

Less-comparatives must be less ambiguous than exactly-differentials, experimental data shows

Fabian Schlotterbeck & Polina Berezovskaya*

Abstract. Scope mobility of comparative operators has been claimed to surface in a narrow class of specific cases where intensional verbs are combined with *less-comparatives* or *exactly-differentials*. Though not uncontroversial and dependent on subtle judgments, this type of ambiguity influenced subsequent compositional semantic analyses of comparatives and was also used as a diagnostics for scope mobility of the comparative operator in cross-linguistic studies. We use judgment data from three acceptability rating experiments to empirically test the (un)availability of this ambiguity in German and English. We discover an empirical difference between *exactly-differentials* and *less-comparatives* which is unexpected under the standard approach to the semantics of comparatives. We discuss the theoretical implications of our findings and highlight recent proposals that can account for our data.

Keywords. scope ambiguity; degree semantics; comparatives; modals; acceptability judgments

1. Introduction. It has been a little over twenty years since the seminal paper by Heim (2000, following the standard degree theory by Stechow 1984) broached the issue of degree operators and scope. According to Heim (2000) and contrary to previous conclusions, e.g. from the comprehensive analysis of Kennedy (1997), scope mobility of comparative operators surfaces in a narrow class of specific cases where intensional verbs are combined with *less-comparatives* or *exactly-differentials*, as in (1). According to this view, (1) has the two readings in (1-a/b). Its less prominent, inverse scope reading in (1-b) imposes no upper limit on the paper’s length, but only a minimal requirement of 15 pp.

(1) (This draft is 10 pages.) The paper is required to be exactly 5 pages longer than that.

- a. linear scope: $\forall w \in Acc : \text{MAX}\{d : \text{long}_w(p, d)\} = 15pp$
‘It is required of the paper that it is exactly 15pp long.’
- b. inverse scope: $\text{MAX}\{d : \forall w \in Acc : \text{long}_w(p, d)\} = 15pp$
‘The minimum length required for the paper is exactly 15 pages.’
(where *Acc* is the set of accessible worlds)

Though not uncontroversial and dependent on notoriously subtle judgments, this type of ambiguity influenced subsequent compositional semantic analyses (e.g., Bhatt & Pancheva 2004, Breakstone et al. 2011, Lassiter 2012, Beck 2012a, a.m.o.) and also served as a diagnostics for scope mobility of the comparative operator in cross-linguistic research (e.g., Beck et al. 2004, 2009).

*This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 75650358. We would like to thank Sigrid Beck, Doris Penka and Paula Menéndez Benito for valuable feedback on this project. We would also like to thank Tamara Stoyke and Robin Edds for their assistance in generating English materials and Hening Wang for support with the application for ethics approval. Last but not least, we thank anonymous reviewers for comments on earlier versions of this paper. Affiliation: University of Tübingen; Email: {fabian.schlotterbeck, polina.berezovskaya}@uni-tuebingen.de).

In the present paper, we investigate the empirical status of this rather subtle ambiguity for German and English. In particular, we systematically collected judgments regarding the availability of the inverse scope reading in three acceptability rating experiments. This embeds our paper in the more general efforts of strengthening the database for theoretical linguistics (for discussion see, e.g., Gibson & Fedorenko 2013, Gibson et al. 2012, Sprouse & Almeida 2013).

2. Theoretical Background. It is common in the semantic literature to analyze degree operators like the comparative, but also their class mates superlatives, equatives or measure phrases, etc., to be quantifiers over degrees (Stechow 1984, Heim 1985, 2000, Beck 2011; a.m.o.). However, we find far less scope interaction than expected between these degree operators and other scope-bearing elements. This led alternative analyses, such as Kennedy (1997), to take a non-quantificational approach, in which the comparative is not a quantifier and cannot undergo quantifier raising (QR). In response, Heim (2000) explains the scarcity of scope interactions involving degree phrases (DegPs) with a number of independently motivated constraints (for further discussion on the constraints of quantifier movement of comparatives, cf. Beck 2011; 1363). With Stateva (2000), Heim (2000) concludes that inverse scope can be detected when intensional verbs (like *require*, *need* or *allow*) scopally interact with non-monotone (e.g. *exactly*-differentials; as in (1)) or downward-monotone (e.g. *less*-comparatives; as in (2)) DegPs.

(2) (**Context:** The draft is 10 pages.) The paper is required to be less long than that.

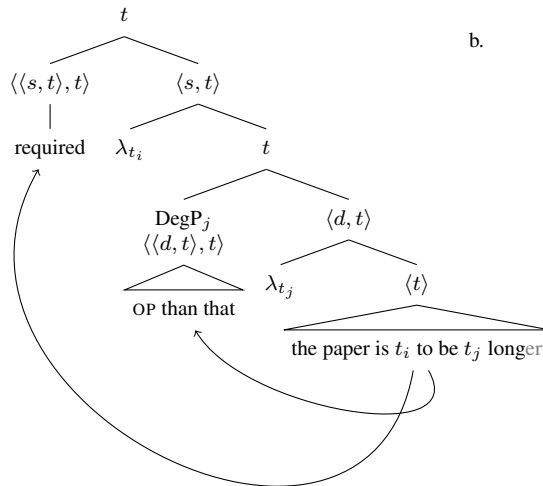
- a. $\forall w \in Acc : \text{MAX}(\lambda d. \text{the paper is } d\text{-long in } w) < 10\text{pp}$
‘It is required of the paper that it is less than 10 pages long (and not longer).’
- b. $\text{MAX}(\lambda d. \forall w \in Acc : \text{the paper is } d\text{-long in } w) < 10\text{pp}$
‘The minimum length required for the paper is less than 10 pages.’

We illustrate the derivation of both readings in a QR-based approach in (3).¹ Under the inverse scope reading, the comparative operator takes scope above the modal verb. For the *exactly*-differential in (1), this leads to a reading where no upper limit is imposed on the paper’s length but only a minimal requirement of 15 pp in our example. The same general point can be made regarding *less*-comparatives as in (3). From a theoretical point of view (via Heim 2000), no difference is thus expected with respect to the investigated ambiguity when comparing *less*-comparatives and *exactly*-differentials in English. The same goes for German, as discussed in the next section.

