A corpus-based analysis of pronoun choice in German relative clauses

Patrick Brandt and Eric Fuß
IDS Mannheim

This paper investigates the conditions that govern the choice between the German neuter singular relative pronouns *das* ‘that’ and *was* ‘what’. We show that *das* requires a lexical head noun, while in all other cases *was* is usually the preferred option; therefore, the distribution of *das* and *was* is most successfully captured by an approach that does not treat *was* as an exception but analyzes it as the elsewhere case that applies when the relativizer fails to pick up a lexical gender feature from the head noun. We furthermore show how the non-uniform behavior of different types of nominalized adjectives (positives allow both options, while superlatives trigger *was*) can be attributed to semantic differences rooted in syntactic structure. In particular, we argue that superlatives select *was* due to the presence of a silent counterpart of the quantifier *alles* ‘all’ that is part of the superlative structure.

1. Introduction

In German, relative clauses that modify a nominal element are typically introduced by a so-called *d*-pronoun that inflects for case (assigned in the relative clause) and agrees in gender and number with the head of the relative clause (cf. e.g. Duden 2016, 303):

(1) a. der Mann, **der** schläft
   the man that.masc.nom sleeps
   b. der Mann, **den** Peter getroffen hat
   the man that.masc.acc Peter met has

---

1. An alternative albeit less frequent and stylistically marked option consists in using inflected forms of the wh-determiner *welch*- ‘which’ to introduce relative clauses. *Welch*- has a similar distribution as *d*-forms and is more readily used in the written language.
While the use of *wh*-forms is generally ruled out in connection with lexical head nouns (*‘der Mann, wer… the man, who…’*), certain neuter singular antecedents trigger relativization by means of the *wh*-pronoun *was ‘what’* (cf. e.g. Duden 2016, 1047ff.; see Citko 2004 on light-headed relatives in Polish; Boef 2012, and Broekhuis and Keizer 2012, 407–420 on *wat* in Dutch), compare the following (non-exhaustive) list in (2) and the examples in (3):\(^2\)

\[
(2) \begin{align*}
\text{a.} & \quad \text{indefinites/quantifiers: alles ‘everything’, eines ‘one thing’, vieles ‘many things’ …} \\
\text{b.} & \quad \text{demonstratives: das ‘that’, dasjenige ‘that thing’, dem ‘that.dat’, …} \\
\text{c.} & \quad \text{deadjectival nouns (conversions): das Gute ‘the good (thing)’, das Beste ‘the best’ etc.}
\end{align*}
\]

\[
(3) \begin{align*}
\text{a.} & \quad \text{Alles, was die Zuschauer dort sehen, ist Lug und Trug.} \\
& \quad \text{everything that the spectators see there is lies and deception.} \\
& \quad \text{‘Everything that the spectators see there is lies and deception.’} \\
& \quad \text{(NON13/JAN.07012 Niederösterreichische Nachrichten, 17.01.2013, NÖN} \\
& \quad \text{Großformat, Ressort: Meinungen; PRO & KONTRA)} \\
\text{b.} & \quad \text{Das, was wir machen, ist das, was uns gefällt.} \\
& \quad \text{that what we make is that what us pleases} \\
& \quad \text{‘What we do is what we like.’} \\
& \quad \text{(BRZ07/JUN.06447 Braunschweiger Zeitung, 04.06.2007; „Das, was wir} \\
& \quad \text{machen, ist das, was uns gefällt“)} \\
\text{c.} & \quad \text{Das Beste, was Microsoft heute tun kann, ist, Yahoo zu kaufen.} \\
& \quad \text{the best what Microsoft today do can is Yahoo to buy} \\
& \quad \text{‘The best that Microsoft can do today is to buy Yahoo.’} \\
& \quad \text{(HAZ08/NOV.01608 Hannoversche Allgemeine, 08.11.2008, S. 15;} \\
& \quad \text{Microsoft lässt Yahoo abblitzen)}
\end{align*}
\]

The exceptional use of a *wh*-pronoun as a relativizer raises a set of theoretical as well as empirical issues, two of which figure prominently in this paper. First, we

\[2. \text{But note that in colloquial use, was may occur in connection with neuter singular antecedents as in (1e) (das Auto, was…), cf. e.g. Murelli (2012).}\]
will address the question of whether the alternation between *das* and *was* must eventually be captured in terms of a list of exceptions (as is often the approach taken in traditional (descriptive) works, cf. e.g. Duden 2016, 1047ff.), or whether it is possible to identify a grammatical rule and/or more abstract property that unifies the contexts in (2). In addition, it will be shown that the elements that trigger relativization by means of *was* do not behave uniformly with respect to pronoun choice, giving rise to an amount of linguistic variation that has not been properly appreciated in earlier works. The examples in (4) and (5) show that deadjectival nouns and elements like *etwas* ‘something’ tolerate both *das*- and *was*-relatives and thus differ from both lexical nouns (which select *das*) and determiners/quantifiers (which select *was*).

