

## 1 Results

The judgments obtained for the four target conditions are depicted in Figure 1. We coded the judgments as literal, global or local if they were as expected under one of these readings and as error if not. The distribution of readings is presented in Figure 2. Participants mostly gave judgments indicating literal or local readings. Judgments compatible with a global reading were seldomly obtained. In the *All-Some* conditions participants made hardly any errors, whereas the amount of errors was slightly greater in the *Exactly-One-Some* conditions.

In the *All-Some* conditions (Figure 3(c)) judgments were consistent with literal readings in 65.0% of the trials with neutral prosody and in 67.5% of the trials with accented prosody. The number of global readings was very low, with 1.6% of the neutral and 0.8% of the accented trials. Judgments indicating local readings were given in 22.0% of both the neutral and accented trials. In the *Exactly-One-Some* conditions (see Figure 3(d)) judgments consistent with literal readings were given in 47.2% and 44.7% of the trials with neutral and accented prosody, respectively. Global readings were observed in 0.8% of the trials in both the neutral and the accented condition. Judgments indicating local readings were given in 22.8% of the neutral and 27.6% of the accented trials.

In order to test whether local readings exist, the amount of local responses was compared to chance. Local responses had to be given on the last position in each trial. If local readings did not exist we would not expect participants, who reached the last position (i.e. who did not abort the trial prior to the last position) to give local judgments significantly more often than expected by chance. On the last position there were two responses available to the participants. One was compatible with a local reading the other was not. Therefore, we would expect 50% local responses by chance. As it turns out local responses were given significantly more often than 50% in all target conditions (all  $p < .05$ ). Note that this finding cannot be explained by a response bias on the last position since local readings required a yes-judgment in the *Exactly-One-Some* conditions, but a no-judgment in the *All-Some* conditions.

In order to test whether accentuation or the type of quantifier has an influence on the distribution of readings, log-linear models with the factors *Reading*, *Accentuation* and *Quantifier* were computed (see Scheepers 2003). Two variants of these models were computed. In the first variant, the factor *Item* was additionally included, whereas in the second variant, the factor Participants was added. The inclusion of these two factors allowed us to test whether the distribution of readings was identical across items and participants. We report log-likelihood ratio Chi-squares (LRCS<sub>1</sub> and LRCS<sub>2</sub>), degrees of freedom (df<sub>1</sub>, df<sub>2</sub>) and significance levels (p<sub>1</sub> and p<sub>2</sub>).

Unsurprisingly, there was an effect of *Reading* (LRCS<sub>1</sub> = 364.77, df<sub>1</sub> = 3,  $p_1 < .001$ , LRCS<sub>2</sub> = 364.77, df<sub>2</sub> = 3,  $p_1 < .001$ .) because the readings were distributed inhomogeneously (see Figure 2). The type of quantifier had an influence on the distribution of readings as revealed by a reliable interaction

Petra 7.5.12 10:35

**Kommentar:** as coded by us are : deleted

Petra 7.5.12 10:36

**Kommentar:** würde ich kursiv machen, hab ich unten durch rote Schriftfarbe markiert

Petra 7.5.12 10:36

**Kommentar:** see deleted

Petra 7.5.12 10:37

**Kommentar:** das könnten wir vielleicht neutraler formulieren, es klingt hier bereits etwas wertend

Petra 7.5.12 10:40

**Kommentar:** wether -> whether, ammount -> amount

Petra 7.5.12 11:04

**Kommentar:** Würde ich umformulieren: „As a negative judgment on the final position in each trial corresponded to a local reading, we expected to observe differences from chance level (i.e. from 50%, as both positive and negative judgments were available at this position) only if participants adapted a local strategy. Indeed, negative (= local) responses were given significantly more often ...“ [how often was it??]

Petra 7.5.12 10:56

**Kommentar:** Satz gestrichen / gekürzt

Petra 7.5.12 10:58

**Kommentar:** Sätze geändert

Petra 7.5.12 11:01

**Kommentar:** sse - > see

of *Reading* and *Quantifier* ( $\text{LRCS}_1 = 49.32$ ,  $p_1 < .01$ ,  $\text{LRCS}_2 = 32.16$ ,  $p_1 < .01$ ). On the basis of the obtained data it did, however, seem unlikely that the type of quantifier affected the amount of local readings. We suspected that the interaction was due to the higher number of errors in the *Exactly-One-Some* conditions as compared with the *All-Some* conditions. In addition to the effect of the quantifier, a difference in the distribution of judgment types between participants was revealed by a reliable interaction of *Reading* and *Participant* ( $\text{LRCS}_1 = 487.70$ ,  $p_1 < .01$ ). We were interested in whether this effect was due to some of the participants being more likely to choose local readings than others. Finally, there was a three-way interaction between *Participant*, *Quantifier* and *Reading* ( $\text{LRCS}_1 = 143.92$ ,  $\text{df}_1 = 117$ ,  $p_1 < .05$ ), which could stem from the fact that some participants were more prone than others to make errors in the *Exactly-One-Some* conditions.

