Disentangling semantic empty-set effects in quantifier comprehension from simple associations – Processing evidence from exceptive-additives

Introduction: Recently, [1] proposed an algorithmic theory of quantifier interpretation in natural language meant to be cognitively realistic. One specific distinguishing feature of this theory, e.g. in comparison to Generalized Quantifier Theory [2], is that it predicts so-called empty-set (ES) quantifiers, which have the empty set as a witness set (i.e. they are true in a situation in which the scope set is empty (ES-situation)), to be more difficult to interpret than quantifiers that do not have the empty set among their witnesses. Since all downward entailing quantifiers are ES-quantifiers, the proposal of [1] explains the well-known processing difficulty associated with downward entailing quantifiers [3]. What's more, the proposal also predicts extraordinary difficulty (the so-called ES-effect) in case ES-quantifiers are verified in ES-situations. In a combined self-paced reading plus picture verification experiment, [1] confirmed these predictions across a range of different types of quantifiers. Moreover, they teased apart difficulty of ES-quantifiers from effects of monotonicity by establishing ES-effects for the non-monotone empty-set quantifier *no or three*. In the current study, we combined ES-quantifiers with the German so called exceptive-additive *außer* (see [4]) to investigate how the additive inference, triggered by *außer*, interacts with ES-effects. The specific motivation for this study was to further test the algorithmic theory of [1] by ruling out an alternative explanation of ES-effects that is based on a superficial mismatch between linguistic and visual features.

Explanations of the ES-effect: In the proposal of [1], ES-quantifiers are difficult because their verification algorithm involves one specific rule intended to handle ES-situations. This rule relies on the explicit representation of objects in the restrictor set that are not in the scope set. [1] argue that this type of encoding is cognitively costly because the cognitive system is tuned to encode the presence of properties rather than their absence, a strategy viable for non-ES-quantifiers but doom to failure for ES-quantifiers. While the data of [1] support this assumption, they cannot rule out the possibility that the observed ES-effects were at least in part due to a superficial mismatch between the visual contexts and the content words in the tested sentences. To illustrate this explanation, consider a sentence like *at most one triangle is red* in the context of a picture like in Fig 2a. The sentence is true although none of the objects has the mentioned color (i.e. none is red). There is thus a conflict between the truth-value and the semantic dis-similarity between visual features and content words. This mismatch in ES-situations is restricted to ES-quantifiers and could thus explain at least some part of the observed empty-set effects.

Testing empty set effects in 'außer' ('except') sentences: Modifying ES-quantifiers with the so called exceptive-additive *außer* (see [4]), allows for testing ES-effects in the absence of the type of superficial mismatch just described. For a sentence such as (1), a picture such as Fig. 2b may constitute an empty-set situation even though there are red triangles in the picture. Exceptive-additives have not received much attention in experimental work and are therefore interesting in their own right. A natural extension of [1] is to assume that the modification of empty-set quantifiers with these phrases adds e.g. domain subtraction as another step of verification to their proposed processing algorithm (see [5]). Since exceptive-additives can trigger also a non-additive depending on the quantifier they are combined with, four control conditions were added to ensure the additive interpretation.

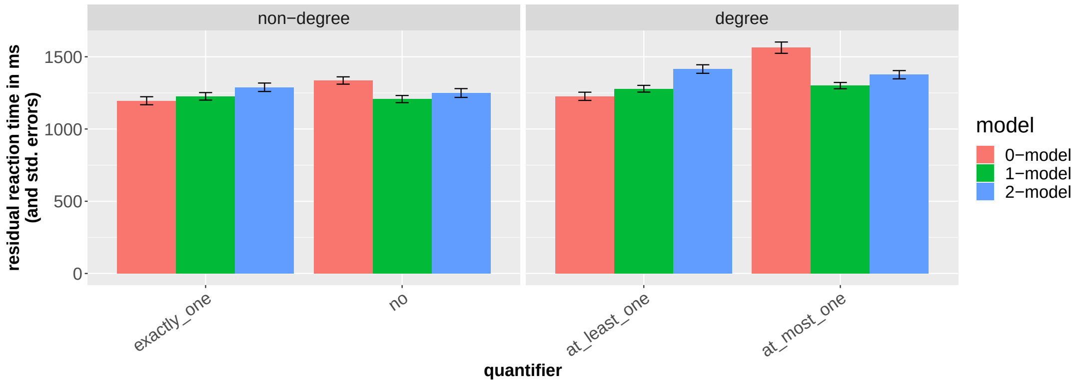
Methods: Adopting the methods from [1], a combined self-paced reading plus picture verification experiment was conducted ($N = 76$). In the web-based experiment participants first received German sentences as in (1) manipulating quantifiers in a 2×2 design. We compared degree quantifiers (*höchstens ein* ('at most one') and *mindestens ein* ('at least one')) to non-degree quantifiers (*kein* ('no') and *genau ein* ('exactly one')) and in both classes compared an ES-quantifier (*höchstens ein* and *kein*, respectively) to a otherwise similar non-ES-quantifier. After reading the sentence, participants had to verify one of the three picture conditions in Fig. 2b–2h.