(4) Denn nicht das Gute, *das* ich will, tue ich, sondern das Böse, *das* ich nicht will, das führe ich aus. [...] Und dass es Kraft braucht, Mut, trotz diesem Widerstreit all das Gute, *was* wir zustande bringen, zu achten, zu verstärken. ‘Since I do not carry out the good things that I want, but rather the evil that I don’t want. [...] And that it requires strength and courage, despite this conflict, to heed and strengthen all the good things that *what* we accomplish.’ 
(SOZ06/FEB.03769 Die Südostschweiz, 18.02.2006; « Gutmensch » – ein Unwort)

(5) Menschen, die einen Vortrag hören, wollen später etwas mit nach Hause nehmen. *Etwas, was* sie neu gelernt haben, *etwas, das* sie bereichert. People who listen to a talk want to take something home later on. Something that *what* they have newly learned, something that enriches them. 
(Bernd Wolfgang Lubbers. 2013. *Das etwas andere Rhetorik-Training oder "Frösche können nicht fliegen"*. Wiesbaden: Gabler Verlag, p. 97)

We will argue that the alternation between *d*- and relative clauses reflects categorial properties of the antecedent: Lexical nouns trigger relativization by means of *das*, while *was* is used as an underspecified default relativizer in contexts where *d*-forms (i.e., *das*) are not licensed. This approach reverts the markedness relations usually assumed for pronoun choice in relative clauses, and builds on ideas in Brandt and Fuß (2014) (see also Behagel 1928 and, more directly, Wiese 2013). Here we go beyond the basic generalization and show that variation between *das* and *was* may be due to (i) structural/subcategorial differences between different types of deadjectival antecedents (only nominalized superlatives trigger *was* across the board), and (ii) semantic/pragmatic factors conforming with more general patterns typical of conversational implicatures (typically in cases where pronoun choice is underdetermined by the grammar). To isolate the various factors governing the alternation between *d*- and *wh*-forms, we iteratively reduce the
domain where variation obtains via subtraction of predictable classes of cases until virtually no free variation remains.

The paper is organized as follows: Section 2 introduces the idea that was functions as a default relativizer in present-day German, developing and modifying earlier theoretical proposals (Brandt and Fuß 2014). Section 3 discusses variation between das vs. was in connection with deadjectival nouns. Section 4 shows that in cases where the choice of relative pronoun is not determined by the grammar, the alternation between das and was can be used to achieve certain semantic/pragmatic effects. Section 5 provides a concluding summary.

2. was as a default relativizer

This section presents a range of basic facts concerning the distribution of das and was as relativizers and an analysis that builds on previous work (Brandt and Fuß 2014) where we argue that the absence/presence of a lexical-nominal antecedent is the most important factor governing pronoun choice in relative clauses with a neuter singular antecedent. This specific approach is supported by corpus studies that investigate the impact of the presence vs. absence of a lexical head noun on pronoun choice in relative clauses, using the COSMAS web-interface to the Deutsches Referenzkorpus (DeReKo, http://www.ids-mannheim.de/cosmas2). Table 1 lists the aggregate results of the relevant searches, which already show that the presence of a lexical head noun overwhelmingly leads to relativization by means of a d-pronoun, whereas was predominates in the absence of a lexically nominal antecedent (see below for some discussion of the exceptional cases).

Even more revealing is Figure 1, which displays the impact of a lexical head noun in combination with selected elements that usually (i.e., if they act as the sole antecedent of the relative clause) trigger the relativizer was (determiner/demonstrative: das ‘the, that’; indefinites/quantifiers: alles ‘all, everything’, vieles ‘many, much’, nichts ‘nothing’; deadjectival nouns: das einzige ‘the only thing’). To pick one example, let’s take a closer look at the numbers given for the quantifying element alles ‘all’. First, it becomes clear that alles by itself quasi-obligatorily leads to a was-relative (compare the third column from the left). However, when a lexical noun is added to the quantifier (e.g., alles Geld/Wasser/Öl ‘all money/water/oil’), relativization by means of a d-pronoun becomes the norm (the amount of das-relatives is marked by light grey).3

---

3. Alles combines with mass nouns (as well as deadjectival nouns, cf. below) but not count nouns; a separate corpus study showed that the mass/count distinction hardly has an impact on the choice between das and was (pace e.g. Curme 1922). See below for marginal cases where was is used in this context.
Table 1. Distribution of das/was dependent on the presence of a lexical head noun (aggregate numbers, $p < 2.2e-16$, Cramér’s $V = 0.965$)

<table>
<thead>
<tr>
<th>Antecedent without N</th>
<th>Das</th>
<th>Was</th>
</tr>
</thead>
<tbody>
<tr>
<td>678</td>
<td>17.006</td>
<td></td>
</tr>
<tr>
<td>Antecedent with N</td>
<td>36.796</td>
<td>152</td>
</tr>
</tbody>
</table>

Figure 1. Impact of the presence/absence of N with elements that require the relativizer was (DeReKo, Connexor-Teilarchiv, June 2014). 4, 5

Based on these findings, Brandt and Fuß (2014) arrive at the following generalization concerning the licensing conditions for a relative $d$-pronoun in the context neuter singular:

(6) Generalization: Relativization by means of $das$

$N_{[\text{neuter singular}]} \rightarrow das$

According to (6), the presence of a lexical head noun is the decisive factor determining the choice between $das$ vs. $was$ in relative clauses. A related insight has already been reached by Behagel (1928:725f.).

In what follows, we will first present an analysis of the findings reported so far before we take a closer look at cases beyond the basic generalization in (6) where $das$ and $was$ covary in one and the same linguistic context.

4. The rare examples of the type $nichts$+$N$ all involve deadjectival nouns (nichts Gutes/Schönes ‘nothing good/beautiful’ etc.), which allow relativization by means of $was$ (see below for details).

5. The distinction between restrictive and non-restrictive relative clauses does not seem to influence the choice between $das$ and $was$ (in contrast to Dutch, cf. Broekhuis & Keizer 2012).
2.1 Analysis: Inherent gender features trigger das

In this section, we present a slightly modified version of the analysis developed in Brandt and Fuß (2014), which shares with the original approach the intuition that the choice between d- and morphology is determined in the course of the syntactic derivation, depending on whether the relativizer enters into an agreement relation with a lexical head noun (Wiese 2013, see also Boef 2012 on Dutch). The updated account, which is fleshed out in the sections immediately below, is based on the following set of core assumptions:

1. (Inherent) gender is the defining characteristic of lexical nouns.
2. The more specified phonological exponent das is used if the relativizer picks up a gender feature via agreement with a lexical nominal antecedent.
3. Elsewhere, was is inserted (as a default relativizer).