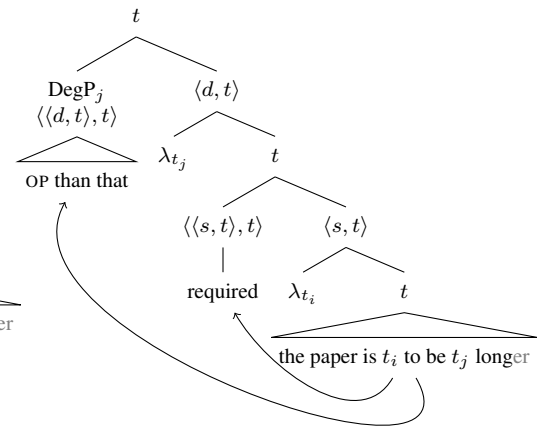
2.1. IMPORTANCE OF THE AMBIGUITY IN CROSS-LINGUISTIC RESEARCH. Under the assumption that DegPs are generalized quantifiers over degrees, the illustrated ambiguity is used in Beck et al. (2004, 2009) to argue for the existence of degree abstraction cross-linguistically. An additional diagnostic are negative island effects. German is on a par with English in displaying both, the purported ambiguity and negative island effects. It has therefore been taken to have the positive setting of the Degree Abstraction Parameter (DAP). The parameter asks whether any given language has binding of degree variables in the syntax. Essentially, the question is whether Logical Forms with the following constellation exist: [DegP_{<<d,t>,t>} [$\lambda d. [...t_d...]$]].

¹In the Logical Forms, we use intensional meanings, i.e. propositions of type $\langle s, t \rangle$ where needed. Modals like ‘require’ are treated as quantifiers over possible worlds. These assumptions do not have an impact on our subject matter.

(3) a.



b.



(OP \in {exactly 5pp COMP, LESS})

It is worth mentioning that there is a substantial body of work on degree constructions in a range of individual languages spurred by this line of research. The suggested diagnostics have been used for cross-linguistic (field)work and L1 acquisition in a range of languages (see e.g. Bochnak 2015 on Washo; Bowler 2016 on Warlpiri; Hohaus et al. 2014 for L1 acquisition of German and English; Howell (2012) on Yorùbá; Berezovskaya (2014) on L1 acquisition in Russian; and Kapitonov (2019) on Kunbarlang).

Among the 17 languages from different language families investigated by Beck et al. (2009), Motu, Japanese, Chinese, Mooré, Samoan and Yorùbá were determined to have the negative setting of the DAP-parameter. German patterned with English, Bulgarian and Hindi-Urdu in terms of the positive parameter setting of the DAP. Here, no differences are expected between English and German with respect to our protagonist, the modal-comparative ambiguity. Combining this with prominent theoretical proposals discussed above, we can thus derive the prediction for the current experiments that the inverse scope reading should be available in German and English for both *exactly*-differentials and *less*-comparatives. A limitation of the cross-linguistic research by Beck et al. (2009) for our current purpose is, however, that their questionnaires included few items and no inferential statistics were reported regarding the comparison between *exactly*-differentials and *less*-comparatives or a baseline condition. Thus, we cannot draw strong conclusions about the existence of the inverse scope reading from this research. Interestingly, recent work by Philipp & Zimmermann (2020, 2023) points to a gradual, non-categorical difference between English and German with regard to scope ambiguity in nominal quantification, with higher acceptance of inverse scope in English. From this perspective, it is an interesting question whether there is a difference in the scope potential of the comparative operator between English and German as well.

3. Experiments on German. In order to test for the availability of inverse scope readings in sentences like (1) and (2) in German, we conducted two web-based questionnaire studies.

3.1. EXPERIMENT 1. The aim of Exp. 1 was to compare judgments indicating linear or inverse scope readings of sentences like (1) and (2) to a baseline obtained from judgments for simple unambiguous control conditions.

3.1.1. METHODS.

Materials, design, procedure & participants: Twelve German items were constructed as exemplified in (4) and (5). All items start with a sentence in which gradable adjectives (e.g. *lang*, ‘long’) are degree-modified by *exactly*-differentials (e.g. *genau 10 Seiten länger als*, ‘exactly 10 pages longer than ...’) or *less*-comparatives (e.g. *weniger lang als*, ‘less long than ...’).² Gradable properties of referents (e.g. ‘the draft’ and ‘the paper’) that are introduced by definite descriptions are compared. In half of the conditions (e.g. (4-a/c)), comparatives are combined with the modal verb *müssen* (‘must’). By hypothesis, the presence of ‘must’ leads to the purported ambiguity. Sentences without modals (e.g. (4-b/d)) were used as unambiguous controls against which responses to the modal conditions can be compared (see predictions below).

The controls were kept similar to the target conditions regarding aspects like syntactic structure, lexical material, length, plausibility, etc. At the same time, we kept them as simple as possible in order to maximize our chances of detecting the ambiguity. The rationale behind this was that comprehension difficulty and ambiguity may have indistinguishable effects on judgments. In particular, ambiguous as well as difficult sentences may lead to less clear-cut judgments than sentences that are easier to understand or unambiguous, respectively. If we found no indication of the ambiguity even when comparing the relatively complex test sentences to very simple controls (i.e. if response categories are differentiated to the same degree in both cases), we would thus have rather strong evidence against the purported ambiguity in the tested type of construction. To keep the *exactly*-controls as easy to understand as possible, we removed not only the modal, but also the comparative morphology from them. The reason was that previous studies found increased processing difficulty of comparative vs. bare forms of the adjective (e.g. Agmon et al. 2019). This was not possible for the *less*-controls because they are inherently comparative, i.e. they need to have a standard of comparison. Sentences in all conditions are followed by a short post-context sentence.