In the main experiment, the four quantifiers were accompanied each by a picture that showed one marked object in the mentioned color and either zero, one or two additional objects in that same color plus objects in another color resulting in 12 experimental conditions in total. In addition, the four sentences were also combined with the pictures in Fig. 2e-2h. These control conditions violate the additivity inference of *außer*. In total, 64 items were created in the 16 conditions each and distributed together with 76 fillers across 16 lists using a Latin square. In the verification stage, both judgments and judgment RTs were recorded.

Results: Prior to further analysis, RTs below 200 ms and above 2.5 standard deviation above participants means were removed. Furthermore, judgment RTs were corrected for differences between response types ("fits" vs. "does not fit") by computing the fixed effect (-154 ms) as well as by-participants random intercepts of response type (cf. [1]) in a linear mixed effects regression (LMER) on a set of designated filler items and subtracting these effects from the corresponding RT in the main experiment. Judgment RT and accuracy across conditions are shown in Fig. 1. LMER analyses of log judgment RT revealed a significant three-way interaction between QUANTIFIER TYPE, ES-PROPERTY and MODEL ($F(3453.5, 2) = 3.08, p = 0.045$). This was due to qualitatively similar but quantitatively more pronounced effects in *degree* as compared to *non-degree* quantifiers: In both QUANTIFIERS TYPES, we found the predicted two-way interaction between ES-PROPERTY and MODEL (*degree*: $F(1650.7, 2) = 26.972; p < .001$; *non-degree*: $F(1731.1, 2) = 8.93; p < .001$). The reason for these interactions was that, for non-ES-quantifiers, additional objects in the mentioned color (cf. Fig. 2b-2d) led to an increase in RT whereas for ES-quantifiers the ES-situation caused deviation from this pattern and led to the longest RT (0- vs. 1-model in *degree* ES-quantifiers: $t = 6.06; p < .001$; and in *non-degree* ES-quantifiers: $t = 3.94; p < .001$). In the accuracy, we observed an ES-effect for the superlative ES *degree quantifier* (*at most one*), which led to only 54.7% correct responses. All the other quantifiers had accuracy above 90% in all models. This pattern in accuracy again led to significant three way interaction ($\chi^2(1) = 9.29, p = .002$) because accuracy was significantly lower in the 0-model compared to the others for *at most one* ($\beta = -1.32; z = -10.6, p < .001$) slightly lower for *no* ($\beta = -0.15; z = -0.84, p = .4$) but higher for the other two quantifiers ($\beta = 0.47; z = 2.43, p = .015; \beta = 0.43; z = 2.11, p = .035$).

The violation of the additive inference in the controls slowed down the verification process and dropped accuracy: residual RTs were between approx. 1500 and 1650 ms and accuracy was between 47% and 61 %. We interpret this as an indication that participants did not know how to deal with the violation in terms of truth, consistent with a presupposition failure.