2.1.1 Syntax

We take it that relative pronouns (RP) enter into a syntactic agreement relation with the head of the relative clause, which serves to determine the RP’s feature content. Adopting fairly standard assumptions, we take it for granted that RP contains a category feature [D], an operator feature [Op], and a set of phi-features (following a widely adopted practice, features that await valuation in the course of the syntactic derivation are marked as ‘uF’). Accordingly, the feature content of a relative pronoun can be characterized as follows:6

\[
\text{RP} \ [D, \ Op, \ Person, \ Number, \ uCase, \ uGender]
\]

Case is assigned internal to the relative clause, while the gender feature is determined/valued via agreement with the head noun:7

\[
\text{head} \ [CP \ RP; \ [CP \ C \ [VP \ \ldots \ \ldots \ ]] \ \uparrow \ \text{AGREE}]
\]

6. [person] might be left unspecified if it is assumed that third person expresses the absence of positively specified person features (Benveniste 1950, 1966). Number seems to play a special role: The finite verb of the RC agrees in number with the RP, which suggests that the RP is inherently specified for number. However, the RP also agrees in gender and number with the head noun, which suggests that number must be checked by the relevant agreement operation; thus, agreement not only involves feature valuation, but also matching of already valued features.

7. See Zeijlstra (2012, 2013) for the idea that agreement involves a relation between a probe and a higher, c-commanding goal. Cf. Heck & Cuartero (2011) for an alternative mechanism based on downward agree that accomplishes agreement between head noun and relative pronoun/relative clause.
Focusing on the context [neuter singular, nominative/accusative] where variation between \(d\)- and \(w\)-forms obtains (das vs. was), there are two possible outcomes of the syntactic derivation, dependent on whether the RP acquires a gender feature from a lexical head noun.\(^8\)

\[(10)\]  
\begin{align*}
\text{a.} & \quad [D, \text{Op, } -\text{pl, } -\text{obl, } -\text{obj}/+\text{obj, Gender: } \neg \text{masc}, \neg \text{fem}] \\
\text{b.} & \quad [D, \text{Op, } -\text{pl, } -\text{obl, } -\text{obj}/+\text{obj, Gender: } \_\_\_ ]
\end{align*}

We propose that the choice between \(d\)- and pronouns can be accounted for in terms of the difference between the feature structures in (10) in combination with a set of (fairly standard) assumptions concerning the phonological realization of syntactic terminal nodes and the featural specifications of the vocabulary items involved.

### 2.1.2 Spelling out RP

In a realizational model of grammar where bundles of abstract morpho-syntactic features are supplied with phonological exponents post-syntactically (cf. e.g. Halle and Marantz 1993), the distribution of das vs. was can be captured by assuming that \(d\)- and \(w\)-forms are linked to vocabulary items that differ in specificity and compete for realizing the feature structures in (10). More precisely, we propose that the vocabulary items linked to (neuter) \(d\)-forms are generally more specified than those linked to (neuter) forms in that only the former carry gender specifications (and, presumably, a categorial D-feature):

\[(11)\]  
\begin{align*}
\text{a.} & \quad [D, +\text{Op, } -\text{obl, } -\text{masc}, -\text{fem}] \leftrightarrow /das/ \\
\text{b.} & \quad [+\text{Op, } -\text{obl}] \leftrightarrow /w\_s/
\end{align*}

In other words, in addition to structural case (nom/acc), relative das is specified for an operator feature \([\text{Op}],\) a category feature and neuter gender; in contrast, was is a pure operator that carries only a subset of the feature content associated with the competing \(d\)-form das (cf. e.g. Bayer and Brandner 2008, Grewendorf 2012; cf. Brandt and Fuß 2014 for further discussion concerning the role of case). Under the standard assumption that the insertion of phonological exponents is governed

---

\(^8\) The feature structures in (10) assume decomposition of case features (basically following Bierwisch 1967; cf. Blevins 1995 and Wiese 1999 for slightly revised systems) based on the features \([\pm\text{oblique, } \pm\text{object}]:\)

\[(i)\]  
\begin{align*}
\text{a.} & \quad \text{nominative: } [-\text{obl, } -\text{obj}] \\
\text{b.} & \quad \text{accusative: } [-\text{obl, } +\text{obj}] \\
\text{c.} & \quad \text{dative: } [+\text{obl, } +\text{obj}] \\
\text{d.} & \quad \text{genitive: } [+\text{obl, } -\text{obj}] 
\end{align*}
by some form of the Elsewhere Condition (Kiparsky 1973, 1982; Halle 1997), the
distribution of das and was can be correctly described:

(12) a. das Buch, das du liest
    the book that you read

b. alles, was du liest
    all what you read

In cases like (12a), the RC combines with a lexical head noun, from which the RP
receives a gender specification under agreement. Note that the vocabulary items
in (11) are both compatible with the insertion context in (12a). However, the
Elsewhere Condition requires that the more specified exponent be used, leading
to insertion of das. Alternatively, the RC may be merged with a non-lexical ele-
ment such as a determiner/quantifier (presumably of the category D), as in (12b).
Arguably, functional elements unlike lexical elements do not carry gender features
by themselves (cf. below). Accordingly, the RP cannot receive a gender feature in
the course of the syntactic derivation. As a result, das does not match the insertion
context since it requires the presence of valued gender features; the only remain-
ing possibility is the insertion of the pure operator marker was, which is under-
specified for [gender]. This analysis not only captures the distribution of das and
was in attributive relative clauses, but also facilitates a unified explanation of the
fact that other types of RCs that lack an appropriate (overt) nominal antecedent
are introduced by forms. The latter constructions include free relatives, cf. (13),
continuative relative clauses (“weiterführende Relativsätze”), which attach to the
matrix VP or a higher clausal node as in (14), as well as relative clauses referring
to quote-like expressions (translations, in particular) as illustrated in (15):