(4) Target sentences and post contexts

- a. Das Papier muss genau 10 Seiten länger sein als der Entwurf. So lautet die Vorgabe der Zeitschrift.
the paper must exactly 10 pages longer be than the draft. So sounds the guideline of the journal
‘The paper is required to be 10 pages longer than the draft. That’s what the journal’s guideline says.’
- b. Das Papier ist genau 10 Seiten lang. Das haben die Autoren gesagt.
the paper is exactly 10 pages long. That have the authors said
‘The paper is exactly 10 pages long. That’s what the authors said.’
- c. Der Entwurf muss weniger lang sein als das Papier. So lautet die Vorgabe der Zeitschrift.
the draft must less long be than the paper. So sounds the guideline of the journal
‘The draft is required to be less long than the paper. That’s what the journal’s guideline says.’
- d. Der Entwurf ist weniger lang als das Papier. Das haben die Autoren gesagt.
the draft is less long than the paper. That have the authors said
‘The draft is less long than the paper. That’s what the authors said.’

²Materials, experimental data and analysis scripts of the experiments can be found in the following repository in the Open Science Framework: https://osf.io/528cr/?view_only=a0c996854f75430585eeace550d3d01b

Each sentence doublet is, furthermore, paired with yes-no comprehension questions, as illustrated in (5). There are two types of questions: “*Matching*” questions probe for the preferred or (in case of the controls) only possible reading. “*Mismatching*” questions ask about propositions that are incompatible with the linear scope readings and would thus receive a “no”-response if this was, in fact, the only available reading. A “yes”-response to a mismatching question is indicative of an available inverse scope reading in the modal conditions or an error, e.g. in comprehension or response execution, in the controls. The pairing of target sentences and comprehension questions is indicated by the labels in (4) and (5); e.g. (4-a) is paired with the matching question in (5-a-m) and mismatching question in (5-a-mm).

(5) Comprehension questions

- a-m Soll das Papier 10 Seiten länger sein als der Entwurf?
should the paper 10 pages longer be than the draft
‘Should the paper be 10 pages longer than the draft?’
- a-mm Darf das Papier auch 15 Seiten länger sein als der Entwurf?
may the paper also 15 pages longer be than the draft
‘Is the paper also allowed to be 15 pages longer than the draft?’
- b-m Ist das Papier 10 Seiten lang?
is the paper 10 pages long
‘Is the paper 10 pages long?’
- b-mm Ist das Papier 14 Seiten lang?
is the paper 14 pages long
‘Is the paper 14 pages long?’
- c-m Soll der Entwurf kürzer sein als das Papier?
should the draft shorter be than the paper
‘Should the draft be shorter than the paper?’
- c-mm Darf der Entwurf auch länger sein als das Papier?
may the draft also longer be than the paper
‘Is the draft also allowed to be longer than the paper?’
- d-m Ist der Entwurf kürzer als das Papier?
is the draft shorter than the paper
‘Is the draft shorter than the paper?’
- d-mm Ist der Entwurf länger als das Papier?
is the draft longer than the paper
‘Is the draft longer than the paper?’

Altogether, we thus manipulated the factors MODIFIER (*exactly* vs. *less*), MODAL (*absent* vs. *present*) and QUESTION (*match* vs. *mismatch*), yielding eight conditions in a $2 \times 2 \times 2$ design. The complete set of experimental items comprised 96 pairs of assertions and questions distributed over eight lists using a Latin square (together with 48 fillers). Target and post-context sentences were presented together on a first display.³ The yes-no-question was presented on a subsequent

³Sentences were presented in a self-paced manner and phrase by phrase using the moving window technique. The reason was that, with regard to the filler items, which were part of another unrelated experiment, we were interested in reading times, which are, however, irrelevant and uninformative here.

display in each trial. Participants indicated their response by pressing one of two response keys. The response key assignment was counterbalanced between participants. Before exclusion, 87 participants were recruited via `prolific.co` (mean age: 23 years; sd: 6; 66 female; 19 male).

Statistical analysis: A consequence of our design decision to keep controls simple was that we observed quasi-complete separation with the lowest error rates in control conditions being 1.1% and 2.3% (corresponding to an absolute frequency of one and two) errors in Exps. 1 & 2, respectively. Because quasi-complete separation can make parameter estimates and resulting p -values unreliable, we analyzed the data using Bayesian logistic regression (although we did not observe the usual symptoms of large estimates or standard errors). In particular, we computed Bayesian logit mixed-effects models (implemented in the `blme` package) with weakly informative priors. We assumed normally distributed priors with a mean of 0 and standard deviation of 5 for the fixed effects, as suggested by Clark et al. (2023). This analysis revealed exactly the same qualitative effects as our original logit mixed-effects models, which we provide in the OSF repository.

3.1.2. PREDICTIONS. According to Heim (2000), the presence of the modal verb *müssen* (‘must’) makes the inverse scope reading available, although it may still be dispreferred. Based on this hypothesis – and assuming the control conditions work as intended – a larger number of incongruent answers was, therefore, expected for the modal as compared to the control conditions, irrespective of the degree modifier (*less* vs. *exactly*). This expectation can be illustrated with a simple example: A pattern of results that would meet the expectation could consist of 10% incongruent answers in the controls (say 90% “yes” in matching and 10% “yes” in mismatching questions) because of errors vs. 30% incongruent answers in the potentially ambiguous conditions (say 80% “yes” in matching and 40% “yes” in mismatching questions) because of the availability of the inverse reading. More generally, we predicted a MODAL \times QUESTION interaction in the relative frequency of “yes” responses, where the frequencies of responses in the two categories diverge less clearly between the two question types for controls than for the potentially ambiguous conditions.

By contrast, proposals that assume no scopal flexibility of the comparative operator would not predict any difference between the pattern of responses to conditions with and without modals. A larger number of incongruent answers in the modal conditions may, however, still be explained as errors due to increased semantic and syntactic complexity. The most informative result here would be one without any difference between the modal and control conditions, as this would be hard to account for under the assumption of scopal flexibility of the comparative operator.