In order to find out whether the interactions just reported affected the amount of local readings, the log-linear models were computed again with a different coding of the readings. Here, we only considered the amount of local readings versus all other kinds of judgments. Judgments were coded as "local" and "other". If the reported interactions affected the amount of local readings, they should show up again. The expected but irrelevant effect of *Reading* was again significant ( $\text{LRCS}_1 = 134.55$ ,  $\text{df}_1 = 1$ ,  $p_1 < .001$ ,  $\text{LRCS}_2 = 144.64$ ,  $\text{df}_2 = 1$ ,  $p_1 < .001$ ). In addition, we found an interaction of *Participant* and *Reading* ( $\text{LRCS}_1 = 487.70$ ,  $p_1 < .01$ ), as well as a three-way interaction of *Quantifier*, *Participant* and *Reading* ( $\text{LRCS}_1 = 487.70$ ,  $p_1 < .01$ ). No other effects were significant. In particular, the interaction of *Quantifier* and *Reading* was not significant ( $\text{LRCS}_1 = 2.10$ ,  $p_1 = .15$ ,  $\text{LRCS}_2 = .74$ ,  $p_2 = .39$ ). The interaction of *Participant* and *Reading* shows that there are indeed varying preferences for local readings. Further examination of the distribution of local readings revealed a clear pattern (see Figure 1). In all four conditions, about seven out of the 40 participants (16%) were consistently giving judgments that indicate local readings. Further, the relative frequency of local judgments per participant strongly correlated between all four conditions (all  $r > .6$ ,  $p < .001$ ). Apart from the localists about half of the participants (and 41.5% of all participants) were consistently giving judgments that indicate literal readings. Only a minority of participants exhibited inconsistency. Reading preferences were more clear-cut in the *All-Some* conditions than in the *Exactly-One-Some* conditions. The number of consistent "literalists" was lower in the *Exactly-One-Some* (25.6%) conditions as compared to the *All-Some* conditions (57.3%). Also, the number of participants that showed inconsistency was higher in these conditions. However, the number of consistent localists did hardly differ between constructions.

The three way-interaction of *Participant*, *Quantifier* and *Reading* might be explained by the fact that for some participants, preferences for local over other readings deviated between the two quantifier types. These deviations were, however, not systematic to any degree. Most participants had exactly the same preferences for both quantifiers. However, a few had a stronger and a few others a weaker preference for local readings in the *Exactly-One-Some* than in the *All-Some* conditions. Since we did not find any systematic patterns here, we speculate that the three-way interaction is due to the *Exactly-One-*

Petra 7.5.12 11:08

**Kommentar:** Den Satz verstehe ich nicht: wieso ist es anhand der Daten unwahrscheinlich? Ich würde den Satz ... [1]

Petra 7.5.12 11:12

**Kommentar:** OK, haben wir auch w ... [2]

Petra 7.5.12 11:10

**Kommentar:** geändert

Petra 7.5.12 11:10

**Kommentar:** komma

Petra 7.5.12 11:13

**Kommentar:** Judgmentes -> Judgments

Petra 7.5.12 11:14

**Kommentar:** Anführungszeichen

Petra 7.5.12 11:14

**Kommentar:** komma

Petra 7.5.12 11:15

**Kommentar:** komma

Petra 7.5.12 11:17

**Kommentar:** Satz modifiziert

Petra 7.5.12 11:19

**Kommentar:** consistently -> consistently

Petra 7.5.12 11:19

**Kommentar:** Apart -> Apart

Petra 7.5.12 11:26

**Kommentar:** → „Apart from thes ... [3]

Petra 7.5.12 11:27

**Kommentar:** literalists -> „literalists“

Petra 7.5.12 11:27

**Kommentar:** inconsistency...

Petra 7.5.12 11:29

**Kommentar:** vorsichtiger formulieren

Petra 7.5.12 11:31

**Kommentar:** komma

Petra 7.5.12 11:31

**Kommentar:** with -> for

Petra 7.5.12 11:36

**Kommentar:** Satz weglassen!

Petra 7.5.12 11:36

**Kommentar:** systematic

*Some* construction being understood non-standardly by a few participants. This speculation is plausible given the higher number of errors in the *Exactly-One-Some* as compared to the *All-Some* conditions.

The absence of an interaction between *Quantifier* and *Reading* indicates that the type of quantifier does not affect the amount of local readings. We assumed based on the observed distribution of judgments that the type of quantifier affected the amount of literal but not local and global readings. To further test this assumption, we also considered literal and global readings versus other judgments in separate analyses. Comparing global readings versus other judgments *Quantifier* and *Reading* were found to interact ( $LRCS_1 = 42.80, df = 1, p < .001; LRCS_2 = 22.14, df = 1, p < .001$ ). Comparing global readings to other judgments no such interaction was observed ( $LRCS_1 = 1.24, df = 1, p = .29, LRCS_2 = .21, df = 1, p = .65$ ).

Petra 7.5.12 11:37

Kommentar: beeing -> being

Petra 7.5.12 11:37

Kommentar: the -> an

Petra 7.5.12 11:39

Kommentar: observed

#### Allgemeine Kommentare von Petra

1. Generell bin ich es gewöhnt, zunächst die Statistik zu berichten und dann die Interpretation der Daten. Wenn es z.B. keine Dreifachinteraktionen gibt, dann berichte ich auch keine Einzelvergleiche. Diese bespreche ich dann höchstens "deskriptiv" in der Discussion. Das betrifft insbesondere auch die Prosodieeffekte, von denen nie wieder geredet wird, weil sie nicht signifikant sind. Damit komme ich auch zu meinem 2. Punkt.
2. Die Prosodie sollte ab und zu mal erwähnt werden. Es wird zumeist berichtet, welche Interaktionen nun nicht significant geworden sind. Manchmal sollten wir also einen Satz einwerfen wie "We did not observe any significant main effects of Accentuation nor any interactions between Accentuation or any of the other factors". Auch hast Du den Ergebnisbericht nicht strikt von dessen Interpretation getrennt. Ich persönlich finde das so auch immer besser, das kommt natürlich auch auf das Journal an. Wenn wir es so machen, dann sollte man wirklich ab und zu was von der Prosodie erwähnen, z.B. immer am Ende der jeweiligen Analyse. Soll ich das noch einfügen?