(13) [Was der Mann auch anpackt], funktioniert.
    what the man ever tackles works
    ‘Whatever the man tackles, works.’
    (HAZ09/AUG.02148 Hannoversche Allgemeine, 14.08.2009)

(14) Wie bei allen anderen Mannschaftssportarten nahmen die Starken
    as with all other team sports took the strong
    Rücksicht auf die Schwächeren, [was den Spass für alle garantierte].
    regards for the weak what the fun for all guaranteed
    (A09/OKT.06424 St. Galler Tagblatt, 23.10.2009, S. 52; Goldener Herbst im
    Simmental)
(15) Von disciplina wird der Begriff discipulus hergeleitet, was so viel wie from disciplina is the notion discipulus derived what so much as Lehrling oder Schüler bedeutet. apprentice or pupil means

(A09/FEB.05129 St. Galler Tagblatt, 18.02.2009, S. 36; Geschichte prägt die Disziplin)

2.1.3 Some further consequences of the analysis: wh-forms vs. d-forms

At this point, let us briefly highlight the findings reached so far and discuss a couple of theoretical implications, focusing on the different licensing conditions for d- and pronouns, and the status of gender. The relativization facts surveyed so far suggest that the insertion of d-forms as relativizers is dependent on the presence of a syntactic agreement relation between the head noun and the RP, which leads to a valuation of [uGender] on the RP and serves to establish strict coreference between these two elements (traditionally marked by coindexing). As a result, gender on d-pronouns is grammatical in nature (i.e., non-interpretable, resulting from agreement). In contrast, forms such as was are used in cases where no agreement relation is established between the RP and the head of the relative clause (i.e., RP’s feature content is not affected by feature valuation in the syntax). More precisely, was, which lacks gender specifications, can be regarded as the spell-out of unvalued gender features. Thus, we assume that neuter gender can result either from negative values for [masculine] and [feminine] or from the absence of gender valuation altogether (cf. e.g. Harley and Ritter 2002). Furthermore, there are reasons to believe that the absence of [gender] is interpreted as [−animate/]

9. Further evidence suggesting that the connection between relative was and its antecedent is less tight than between a d-form and its antecedent comes from the observation that was can be construed with different kinds of antecedents, leading to systematic ambiguities (see also Holler 2005:96):

(i) Adrian hat alles gekauft, was Anton auch hat.
   Adrian has everything bought what Anton also has
   a. ‘Anton has bought everything, too.’ (reference to the matrix predicate/VP)
   b. ‘Adrian has bought everything that Anton already owns.’ (reference to alles)

(ii) Adrian will in die Bretagne fahren, was Anton auch will.
   Adrian wants to go to Brittany go what Anton also wants
   a. ‘Adrian wants to go to Brittany, and Anton also wants that Adrian goes to Brittany’ (reference to the matrix proposition/IP)
   b. ‘Adrian wants to go to Brittany, and Anton wants to go to Brittany, too.’ (reference to the matrix predicate/VP)

To account for the ‘ sloppy’ behavior of wh-relativizers, Brandt & Fuß (2014, 2016) assume that coreference between a wh-pronoun and a possible antecedent is established by pragmatic mechanisms.
human] at the interface to the semantic component. In this way, the lack of an agreement relation that determines gender on the RP is linked to the observation that gender distinctions have a semantic effect with wh-pronouns: The absence of an antecedent with specified gender (and number) features frees up wh-forms to code a semantic difference (as opposed to purely grammatical gender found with d-pronouns), namely, the difference between persons ([+gender], wer) and non-persons (absence of [gender], was), just as in interrogatives (cf. Wiese 2013).

3. Beyond the basic generalization

3.1 Distributive quantifiers: Cases of noun ellipsis

Certain indefinite pronouns or quantifiers like jedes ‘each’, eines ‘one thing’ or keines ‘none’ appear to trigger relativization by means of das independent of the absence/presence of a lexical noun; the dendogram in Figure 2, which is based on a heat map using normed frequencies, clearly shows that regarding the ratio of das/was, they cluster with antecedents featuring a lexical nominal.

Zooming in, we observe that combining a lexical nominal with jedes, eines or keines only very slightly strengthens the already strong tendency of these elements to trigger relativization by means of das, cf. Figure 3.

When we look at the cases of ‘naked’ jedes/eines/keines followed by das, we observe that virtually without exception, the lexical restriction of the quantifier corresponds to an element mentioned in the immediately preceding discourse context as in (16). In other words, jedes/eines/keines anaphorically refer to a discourse-given antecedent.

(16) Ein richtiges Fußballspiel. Keines, das ich nur im Fernsehen anschauen kann […]

‘A real football match. Not one that I can only watch on TV.’

(BRZ10/MAR.05983 Braunschweiger Zeitung, 12.03.2010)

There are different ways of accounting for the corefential interpretation of elements such as keines in examples like (16). In Brandt and Fuß (2014), we regard such examples as cases of noun ellipsis (see also Duden 2016,957); specifically, we argued that jedes/keines/eines always select for a lexical noun that may remain silent to the extent that its reference can be reconstructed from the discourse context, as sketched in (17):

(17) … Ni,… [jedes/eines/keines Ni [SRel das …]]
Figure 2. Dendrogram of selected antecedents based on the ratio of das/was (normed frequencies).

