicht in Worte fassen, aber sie haben gemerkt, dass da mitten unter ihnen etwas Schlimmes geschehen ist, etwas, was nicht wiedergutzumachen ist. [C -2] Der einzige, der für dieses Gefühl Worte findet, ist **der römische Offizier, der Kommandant des Hinrichtungskommandos**. [C -1] **Der** hat schon viele sterben sehen. **Der** hat sie alle erlebt, wie sie fluchen, schimpfen, betteln und flehen, [...].

> '[C -5] And when Jesus died, they hit their breast as a wordless sign of remorse and shame. [C -4] They silently dispersed. [C -3] They could not describe it in words, but they noticed that something bad happened amongst them that could not be made up for. [C -2] The only one finding words for this feeling is **the Roman officer, the commander of the execution squad**. [C -1] **He**[d-pronoun] has seen a lot of people die. [T] **He**[d-pronoun] has witnessed their swearing, insulting, begging, and craving.'
>
> (corpus = "DeWaC-7" text = "659508" id = "http://www.evlka.de")

87 Here again, emphasis seems to play a role. Furthermore, there is a kind of parallelism effect. The two sentences following the introduction of the antecedent in [C -2] are both elaborations of the antecedent NP and thereby form a list-like structure.

88 A p-pronoun could also have been used in the examples in [18] and [19], but this would have put less emphasis on the referent. A stressed pronoun, in contrast, could not have been used, because then a contrastive reading would have resulted, but contrastiveness does not seem to play a role in these examples. Emphasis might also play a role in other examples with the d-pronoun, but we do not think that the use of the d-pronoun always involves emphasis. Capturing emphasis in a corpus study is very difficult, if possible at all. In contrast to the other properties considered in this paper, it is not a property of the antecedent NP, but a property which derives from the sentence containing the d-pronoun as a whole and cannot be tied to any surface properties of the sentence.

6. General discussion

89 This paper has presented a corpus study and a production experiment investigating the choice between p-pronoun and d-pronoun in German. Our analysis was phrased in terms of accessibility theory (Ariel, 1990), which has been fruitfully applied before to the interpretation of p- and d-pronouns. In contrast to language comprehension, language production requires an absolute notion of accessibility because a choice between p-pronoun and d-pronoun is necessary even when ambiguity is not at issue. The corpus data show that such an absolute notion of accessibility can be defined in terms of givenness and syntactic prominence, which comprises both syntactic function and clausal position. Thus, as claimed by accessibility theory, accessibility cannot be defined in terms of a single property of the antecedent, but must be considered as a complex property.

90 In the remainder of this general discussion, we address two issues that are in need of clarification. The first issue concerns the relationship between givenness and topichood. In our corpus study, we coded whether the antecedent was given or not, but we did not code whether it is a (sentence) topic or not. We made this decision for the following reason. Whereas the givenness of an NP can be determined uncontroversially with few exceptions, determining whether an NP is the topic of a

sentence or not is often not possible in an objective way, as shown by experiments on inter-annotator agreement (e.g., Cook & Bildhauer, 2013). With regard to the choice between p-pronoun and d-pronoun, this issue is of relevance because, as discussed at the beginning, d-pronouns have been claimed to refer to non-topics. Our corpus studies revealed a number of examples where the d-pronoun refers back to a given antecedent, mainly when the antecedent occurred as object in clause-final position. In a similar way, the participants in our production experiment used a d-pronoun to refer to a clause-final object whether the object was given or new.

91 The question thus is whether the antecedent in these cases was not only given but also a topic. A corpus example where this indeed seems to be the case is provided in [20].

[20] [C -5] **Kasimir** ist etwas besonderes. [C -4] **Seine** Eltern geben **ihn** deshalb schweren Herzens in die Obhut von Wieland von Waghals, der als Historiker in der Gespinsterschule arbeitet. [C -3] Weil **Kasimir** noch zu klein für den regulären Schulbetrieb ist, gibt Wieland **ihm** Privatunterricht. [C -2] Und das ist auch bitter nötig: [C -1] Wie sich bald herausstellt, ist wieder ein Ranzenmann unterwegs – und er hat es auf **Kasimir** abgesehen. [T] **Der** hat also gleich zwei schwere Aufgaben: [...].