3.1.3. RESULTS. After applying predefined exclusion criteria, data from 62 participants were passed on to the statistical analysis. Participants were excluded if they (i) took more than two standard deviations longer than the mean duration of an experimental session, (ii) had individual RTs of more than 10 s, (iii) were below chance in answering comprehension questions in filler trials, (iv) indicated German was not their native language, or (v) missed more than one break (i.e. pressed the same button as in experimental trials instead of a designated button to end the break).

Descriptive results are shown in Fig. 1. Control conditions with the MODIFIER *exactly* led to fewer errors on average (1%) than controls with *less*-comparatives (11.3%). In the modal conditions, questions matching the preferred linear scope interpretation were overwhelmingly answered with “yes” (94.1%) and mismatching questions with “no” (83.3%).

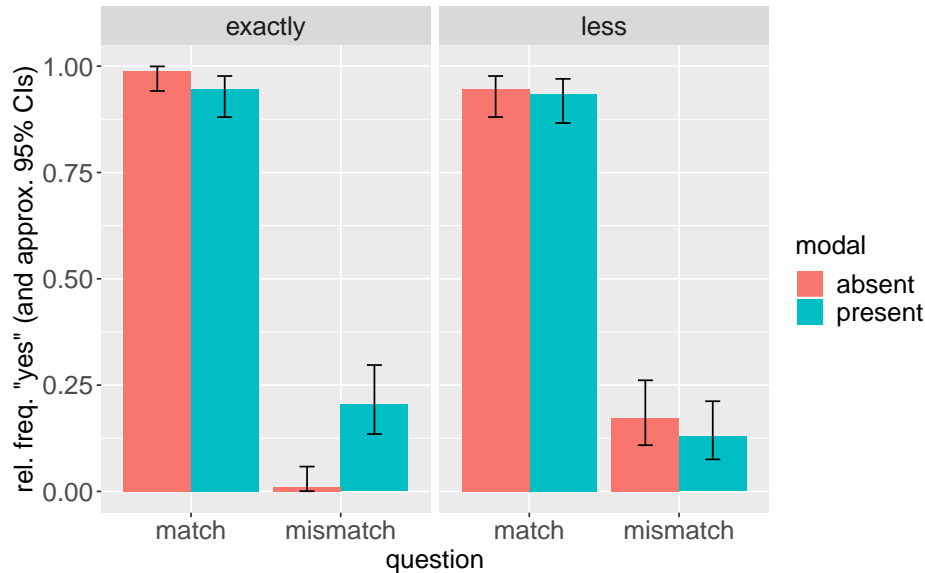


Figure 1: Relative frequency of “yes” responses across conditions in Exp. 1 and approximate confidence intervals (CIs). Confidence intervals were computed for binomial probabilities but ignoring the repeated-measures design.

The statistical analysis revealed a significant three-way interaction ($z = -3.05, p = .002$). To resolve the three-way interaction, we conducted separate analyses of the two MODIFIERS. In the analysis of the *exactly*-conditions, we found the predicted MODAL \times QUESTION interaction ($z = 3.36, p = .001$). This effect was due to the fact that there was less variation in judgments for conditions with MODAL *absent* vs. *present*. Pairwise comparisons showed a significant effect of the MODAL in *mismatching* ($z = 5.19, p < .001$) but not in *matching* ($z = -1.44, p = .15$) conditions. This contrasts with the *less*-comparatives where no MODAL \times QUESTION interaction was found ($z = -0.32, p = .747$). In addition, the expected main effect of QUESTION, i.e. more positive responses to *matching* vs. *mismatching* questions, was found for both MODIFIERS ($z = -7.84$ and $z = -9.26$, for *exactly* and *less*, resp.). However, the effect of MODAL was not significant in either subset of the data.

3.1.4. DISCUSSION. Indication of the purported ambiguity was limited to *exactly*-differentials. This difference between the two modifier types was surprising to us, as they are usually not distinguished in the theoretical literature with regard to their scope taking potential (but see the General Discussion below for possible explanations of the difference). However, even in the *exactly*-differentials, the indication of the purported ambiguity that we found may, in fact, be an artefact driven by the simplicity of the *exactly*-controls rather than ambiguity in the corresponding modal conditions: The *exactly*-controls were the only conditions without comparatives. They were therefore relatively easy to judge and led to almost flawless performance. As a matter of fact, a post-hoc analysis supports this interpretation. In particular, we found a MODIFIER TYPE \times QUESTION-interaction in controls ($z = 3.062, p = .002$), but not in the conditions with modals ($z = -0.449, p = .653$). These results stem from the fact that in the control conditions, response options were differentiated more clearly for the *exactly*-differentials (in the positive form

of the adjective) as compared to the *less*-comparatives (in the comparative form of the adjective), whereas no such interaction effect was found in the target conditions (all in comparative form). Thus, we conclude at this point that the data of Exp. 1 provide little evidence for the purported ambiguity, at least in German.

3.2. EXPERIMENT 2. Exp. 2 was a quasi-replication intended to test whether the results of Exp. 1 reflect differences between the two types of comparatives or were due to specific characteristics of the controls, i.e. potential differences in comprehension difficulty of the comparative vs. the positive form of the adjectives. To this end, *exactly*-controls, e.g. (4-b), were put into comparative form, e.g. (5-e) – like all the other conditions.

(5) Modified target sentences and post context in *exactly*-control of Exp. 2

- e. Das Papier ist genau 10 Seiten länger als der Entwurf. Das haben die Autoren gesagt.
 the paper is exactly 10 pages longer than the draft. That have the authors said
 ‘The paper is exactly 10 pages longer than the draft. That’s what the authors said.’

(6) Modified comprehension questions in *exactly*-control of Exp. 2

- e-m. Ist das Papier 10 Seiten länger als der Entwurf?
 is the paper 10 pages longer than the draft
 ‘Is the paper 10 pages longer than the draft?’
 e-mm. Ist das Papier 14 Seiten länger als der Entwurf?
 is the paper 14 pages longer than the draft
 ‘Is the paper 14 pages longer than the draft?’