As a reviewer points out, however, the form *keines* featuring strong adjectival inflection is unexpected under the ellipsis account as it differs from the form we observe in connection with an overt head noun (*keines Fußballspiel vs. kein Fußballspiel ‘no football match’). This might suggest that no ellipsis takes place but that *keines* itself is (pro)nominal. Interestingly, it is exactly when a noun is locally invisible that the quantifier carries strong inflection, e.g. in connection with so-called split noun phrases where the quantifier is stranded by the fronted head noun:

    money have I none anymore

As a general rule, the inflectional features of noun phrases have to be marked at least once (Duden 2016, 955); marking can appear on the article, an adjective, or
Figure 3. $D$-elements that always trigger $das$ (DeReKo, Connexor-Teilarchiv, June 2014).

the head noun. Only if the noun phrase is realized discontinuously can the strong marking appear twice:

    green-nom.sg.masc asparagus is none-nom.sg.masc anymore there  
    ‘There is no green asparagus anymore.’

b. Ist [kein-∅/*-er grün-er Spargel] mehr da?  
    is no.nom.sg.masc green-nom.sg.masc asparagus anymore there

To the extent that the local absence of a noun is compensated for by the strong inflection on the quantifier, the quantifier itself may assume a pronominal function. With forms like kein-, ein- or jed-, the singular endings (-er ‘masc.’ -es ‘neut.,
-e ‘fem.’) unambiguously signal gender, which is the defining characteristic of lexical nouns. In contrast, all- does not signal gender distinctions in the same manner as there is only a single neuter singular ‘pronominal’ form (all-er ‘all-nom.sg.masc’ or all-e ‘all-nom.sg.fem’ occur only together with lexical nouns). As is well-known, gender agreement is a necessary ingredient to syntactically determined anaphoric reference. *Alles* (and *nichts* ‘nothing’) cannot serve as an anaphor since it lacks gender distinctions (but is always neuter). In contrast, jedes/eines/keines exhibit the full gender paradigm; these elements may thus be identified with a nominal antecedent. We thus expect relativization by means of $d$-forms no matter where the gender features come from (from a silent/elided noun under agreement or directly from the lexicon).

In other respects as well, *alles* lacks the specifically nominal quality of *eines* or *keines*; note, thus, that it at best very marginally allows for (post-nominal) attributive genitives, which regularly attach to N only.
3.2 Deadjectival nouns

As we noted above, deadjectival nouns that are derived by conversion seem to be compatible with both das and was, cf. the following statement in the Duden grammar (2016, 1049): “Die Form was steht überwiegend bei Bezug auf substantivierte Adjektive mit Genus Neutrum. Dabei kommt ebenfalls die Form das vor” [The form was ‘what’ prevails with reference to nominalized adjectives in the neuter gender. The form das ‘that’ is also possible. our translation]. However, a corpus study carried out in the DeReKo (W-Gesamt, November 2014) shows that the situation is actually even more complex. We propose that three classes of cases must be distinguished along the lines in (21).

(21) Distribution of das/was with deadjectival nouns (neuter singular):
   a. anaphoric/elliptical readings ⇒ das
   b. (non-anaphoric) nominalized positives (das Gute ‘the good (one)’) ⇒ das & was
   c. (non-anaphoric) nominalized superlatives (das Beste ‘the best’): ⇒ was

These observations bear on the more general theoretical question whether ‘nominalized’ adjectives come with an elided N-head or are categorically nominal themselves (or may instantiate a mix of both options), cf. e.g. Kester (1996) vs. Sleeman (2013). In the following subsections, we turn to the respective classes one by one.

3.2.1 Anaphoric/elliptical readings

To the extent that a nominal restriction can be reconstructed from the context, no variation is observed and das is used without exception; (22) gives another relevant corpus example.

(22) Das bisher bestehende Kraftwerk ist bekanntlich völlig überaltet […] Das neue, das Ende 1994 fertiggestellt sein soll, wird nach dem neuesten Stand der Technik errichtet.

As is well-known, the existing power plant is completely outdated […] The new one, which will be completely by the end of 1994, will be built using the latest technology.

(N93/MAR.08195 Salzburger Nachrichten, 05.03.1993; Heizkraftwerk Nord 88 Mill. S teurer)
Like in the case of *jedes/eines/keines* discussed above, we propose that the use of *das* is triggered here by a silent lexical head noun, as sketched again in (23).

(23) … *Kraftwerk… [Das neue *Kraftwerk* [SRel das …]]*

Elliptical readings thus do not present relevant cases altogether: in the present context, the adjectives accompany lexical if silent head nouns that trigger relativization by means of *das*.

### 3.2.2 Non-elliptical readings: Nominalized positives

We do observe variation in the context of non-elliptical readings of nominalized positives, cf. (24a) vs. (24b).

(24) a. oft sei es *das Neue, was* den Anstoß gebe, eingefahrene Fahrrasser endlich zu verlassen

‘Often it is the new [ideas] that [what] motivate people to finally leave the beaten track.’

(RHZ97/JUL.03252 Rhein-Zeitung, 05.07.1997; Gute Ideen sind in der Tat gefragt)

b. *Das Neue, das* sich aus solchen Veränderungen ergeben könnte, ist vielerorts allerdings noch nicht in Sicht.

‘In many places, the new [things] that could result from such changes are not visible yet.’

(T92/JUN.24015 die tageszeitung, 24.06.1992, S. 16; Zensur in der neuen Weltunordnung)

The quantitative analysis shows that *das* is actually more frequent than *was* in this context:

We propose to attribute the observed variation to the categorial status of nominalized positives that can be characterized as a mix of nominal as well as adjectival properties. On the one hand, nominalized adjectives have a distribution typical of nouns; specifically, they combine with determiners or determiner-like elements, cf. (25), and they allow adjectival modification exhibiting the agreement patterns otherwise found with nominal heads, cf. (26).