'[C -5] **Kasimir** is special. [C -4] **His** parents therefore heavy-heartedly left **him**[p-pronoun] in the hands of Wieland von Waghals, who works in the ghosts school as a historian. [C -3] Since **Kasimir** is to young for regular lessons, Wieland teaches **him**[p-pronoun] private lessons. [C -2] And there is desperate need for it: [C -1] It turned out that there is again a satchel man on the run – and he aims for **Kasimir**. [T] **He**[d-pronoun] thus has two difficult tasks: [...].'

(corpus = "DeWaC-9" text = "862984" id = "http://www.aviva-berlin.de")

92 Here, the d-pronoun *der* is used to refer to Kasimir. This referent is mentioned several times in the preceding context, and it is clearly the topic of this discourse. Furthermore, we consider Kasimir also as the sentence topic of the last context sentence [C -1]. Like the preceding sentences, this sentence adds further information about Kasimir.

93 In a similar way, we assume that in our experimental texts (see [12]) the given NP of the final context sentence was the sentence topic. In contrast to the corpus example in [20], this sentence topic is not the discourse topic, which can be taken to be the referent introduced at the beginning (*Maria*). However, the last context sentence does not make an assertion about the discourse topic, but about the sentence topic, which is then taken up again by a pronoun in the continuation sentence [10].

94 A corpus example where a clause-final antecedent is given but arguably not a sentence topic is given in [21].

10. Alternatively, as proposed by a reviewer, the referent introduced in the second context sentence becomes the new discourse topic. In this case, the NP under consideration is not only an aboutness topic but a discourse topic too.

[21] [C -4] Der Zoll hat einen Briten mit 30 Kilo Marihuana im Gepäck festgenommen. [C -3] Der 32-Jährige hatte **seinen Koffer** mit den 30 ein Kilogramm schweren Drogenpäckchen in Johannisburg aufgegeben. [C -2] Auf dem Weg nach Brüssel machte der Flieger in Frankfurt Zwischenstation. [C -1] Beim Überprüfen des Gepäcks schlug ein Spürhund auf **den Koffer** an. [T] **Der** wurde geröntgt und von den Zöllnern anschließend als Falle auf das Gepäckband gestellt.

'[C -4] Costums officers have arrested a British man with 30 kilograms of cannabis in his luggage. [C -3] The 32-year-old had checked in **his suitcase** containing the 30 packages of drugs in Johannesburg. [C -2] On its way to Brussels, the airplane made a stop in Frankfurt. [C -1] When controlling the luggage, a sniffer dog responded to **the suitcase**. [T] **It**[d-pronoun] was X-rayed and then put on the luggage conveyor belt to lay a snare.'

(corpus = "DeWaC-4" text = "307231" id = "http://www.zoll.de")

95 In this example, a suitcase is introduced in context sentence [C -3]. This referent is not mentioned in the next context sentence [C -2]. In context sentence [C -1], where it is mentioned again, the initial phrase can be considered a stage-setting topic.

96 In sum, in cases where a given clause-final object is the antecedent of the d-pronoun, the antecedent can be a topic or not. Given the difficulties of identifying the topic in each case, we refrain from presenting numerical data on this issue and leave it as a question for future research whether the topichood of the antecedent affects the probability of using the p- or the d-pronoun.

97 A final issue that we want to discuss concerns the finding that d-pronouns are used much less often than p-pronouns when establishing pronominal reference. Even when the use of a d-pronoun was highly favored, both the corpus study and the completion experiment showed a preference for the p-pronoun. This strong preference for the p-pronoun seems to be closely tied to the written language. D-pronouns are much more frequent in spoken German than in written German (see Ahrenholz, 2007; Bosch et al., 2007). This does not exclude that the same factors are at work in spoken as in written language. As noted in the preceding section, when we use accessibility in the most straightforward way for predicting the choice between the p- and the d-pronoun, we strongly overpredict the use of d-pronouns with regard to our sample of written texts.