3.2.1. METHODS. Except for the modified *exactly*-controls, the design, materials, procedure and statistical analysis of Exp. 2 were identical to Exp. 1. The predictions were also the same as in Exp. 1. We used the same type of statistical analysis to test for those predictions. In total, 87 new participants were recruited via `prolific.co` (mean age: 33.17 years; sd: 11.2; 46 female; 41 male).

3.2.2. RESULTS. After applying the same exclusion criteria as in Exp. 1, data from 61 out of 87 participants were passed on to the statistical analysis. Mean judgments are depicted in Figure 2. They are by and large comparable to the results from Exp. 1, except for a bit more errors in the mismatching questions in the *exactly*-controls (Exp. 1: 1%; Exp. 2: 5.6%). There were, however, again fewer errors on average in the *exactly*-controls than in controls with *less*-comparatives (4% vs. 13.7%). In the *modal* conditions, matching questions were, as in Exp. 1, overwhelmingly answered with “yes” (88%) and mismatching questions with “no” (92.9%).

The statistical analysis revealed a significant three-way interaction ($z = -2, p = .046$). Again, we resolved this interaction by using separate analyses of the two MODIFIERS. In the analysis of the *exactly*-conditions, we found a marginal MODAL \times QUESTION interaction ($z = 1.86, p = .063$), reflecting a larger difference between *matching* vs. *mismatching* questions in the *modal* than in the *control* conditions. Pairwise comparisons showed a marginal effect of the MODAL within the two conditions with *matching* questions ($z = -1.87, p = .061$), with less variation in judgments for conditions with MODAL *absent* vs. *present*. No such effect was found in *mismatching exactly*-conditions ($z = .68$). As in Exp. 1, this contrasts with the *less*-comparatives where no interaction

was found ($z = -0.9, p = .38$). Furthermore, the expected main effect of QUESTION, i.e. more positive responses to *matching* vs. *mismatching* questions, was found for both modifier types ($z = -8.15$ and $z = -9.8$, for *exactly* and *less*, resp.). Finally, there was an effect of MODAL that was only significant in *less*-comparatives ($z = -2.97, p = .003$). The latter effect was due to the fact that there were fewer “yes”-responses for MODAL than CONTROL conditions. This numerical pattern was also observed in Exp. 1, but there it did not lead to a significant effect.

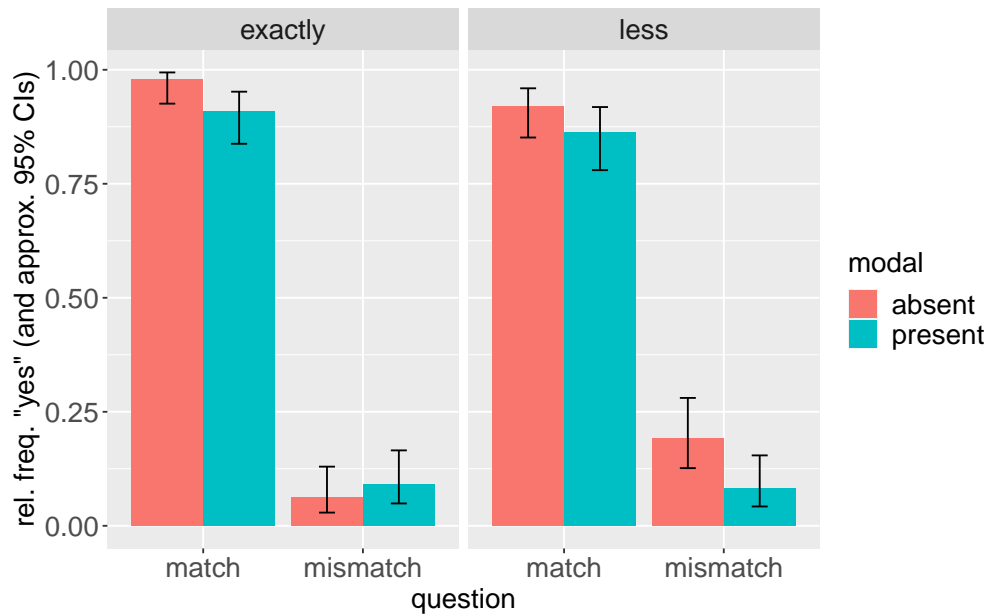


Figure 2: Relative frequency of “yes” responses across conditions in Exp. 2 and CIs computed for binomial probabilities ignoring repeated-measures.

3.3. DISCUSSION. Although the comparative *exactly*-controls in Exp. 2 did in fact lead to a few more errors as compared to Exp. 1, as we expected, the general pattern of results was the same in both experiments. Indication of the ambiguity was again limited to the *exactly*-conditions. In these conditions, it hinges on a marginal difference in the generally rather high proportion of “yes”-responses to the matching questions in *exactly*-controls vs. *exactly*-targets. A difference between the two experiments was a significant effect of MODAL in Exp. 2 that was absent from Exp. 1. Given, its inconsistency, we cannot be confident that it is a real effect. Moreover, this was an unexpected effect and we therefore refrain from speculating about potential reasons.

4. Experiment 3 on English: English materials with contextual embedding. The first two experiments yielded consistent and highly comparable results. Nevertheless, there are still questions regarding their generality. Exp. 3 addressed four of these questions. The first question concerns the tested language. Exps. 1 & 2 used German materials, although the original considerations of Kennedy (1997) and Heim (2000) were focused on English. While we have no specific reason to believe that there are relevant differences in the semantics of comparative operators between the two languages (see section 2 for a brief discussion), there is empirical evidence that scope ambiguities can be more pronounced in English as compared to German (e.g., Philipp & Zimmermann

2020, 2023). In Exp. 3, we therefore tested English translations of the items from the previous two experiments. Secondly, we embedded the translated items into contexts that were constructed in such a way as to make the inverse reading salient. The intention was to boost the potentially dispreferred inverse reading as much as possible. Thirdly, we used demonstrative *that* in the *than*-clause to refer to the reference degree that was introduced in the context, as was also the case in the original examples discussed by Heim (2000). If the *than*-clause contains a definite description, a *de-dicto* reading is possible (e.g. ...*than the paper is required to be*; cf Stechow 1984). While we do not see how this could have weakened the indication of ambiguity in the observed judgments of Exps. 1 & 2, it may still cause differences between the materials tested in these experiments and the examples discussed by Heim (2000). Finally, we removed the *auch* ('also') from the mismatching questions in the MODAL conditions because it may have triggered a presupposition that the preferred linear scope reading is true.