(25) a. *das/viele/alles Gute*

    the/much/all  good (weak inflection)

b. *ein/viel/nichts Gutes*

    a/much/nothing good (strong inflection)

(26) a. *das vermeintliche/einzige/vollständige Neue*

    the alleged/only/complete  new (one)

b. *das vermeintliche/einzige/vollständige Opfer*

    the alleged/only/complete  victim/sacrifice
Table 2. *das* vs. *was* with nominalized positives (non-elliptical readings)

<table>
<thead>
<tr>
<th></th>
<th>Das</th>
<th>Was</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Gute(s)</em> 'good'</td>
<td>490 (74%)</td>
<td>172 (26%)</td>
</tr>
<tr>
<td><em>Schöne(s)</em> 'beautiful'</td>
<td>132 (73.7%)</td>
<td>47 (26.3%)</td>
</tr>
<tr>
<td><em>Neue(s)</em> 'new'</td>
<td>502 (74.2%)</td>
<td>175 (25.8%)</td>
</tr>
</tbody>
</table>

On the other hand, nominalized positives show adjectival behavior. They alternate between strong and weak inflection, cf. again (25), they allow adverbial modification cf. (27), and they retain the case-assigning properties of the adjectives that form their base, cf. (28) with dative case on the nominal complement.

(27) a. das sehr/vermeintlich/einzig/vollständig Neue
     the very/allegedly/only/completely new (one)

     b. *das sehr/vermeintlich/einzig/vollständig Opfer
     the very/alleged/only/complete victim/sacrifice

(28) die dem König Treuen
     the.PL the.DAT king faithful.PL

In sum, the variation between *das* and *was* in the case of nominalized adjectives likely is a consequence of their mixed lexical status between nouns and adjectives (cf. Fuß 2017 for further discussion). This approach is in line with proposals by Sleeman (2013) or Alexiadou (2011, 2015) who argue that the ‘mixed’ categorial properties of certain deadjectival nouns follow from the presence of both nominal and adjectival functional structure in the internal make-up of these forms (giving rise to the impression of different degrees of ‘nouniness’ at the surface). Alternative analyses that posit the presence of silent head noun (cf. e.g. Kester 1996a,b) face a number of problems. For example, they cannot account for the relativization facts (if a silent noun were always present, we would expect relativization by means of *das* across the board), and do not provide an explanation for the observation that in contrast to lexical nouns, gender is not merely a grammatical category, but has a semantic effect: neuter forms refer to abstract, non-animate entities, while masc./fem. forms refer to persons:

(29) das Schöne vs. der/die Schöne
     the.NEUT beautiful     the.MASC/ the.FEM beautiful (= a beautiful person)

3.2.3 Non-elliptical readings: Nominalized superlatives

Finally, nominalized superlatives overwhelmingly trigger relativization by means of *was* (in non-elliptical readings), cf. the example in (30) and the figures in Table 3 (cf. as well Cutting 1902).
Abschließend gibt Angela Merkel ihrer Überzeugung Ausdruck, „dass der Blick von oben auf die Erde das Schönste ist, was es gibt.”

‘Concluding, A.M. expresses her conviction that the view of earth from above is the most beautiful [thing] that [what] exists.’

(B06/JUL.52314 Berliner Zeitung, 21.07.2006, Ressort: Blickpunkt; „Wie klappt’s mit dem Schlafen?”)

<table>
<thead>
<tr>
<th>Table 3. das vs. was with nominalized superlatives (non-elliptical readings)</th>
<th>Das</th>
<th>Was</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bestes ‘best’</td>
<td>413 (6.6%)</td>
<td>5838 (93.4%)</td>
</tr>
<tr>
<td>Schönste ‘most beautiful’</td>
<td>86 (9.9%)</td>
<td>783 (90.1%)</td>
</tr>
<tr>
<td>Neueste ‘newest’</td>
<td>11 (16.9%)</td>
<td>54 (83.1%)</td>
</tr>
</tbody>
</table>

Following Bobaljik (2012), we assume that superlatives decompose into a part meaning ’more’ (= a comparative head) and a component meaning (roughly) ’than all (others)’ (= a ’superlative’ head). The latter provides a standard of comparison containing a universal quantifier. The meaning of a superlative can then be described as in (31), with X a property.

(31) more X than all others (X = a property)

On the basis of this structure, the use of was can be attributed to the presence of a universal quantifier in the internal structure of superlatives. Specifically, the strong preference for was in connection with nominalized superlatives follows if the relative clause standardly modifies the universal quantifier located in the superlative category (recall that alles (almost) obligatorily selects was as a relativizer, cf. Section 2. above) as illustrated by means of the paraphrase in (32).

(32) das Beste, was ich kenne = ’etwas, das besser ist als [alles (andere), the best I know, some X which is better than all (others) [was ich kenne]]’

In (32), the relative clause provides the lexical restriction for the universal quantifier. Asking why the relative clause attaches to the universal quantifier (and not to NP/DP), we note that higher attachment (to NP or DP) does not appear to yield

10. In languages like Russian, this meaning is transparently coded in the syntax (Bobaljik 2012:61):

(i) positive: comparative: superlative:
    xoros-ij   luč-še   luč-še   vse-go/-x
    good-MASK.SG better-CMPR better-CMPR all-GEN.SG/GEN.PL
the correct interpretation. For instance, in a construction like *the best that I know*, the relative clause does not serve to single out a best thing from a set of other best things, but rather identifies a best element (i.e., the element which exhibits a certain property to the highest degree) in a set of elements defined or delimited by the (predicate expressed by the) relative clause.\footnote{Note that this analysis does not predict that [+masc]/[+fem] nominalizations such as *der Beste* ‘the best.masc’ trigger relativization by means of *wh*-forms:}

\begin{enumerate}
\item *der Beste, \quad* den ich kenne
\textquoteleft the best.masc that I know\textquoteright
\end{enumerate}

In cases like (i), the lexical restriction of the universal quantifier is furnished by the feature [+person] on the deadjectival noun. Accordingly, examples like (i) can be paraphrased as in (ii). Note as well that the most nearby natural language paraphrase of (ii) in German contains the masculine singular quantifier *jeder* ‘each person’ which obligatory triggers relativization by means of *d*-pronouns.

\begin{enumerate}
\item X is better than anybody I know.
\end{enumerate}

(33) Of all things that I know, X is the best.