98 The easiest way to adjust the accessibility-based criterion to both modalities would be to introduce a further factor *modality*, which reduces the rate of d-pronouns in written language and increases this rate in spoken language. However, it is probably not sufficient to only differentiate between the two modalities. We further should take into account different text types and their various stylistic characteristics. The use of anaphoric expressions in social media communication or chats, for example, is certainly more strongly orientated towards spoken language than the language we find in scientific texts or in newspapers. It remains a task for future research to

investigate the use of d-pronouns in different discourse settings, and to test whether there are only quantitative or also qualitative differences with regard to the choice between p- and d-pronouns.

References

Abraham, W. 2002. Pronomina im Diskurs: deutsche Personal-und Demonstrativpronomina unter "Zentrierungsperspektive". Grammatische Überlegungen zu einer Teiltheorie der Textkohärenz. *Sprachwissenschaft* 27 (4): 447-491.

Ahrenholz, B. 2007. *Verweise mit Demonstrativa im gesprochenen Deutsch: Grammatik, Zweitspracherwerb und Deutsch als Fremdsprache*. Berlin: De Gruyter.

Ariel, M. 1990. *Accessing Noun-Phrase Antecedents*. London – New York: Routledge.

Ariel, M. 2001. Accessibility Theory: An Overview. In T. Sanders, J. Schilperoord & W. Spooren (eds.), *Text Representation: Linguistic and Psycholinguistic Aspects*. Amsterdam: J. Benjamins: 29-88.

Arnold, J.E. 2010. How Speakers Refer: The Role of Accessibility. *Language and Linguistics Compass* 4 (4): 187-203.

Bader, M. & Portele, Y. 2015. Prominence and Anti-prominence in Pronoun Resolution. Paper presented at the *International Conference "Prominence in Language"* (University of Cologne, 15-17 June 2015).

Baroni, M., Bernardini, S., Ferraresi, A. & Zanchetta, E. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation* 43 (3): 209-226.

Barr, D.J., Levy, R., Scheepers, C. & Tily, H.J. 2013. Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal. *Journal of Memory and Language* 68 (3): 255-278.

Bates, D., Mächler, M., Bolker, B. & Walker, S. 2015. "lme4": Linear Mixed-Effects Models Using Eigen and S4. Available online: http://CRAN.R-project.org/package=lme4.

Bittner, D. & Dery, J.E. 2015. The Information Structural Effects of German P- and D-pronouns in Discourse. In A. Meinunger (ed.), *Byproducts and Side Effects / Nebenprodukte und Nebeneffekte*. ZAS Papers in Linguistics 58. Berlin: Zentrum für Allgemeine Sprachwissenschaft (ZAS): 49-71.

Bosch, P. 2013. Anaphoric Reference by Demonstrative Pronouns in German. Paper presented at the *Workshop on the Impact of Pronominal Form on Interpretation* (University of Tübingen, 15-17 November 2013). Presentation available online: http://cogsci.uni-osnabrueck.de/~pbosch/download/TUE_DPro2013-11-16.pdf.

Bosch, P., Katz, G. & Umbach, C. 2007. The Non-Subject Bias of German Demonstrative Pronouns. In M. Schwarz-Friesel, M. Consten & M. Knees (eds.), *Anaphors in Text: Cognitive, Formal and Applied Approaches to Anaphoric Reference*. Amsterdam: J. Benjamins: 145-164.

Bosch, P., Rozario, T. & Zhao, Y. 2003. Demonstrative Pronouns and Personal Pronouns: German "Der" vs. "Er". In *Proceedings of the 2003 EACL Workshop on the Computational Treatment of Anaphora*. Stroudsburg: Association for Computational Linguistics: 61-68.

Bosch, P. & Umbach, C. 2007. Reference Determination for Demonstrative Pronouns. In D. Bittner & N. Gargarina (eds.), *Intersentential Pronominal Reference in Child and Adult Language*. ZAS Papers in Linguistics 48. Berlin: Zentrum für Allgemeine Sprachwissenschaft (ZAS): 39-51.