4.1. METHODS. The items from Exps. 1 & 2 were translated into English and embedded into contexts that were intended to make the inverse reading salient. Demonstrative *that* was included in the *than*-clause and *auch* ('also') was removed from the mismatching questions. The context and target sentences from an example item are shown in (7). The *than*-clause referred to a degree that was introduced in the context. The final two sentences of the context varied depending on the condition of the target sentence.

- (7) Alex's company has been assembling motorbikes, scooters and mopeds for well-known manufacturers for years. After a major refurbishment of the main production facility, the machinery on the assembly line needs to be recalibrated. In order to ensure that the minimum height standards for different vehicle types are met, Alex and his engineers are consulting various designs provided by the manufacturers. ...
- a. ...The specifications from one particular manufacturer are especially puzzling. For instance, one of the mopeds is 90 cm high. **The motorbike is required to be exactly 50 cm higher than that.** That's what the manufacturer specified.
 - b. ...They are also trying to incorporate their customers' feedback. One of the customers indicated that their moped is 90 cm high. **The motorbike is exactly 50 cm higher than that.** That's what the motorcyclist said.
 - c. ...The specifications from one particular manufacturer are especially puzzling. For instance, one of the motorbikes is 140 cm high. **The moped is required to be less high than that.** That's what the manufacturer specified.
 - d. ...They are also trying to incorporate their customers' feedback. One of the customers indicated that their motorbike is 140 cm high. **The moped is less high than that.** That's what the motorcyclist said.

Participants read the context on a first display and the target sentence on a second one. Experimental items were distributed across eight lists as in Exps. 1 & 2 and 18 filler items were added to each list. The exclusion criteria were identical to Exps. 1 & 2 except for the fact that no RT-based criterion was used and participants were excluded if the 'accuracy' in their ratings on designated filler trials was more than 2 standard deviations away from the mean. Apart from these differences, the materials and procedure were identical. We recruited 202 participants over `prolific.co` (mean age: 45.6 years; sd: 14.8; 121 female; 79 male).

4.1.1. RESULTS. After exclusion, data from 199 participants were passed on to the statistical analysis. Mean judgments are shown in Fig. 3. As compared to Exps. 1 & 2, we observed slightly more variation in responses across the board in the current experiment (mean judgments ranged

between 8.6% and 91.4%). Nevertheless, we found a similar pattern of results overall, as evidenced by a significant three way interaction ($z = 2.02, p = .044$) and also similar effects within *exactly*-differentials and *less*-comparatives. In the *less*-comparatives, we found main effects of QUESTION TYPE ($z = -18.23, p < .001$) and MODAL ($z = -3.57, p < .001$) again but no interaction between them ($z = -.18, p = .86$). In the *exactly*-differentials, we found a significant interaction between QUESTION TYPE and MODAL ($z = 2.86, p = .004$). This was due to more “yes”-responses in the mismatching condition in the *presence* vs. *absence* of MODALS ($z = 4.08, p < .001$). In the *matching* QUESTION TYPE, there was again no difference due to MODALS ($z = -.04, p = .97$).

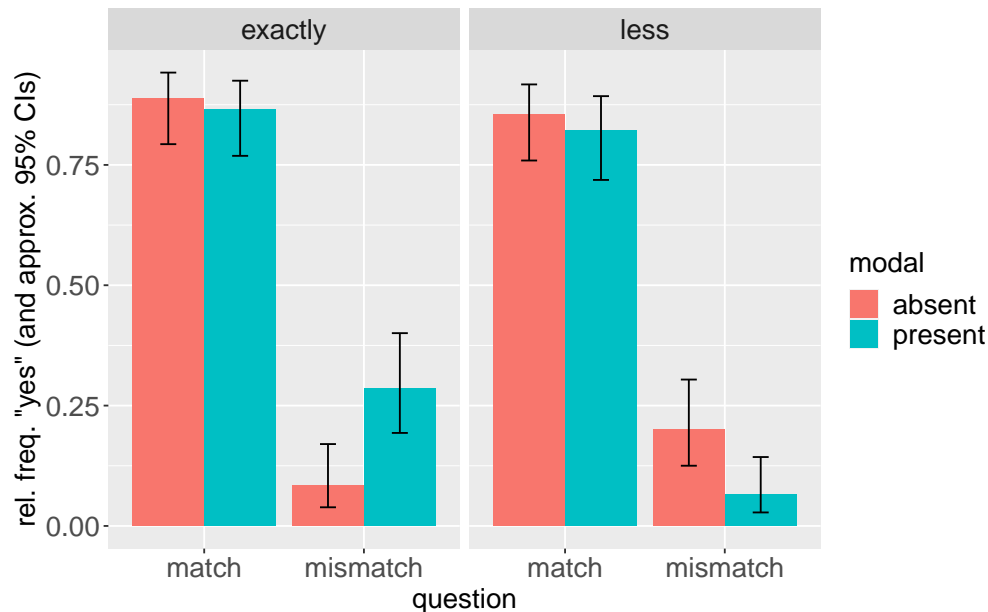


Figure 3: Relative frequency of “yes” responses across conditions in Exp. 3 and CIs computed for binomial probabilities ignoring repeated-measures.