According to this analysis, superlatives contain more functional structure characteristic of adjectival elements and are therefore less ‘nominal’ than nominalized positives; note that the presence of superlative morphology is generally taken to be a defining property of adjectives.

4. Semantics/pragmatics and choice between *das* vs. *was*

Beyond the basic generalization that a lexically nominal antecedent triggers relativization by means of *das*, semantic/pragmatic factors can be observed to influence the choice of relativizer under specific circumstances. In cases not already determined by the basic rule, it appears that the choice of *das* vs. *was* may depend on the ontological status of what the antecedent refers to; in a similar vein, we can observe that the use of *das* where there is no nominal antecedent present may have an individuating effect while use of *was* can be characterized as having an overall totalizing effect.

4.1 Reference to objects as opposed to properties or propositional meanings

In (34), reference of *was* in the so-called predicate-nominal function is not to concrete individuals falling under the nominal restriction *Kleinkind* ‘toddler’ but to
the corresponding property; the former interpretation would be nonsensical, in fact.

(34) Wie soll sich ein Kleinkind, was 4-jährige sind, über einen so langen Zeitraum konzentrieren?
    How should a toddler, [which is] what 4-year-olds are, concentrate over such a long period?
    (A08/OKT.08813 St. Galler Tagblatt, 29.10.2008, S. 31; Kinder werden überfordert)

A corpus research in KoGraDB showed an altogether balanced frequency of *das* and *was* respectively as relativizers of *etwas*. *Etwas* can refer to both (particular) individuals as well as to properties or propositions; a close manual investigation revealed that *was* prototypically serves to pick up reference to properties or propositional meanings, while *das* serves to pick up individual reference. (35) gives corpus examples for reference to properties (viz. predicates, (35a)) or propositions (35b).

(35) a. Ich erfuhr, wie wichtig es ist, die Hypotheken der Vergangenheit anzuerkennen, überhaupt einen Gaumen für die Bedeutung von Vergangenheit zu zeigen – *etwas, was* mein literarisches Vorbild nicht tat, nicht tun konnte.
    'I learned how important it is to appreciate the mortgages of the past, to show a palate for the meaning of the past – something that [what] my literary idol neither did nor could do.'

b. Also war da *etwas, was* ich nicht wissen soll?
    'So, was there something that [what] I should not know?'
    (Rolf F. Schütt: Die Irren sind auch nicht mehr die einzig Normalen. Oberhausen, 1997)

The raw numbers of the survey are given in Table 4; statistical evaluation reveals that selection of *das* vs. *was* depends in a manner that is highly significant on the way of referring \( (p = 1.774 \times 10^{-7}, \text{Cramer's } V = 0.304) \).

<table>
<thead>
<tr>
<th></th>
<th>Etwas, das</th>
<th>Etwas, was</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of cases</td>
<td>141</td>
<td>154</td>
</tr>
<tr>
<td>Property/proposition reference</td>
<td>12</td>
<td>53</td>
</tr>
<tr>
<td>object reference</td>
<td>129</td>
<td>101</td>
</tr>
</tbody>
</table>
Antecedents like *dasselbe* or *das gleiche* ‘the same’ that are ambiguous in the same way as *etwas* lead to analogous results, supporting the conclusion that *das* ranges over individuals while *was* ranges over properties.

4.2 Particularization by means of *das* vs. totalization by means of *was*

According to our analysis, *das* and *was* form what is known as a ‘Horn-scale’ as the possible interpretations of *was* subsume those of *das*. Accordingly, use of *was* may give rise to a conversational implicature that the speaker is not in a position to assert the stronger (more specific, more exclusive) meaning that is associated with *das* (cf. e.g. Horn 2000, 306). The effect is illustrated in the following contrasting examples. In (36), the reference of *das, das* is restricted to pictures, unlike the reference of *das, was* in (37) that is less restricted.¹²

(36) Das Bild, das Peter gekauft hat, war teurer als das, das Maria gekauft hat.
   ‘The picture that Peter bought was more expensive than the picture that Mary bought.’

(37) Das Bild, das Peter gekauft hat, war teurer als das, was Maria gekauft hat.
   ‘The picture that Peter bought was more expensive than the thing(s)/stuff that Mary bought’

(36’) and (37’) sketch the analysis of the contrast in the by now familiar terms: in (36), the relative clause modifies a nominal antecedent, while in (37) it combines directly with the determiner element.

(36’) ... als [das Bild, [das Maria gekauft hat]]

(37’) ... als [das, [was Maria gekauft hat]]

In support, it can be observed that the use of *das* has an individualizing/particularizing effect on the interpretation of a nominalized adjective (Sanders 1879, 279f., Cutting 1902, Curme 1922, Behaghel 1928). In contrast, the use of *was* has a generalizing effect, leading to an interpretation of the deadjectival nouns in terms of a general property, or a totality/collection of things. Sanders (1879, 279) gives the example in (38), commenting that *das Böse* ‘the evil’ signifies a whole class in general, a totality of rigors (= alles (das) Böse ‘all (that) evil’), and therefore triggers the generalizing *was* as relative pronoun. In contrast, *das eine Gute* ‘the single good (thing)’ is a certain good deed, which is reflected by the particularizing relative pronoun *das* or *welches*.