Bouma, G. & Hopp, H. 2007. Coreference Preferences for Personal Pronouns in German. In D. Bittner & N. Gargarina (eds.), *Intersentential Pronominal Reference in Child and Adult Language*. ZAS Papers in Linguistics 48. Berlin: Zentrum für Allgemeine Sprachwissenschaft (ZAS): 53-74.

Colonna, S., Schimke, S. & Hemforth, B. 2012. Information Structure Effects on Anaphora Resolution in German and French: A Crosslinguistic Study of Pronoun Resolution. *Linguistics* 50 (5): 991-1013.

Cook, P. & Bildhauer, F. 2013. Identifying "Aboutness Topics": Two Annotation Experiments. *Dialogue and Discourse* 4 (2): 118-141.

Dudenredaktion 2011. *Duden, richtiges und gutes Deutsch: Wörterbuch der sprachlichen Zweifelsfälle*. Duden 9. Mannheim: Dudenverlag.

Ellert, M. 2013. Information Structure Affects the Resolution of the Subject Pronouns "Er" and "Der" in Spoken German Discourse. *Discours* 12: 1-24. Available online: https://discours.revues.org/8756.

Fraurud, K. 1990. Definiteness and the Processing of Noun Phrases in Natural Discourse. *Journal of Semantics* 7 (4): 395-433.

Frey, W. 2004. A Medial Topic Position for German. *Linguistische Berichte* 198: 153-190.

Fukumura, K. & Van Gompel, R. 2010. Choosing Anaphoric Expressions: Do People Take into Account Likelihood of Reference? *Journal of Memory and Language* 62 (1): 52-66.

Givón, T. 1992. The Grammar of Referential Coherence as Mental Processing Instructions. *Linguistics* 30 (1): 5-55.

Grosz, B.J., Joshi, A.K. & Weinstein, S. 1995. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics* 21 (2): 203-225.

Gundel, J.K., Hedberg, N. & Zacharski, R. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69 (2): 274-307.

Hinterwimmer, S. 2014. A Unified Account of the Properties of German Demonstrative Pronouns. In P. Grosz, P. Patel-Grosz & I. Yanovich (eds.), *NELS 40: Proceedings of the Semantics Workshop on Pronouns*. Amherst: GLSA Publications: 61-107.

Kaiser, E. & Trueswell, J. 2008. Interpreting Pronouns and Demonstratives in Finnish: Evidence for a Form-Specific Approach to Reference Resolution. *Language and Cognitive Processes* 23 (5): 709-748.

Kehler, A. & Rohde, H. 2013. A Probabilistic Reconciliation of Coherence-Driven and Centering-Driven Theories of Pronoun Interpretation. *Theoretical Linguistics* 39 (1-2): 1-37.

Lenerz, J. 1992. Zur Syntax der Pronomina im Deutschen. *Sprache und Pragmatik* 29: 1-54.

Rambow, O. 1993. Pragmatic Aspects of Scrambling and Topicalization in German: A Centering Approach. Paper presented at the *Institute for Research in Cognitive Science (IRCS) Workshop on Centering Theory in Naturally-Occurring Discourse* (University of Pennsylvania, 20-28 May 1993).

R Development Core Team 2015. The R Project for Statistical Computing. Vienna: R Foundation for Statistical Computing. Available online: http://www.R-project.org.

Van Bergen, G. & Swart, P. de 2010. Scrambling in Spoken Dutch: Definiteness versus Weight as Determinants of Word Order Variation. *Corpus Linguistics and Linguistic Theory* 6 (2): 267-295.

Wiemer, B. 1996. Die Personalpronomina "er" vs. "der" und ihre textsemantischen Funktionen. *Deutsche Sprache* 24: 71-91.

Wiltschko, M. 1998. On the Syntax and Semantics of (Relative) Pronouns and Determiners. *The Journal of Comparative Germanic Linguistics* 2 (2): 143-181.

Zifonun, G., Hoffmann, L. & Strecker, B. 1997. *Grammatik der deutschen Sprache*. Berlin: De Gruyter.