4.1.2. DISCUSSION. Contextual embedding led to a bit more variance in judgments as compared to Exps. 1 & 2. This is likely due to contextual embedding and increased memory load resulting from the presentation on two consecutive displays. Apart from this difference, the qualitative pattern of results was highly comparable between all three experiments. Thus, the results from the previous two experiments seem neither specific to German comparatives nor attributable to missing contextual support. Moreover, our findings cannot be attributed to *de dicto* readings of the definite expressions in the *than*-clauses in Exps. 1 & 2 or to the presence of the presuppositional item *auch* (‘also’) in the mismatching questions. This is because both features were absent in the current experiment. Next, we turn to the theoretical implications we draw from the three experiments.

5. General Discussion. Although the theory of natural language quantifiers comes in different flavours (e.g., Barwise & Cooper 1981, Heim & Kratzer 1998, Barker 2002, Sternefeld 2020), it has been extremely successful in the sense that its basic assumptions explain a large number of different linguistic phenomena across languages. The approach of Heim (2000; drawing on Stechow 1984, Heim 1985; a.m.o.) that was tested in the current study can be viewed as an attempt to increase

the coverage and test the explanatory power of quantifier theory by applying it to comparative expressions as well (see also Lassiter 2012; for a similar attempt), at least in languages with degree abstraction (cf. Beck 2009). However, despite the granted elegance and predictive power of this general approach, our experimental data indicate that with regard to comparatives, it has to be critically evaluated when confronted with systematically collected speaker judgments.

In both experiments on German and the experiment on English, indication of the purported ambiguity was limited to *exactly*-differentials, and even in these cases, evidence for the minimal requirement reading is not particularly strong. We were able to replicate the results for English even after changing the dependent measure and fixing potential problems such as the possibility of *de dicto* readings.

The question is what sets *exactly*-differentials and *less*-comparatives apart with respect to scope taking. One possibility could be their monotonicity properties (non-monotone vs. downward entailing). Potential effects of monotonicity were discussed at length by Heim (2000). She concluded that inverse scope can be detected with both non-monotone and downward entailing quantifiers. However, it is not clear how an explanation of our current findings in terms of monotonicity might look like. Another difference between the two modifiers is that decompositional analyses have been suggested for *less than* (e.g., Heim 2006) but we are not aware of similar proposals for *exactly*-differentials (beyond the obvious; cf. below). Decompositional analyses have in fact been related to ambiguities in *less*-comparatives containing modals (e.g., Heim 2006, Rullmann & Beck 1996). Comparable to our results, the prevalence of such ambiguities was, however, challenged in an empirical investigation by Beck (2012b) using German materials. Furthermore, decompositional analyses would likely predict more rather than less potential for scope interaction in *less*-comparatives, as compared to other, non-decompositional approaches.

One possible explanation of our data could be derived from Oda (2008; chap. 2), who accounts for the ambiguity in (1) in terms of scope mobility of the differential measure phrase *exactly 5 pp* rather than the comparative operator. She considers the possibility that *exactly*-differentials allow for scope interaction, because the *exactly*-measure phrase contained in them is a generalized quantifier over degrees (type $\langle\langle d, t \rangle, t\rangle$) and can move. She derives the relevant readings from this assumption. By contrast, *less*-comparatives are not ambiguous in the same way. In the domain of equatives, Penka (2024) in a response to Hohaus & Zimmermann (2021) entertains a similar possibility: Namely that the Heim-ambiguity is due to the scope of *genau* ('exactly') rather than the scope of the equative operator. Another promising avenue for explanation could potentially be derived from the following difference: *Exactly* is assumed to trigger obligatory implicatures, for instance by insertion of a covert exhaustification operator (Landman 1998, Gajewski 2008). Another related account is that of Beck (2012a) who derives scope ambiguity of *less*-comparatives using an alternative-semantics along the lines of Rooth (1992), but also uses the proposal of Oda (2008) to explain ambiguity in *exactly*-differentials. With respect to scope ambiguity in *exactly*-differentials, there is thus no substantial difference between the proposals of Oda (2008) and Beck (2012a). Irrespective of the explanation that may turn out to be correct, our data indicate that the empirical basis for scope mobility of comparative operators is even more restricted than previously assumed. We suggest that scope mobility of comparative operators across languages should be subject of further empirical scrutiny.

References.