---

¹² Duden (2009, 1049) singles out the former as an elliptical use, see also (16) above.
(38) Ich verzeihe ihm das Böse, was er mir zugefügt hat und künftig noch zufügen wird, in dankbarer Erinnerung an das eine Gute, das (welches) er mir früher erwiesen hat.
'I forgive him the evil [deeds] that [what] he did an will do to me in grateful memory of the single good [thing] that (which) he did to me earlier.'

Corroborating these observations, adding the totalizing element alles 'all, every' to a nominalized positive significantly increases the rate of relativization by means of was; (39) gives a representative example; the numbers in Table 5 show that the addition of alles to a nominalized adjective correlates with increased use of the relativizer was (cf. again Section 2 above as well as Section 3.2.3)

(39) Für den Mundart-Dichter verkörpern die Früchte alles Gute, was einem im Leben so passieren kann. Denn Zwetschgen bedeuten Glückseligkeit.
To the dialect poet the fruit represent all good [things] that [what] may happen to you in life, since plums mean happiness.'

(NUN12/NOV.00444 Nürnberger Nachrichten, 05.11.2012, S. 37; Bald regnet es Zwedschgä – Dialekt-Dichter Fitzgerald Kusz stellt seinen neuen Lyrikband vor)

Table 5. Impact of alles 'all' on the relative frequency of das vs. was with nominalized positives (p < 2.2e–16, Cramér's V = 0.344)

<table>
<thead>
<tr>
<th></th>
<th>Das</th>
<th>was</th>
</tr>
</thead>
<tbody>
<tr>
<td>(das, vieles) Gute/Schöne/Neue</td>
<td>681 (84.6%)</td>
<td>124 (15.4%)</td>
</tr>
<tr>
<td>alles Gute/Schöne/Neue</td>
<td>96 (56.5%)</td>
<td>74 (43.5%)</td>
</tr>
</tbody>
</table>

Interestingly, a similar effect can be observed in connection with mass nouns, which in more of 99% of the cases require relativization by means of das (cf. footnote 4 above). However, when the mass noun is modified by the universal quantifier alles (e.g., alles Wasser/Geld etc. 'all water/money' etc.), the frequency of was drops to 80%:

Table 6. Impact of alles 'all' on the relative frequency of das vs. was with mass nouns (p < 2.885e–16, Cramér's V = 0.262)

<table>
<thead>
<tr>
<th></th>
<th>Das</th>
<th>was</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geld, Wasser etc.</td>
<td>5,297 (99.3%)</td>
<td>38 (0.7%)</td>
</tr>
<tr>
<td>alles Geld, Wasser etc.</td>
<td>329 (80%)</td>
<td>82 (20%)</td>
</tr>
</tbody>
</table>

Interestingly, and showing the connection to superlatives again, modification by einzig 'only' as implying uniqueness yields an analogous effect.
That the qualitative difference between *das* and *was* can be used by the speaker to achieve certain communicative effects was noted as well by Curme (1922, 199), who similarly points to something like an individuating or particularizing effect associated with the use of *das*:

> It is possible that there is often here an intentional use of *das* or *welches* by way of differentiation, to refer to something definite, definite at least to the speaker. […] We cannot, however, in many cases on account of the lack of clearness in the thought absolutely determine whether the *das* or *welches* is used merely as a survival of older usage to indicate a collective idea or something indefinite or general, or is employed intentionally in accordance with modern usage elsewhere, to refer to something definite.

5. **Concluding summary**

We have argued that the alternation between the relativizers *das* and *was* reflects categorial properties of the antecedent of the relative clause (Brandt and Fuß 2014):

- *das* is inserted in the presence of a lexical head noun (characterized by specified gender features on *N*)
- *was* is the underspecified elsewhere case

Cases where *das* is used in the absence of an overt nominal typically turn out to be cases of ellipsis or can be attributed to the presence of lexical gender features on the head of the relative clause as in the case of certain indefinites and quantifiers (e.g. *jedes* ‘each’, *keines* ‘none’).

Regarding deadjectival nouns, we noted that three cases must be distinguished: For one, adjectives that modify silent (viz. elided) nouns go with *das* no matter what. Second, nominalized positives allow both *das* and *was* due to their mixed categorial status. On top of this, the distinction between *d*- and *wh*-morphology can be used to express subtle semantic distinctions (e.g., individuation using *das*). Third, nominalized superlatives select *was* which according to the analysis presented combines with a silent counterpart of the quantifier *alles* that is part of the superlative structure.

We have reviewed evidence that beyond the basic generalization, the variation between *das* vs. *was* points to semantic and pragmatic distinctions related to the question of what it takes to be a noun. In the domain investigated here, the crucial distinction is that between ‘ordinary individuals’ going with *das* and everything else, in particular, property instantiations and propositional meanings going with *was*. 

All rights reserved
Acknowledgements

We wish to thank two anonymous reviewers for their insightful and helpful comments as well as the audience of NWASV2. In addition, we want to express our gratitude to a number of colleagues at the Institut für Deutsche Sprache in Mannheim and elsewhere: Hardarik Blühdorn, Fabian Heck, Franziska Münzberg, Anna Volodina, Angelika Wöllstein, and especially Bernd Wiese.

References

Brandt, Patrick, and Eric Fuß. 2014. “Most questionable pronouns: Variation between das- vs. was-relatives in German”. Linguistische Berichte 239, 297–329.
Brandt, Patrick, and Eric Fuß. 2016. Relativpronomenselektion und grammatische Variation: was vs das. in attributiven Relativsätzen. Ms., IDS Mannheim.


Authors’ addresses

Patrick Brandt
Institut für Deutsche Sprache
R5 6-13
68161 Mannheim
Germany
brandt@ids-mannheim.de

Eric Fuß
Institut für Deutsche Sprache
R5 6-13
68161 Mannheim
Germany
fuss@ids-mannheim.de