- Agmon, Galit, Yonatan Loewenstein & Yosef Grodzinsky. 2019. Measuring the cognitive cost of downward monotonicity by controlling for negative polarity. *Glossa: A Journal of General Linguistics* 4(1). 1–18. <https://doi.org/10.5334/gjgl.770>.
- Barker, Chris. 2002. Continuations and the nature of quantification. *Natural Language Semantics* 10. 211–241. <https://doi.org/10.1023/A:1022183511876>.
- Barwise, John & Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4(2). 159–219. <https://doi.org/10.1007/BF00350139>.
- Beck, Sigrid. 2009. Positively comparative. *Snippets* 20. 4–6.
- Beck, Sigrid. 2011. Comparison constructions. In Claudia Maienborn, Klaus von Stechow & Paul H. Portner (eds.), *Semantics: An International Handbook of Natural Language Meaning*, vol. 2, 1341–1389. De Gruyter. <https://doi.org/10.1515/9783110255072.1341>.
- Beck, Sigrid. 2012a. DegP scope revisited. *Natural Language Semantics* 20. 227–272. <https://doi.org/10.1007/s11050-012-9081-6>.
- Beck, Sigrid. 2012b. Lucinda driving too fast again – the scalar properties of ambiguous *than*-clauses. *Journal of Semantics* 30(1). 1–63. <https://doi.org/10.1093/jos/ffr011>.
- Beck, Sigrid, Svetlana Krasikova, Daniel Fleischer, Remus Gergel, Stefan Hofstetter, Christiane Savelsberg, John Vanderelst & Elisabeth Villalta. 2009. Crosslinguistic variation in comparison constructions. *Linguistic Variation Yearbook* 9. 1–66. <https://doi.org/10.1075/livy.9.01bec>.
- Beck, Sigrid, Toshiko Oda & Koji Sugisaki. 2004. Parametric variation in the semantics of comparison: Japanese versus English. *Journal of East Asian Linguistics* 13(4). 289–344. <https://doi.org/10.1007/s10831-004-1289-0>.
- Berezovskaya, Polina. 2014. Acquisition of Russian degree constructions: A corpus-based study. In *Proceedings of Formal Approaches to Slavic Linguistics (fasl)* 22, 1–22.
- Bhatt, Rajesh & Roumyana Pancheva. 2004. Late merger of degree clauses. *Linguist Inquiry* 35(1). 1–45. <https://doi.org/10.1162/002438904322793338>.
- Bochnak, Ryan. 2015. The degree semantics parameter and cross-linguistic variation. *Semantics and Pragmatics* 8(6) 1–48. <https://doi.org/10.3765/sp.8.6>.
- Bowler, Margit. 2016. The status of degrees in Warlpiri. In *Proceedings of the Semantics of African, Asian and Austronesian languages (Triple A)* 2, 1–17.
- Breakstone, Micha Y., Alexandre Cremers, Danny Fox & Martin Hackl. 2011. On the analysis of scope ambiguities in comparative constructions: converging evidence from real-time sentence processing and offline data. In *Proceedings of Semantics and Linguistic Theory (SALT)* 21, 712–731. <https://doi.org/10.3765/salt.v21i0.2609>.
- Clark, Robert G., Wade Blanchard, Francis K.C. Hui, Ran Tian & Haruka Woods. 2023. Dealing with complete separation and quasi-complete separation in logistic regression for linguistic data. *Research Methods in Applied Linguistics* 2(1). 1–11. <https://doi.org/10.1016/j.rmal.2023.100044>.
- Gajewski, Jon. 2008. More on quantifiers in comparative clauses. *Proceedings of Semantics and Linguistic Theory (SALT)* 18. <https://doi.org/10.3765/salt.v0i0.2494>.
- Gibson, Edward & Evelina Fedorenko. 2013. The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes* 28(1–2). 88–124.

- <https://doi.org/10.1080/01690965.2010.515080>.
- Gibson, Edward, Steven T. Piantadosi & Evelina Fedorenko. 2012. Quantitative methods in syntax/semantics research: A response to Sprouse and Almeida (2013). *Language and Cognitive Processes* 28(3). 229–240. <https://doi.org/10.1080/01690965.2012.704385>.
- Heim, Irene. 1985. Notes on comparatives and related matters. Manuscript. Austin: University of Texas.
- Heim, Irene. 2000. Degree operators and scope. In *Proceedings of Semantics and Linguistic Theory (SALT) 10*, 40–64. <https://doi.org/10.3765/salt.v10i0.3102>.
- Heim, Irene. 2006. Little. In *Proceedings of Semantics and Linguistic Theory (SALT) 16*, 35–58. <https://doi.org/10.3765/salt.v16i0.2941>.
- Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Malden: Blackwell.
- Hohaus, Vera, Sonja Tiemann & Sigrid Beck. 2014. Acquisition of comparison constructions. *Language Acquisition* 21(3). 215–249. <https://doi.org/10.1080/10489223.2014.892914>.
- Hohaus, Vera & Malte Zimmermann. 2021. Comparisons of equality with German *so...wie*, and the relationship between degrees and properties. *Journal of Semantics* 38(1). 95–143. <https://doi.org/10.1093/jos/ffaa011>.
- Howell, Anna. 2012. *Comparatives and scope in Yoruba*: University of Tübingen MA thesis.
- Kapitonov, Ivan. 2019. Degrees and scales of Kunbarlang. In *Proceedings of Triple A5*, 91–105.
- Kennedy, Christopher. 1997. *Projecting the adjective: The syntax and semantics of gradability and comparison*: University of California at Santa Cruz dissertation.
- Landman, Fred. 1998. *Plurals and maximalization* 237–271. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-011-3969-4_10.
- Lassiter, Daniel. 2012. Quantificational and modal interveners in degree constructions. In Anca Chereches (ed.), *Proceedings of Semantics and Linguistic Theory (SALT) 22*, 565–583. <https://doi.org/10.3765/salt.v22i0.2649>.
- Oda, Toshiko. 2008. *Degree constructions in Japanese*. Storrs: University of Connecticut dissertation.
- Penka, Doris. 2024. German *so ...wie*-constructions as definite descriptions. a reply to Hohaus & Zimmermann (2021). Manuscript.
- Philipp, Mareike & Malte Zimmermann. 2020. Empirical investigations on quantifier scope ambiguities in German. In *Proceedings of Sinn und Bedeutung 24*, 145–163. <https://doi.org/10.18148/sub/2020.v24i2.914>.
- Philipp, Mareike & Malte Zimmermann. 2023. An experimental comparison of the availability of inverse scope in English and German. *Linguistic Inquiry* 1–37. https://doi.org/10.1162/ling_a.00493.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 4(1). 75–116. <https://doi.org/10.1007/BF02342617>.
- Rullmann, Hotze & Sigrid Beck. 1996. Degree questions, maximal informativeness, and exhaustivity. In Paul Dekker & Martin Stokhof (eds.), *Proceedings of the Amsterdam Colloquium 10*, 73–92.
- Sprouse, Jon & Diogo Almeida. 2013. The empirical status of data in syntax: A reply to Gibson and Fedorenko. *Language and Cognitive Processes* 28(3). 222–228.

<https://doi.org/10.1080/01690965.2012.703782>.

Stateva, Penka. 2000. In defense of the movement theory of superlatives. In Rebecca Daly & Anastasia Riehl (eds.), *Proceeding of Escol 1999*, 219–226. CLC Publications.

Stechow, Arnim von. 1984. Comparing semantic theories of comparison. *Journal of Semantics* 3(1-2). 1–77.

Sternefeld, Wolfgang. 2020. *Quantifiers, scope, and pseudo-scope*. 1–44. John Wiley and Sons, Ltd. <https://doi.org/10.1002/9781118788516.sem124>.