Discussion: We take the long RT and judgments from the four control conditions as evidence that the tested *außer*-sentences do, as intended, trigger an additivity inference (e.g. that the marked object is also red) that is, however, not part of the asserted meaning but rather non-at-issue or presupposed content [cf. 4]. In the twelve experimental conditions, we found marked ES-effects for both degree and non-degree quantifiers. This indicates that these effects are related to compositional meaning and cannot be explained based on superficial features of the stimuli alone. The finding that ES-effects were larger within degree than within non-degree quantifiers is consistent with the results of [1] and may be related to the fact that the *non-degree* quantifier *kein* is the simplest possible ES-quantifier. Moreover, the absence of an ES-effect in accuracy for *kein* may indicate that this quantifier is more difficult to interpret in a Boolean combination like *none or three*, as tested by [1], than in the predication of an additive-exceptive construction. In our view, exhausting the whole compositional potential of combining domain subtraction with the ES-property (e.g. by putting ES-quantifiers inside the *außer*-phrase) is theoretically challenging and empirically promising at the same time, as it allows, e.g., for upward entailing ES-quantifiers.



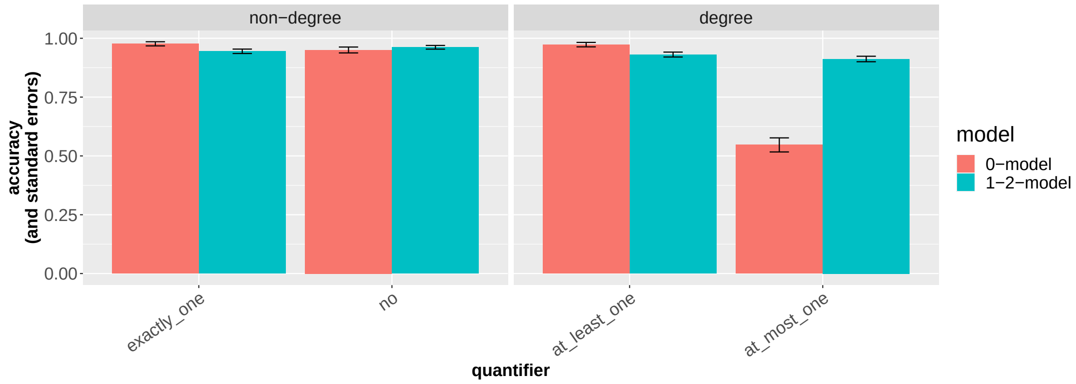
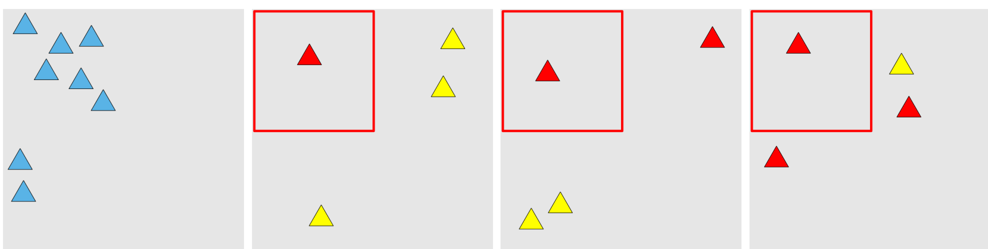


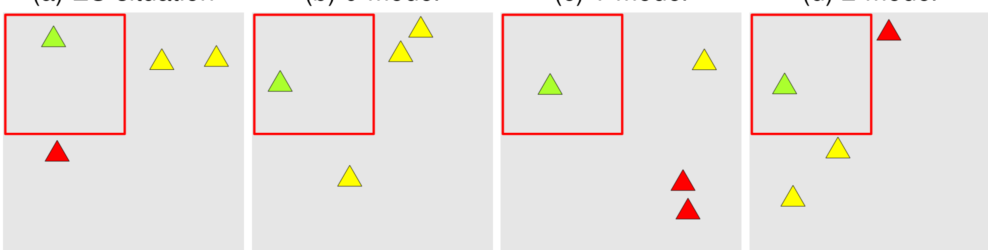
Figure 1: Verification data from target conditions in the experiment. Top: RT; bottom: accuracy.

Sample Item:

(1) Außer dem markierten | ist | Q | weiteres | Dreieck | rot.
Besides the marked | is | Q | more | triangle | red
'Besides the marked one there are Q more triangles that are red.'



(a) ES-situation (b) 0-model (c) 1-model (d) 2-model



(e) additivity control for *exactly one* (f) additivity control for *no* (g) additivity control for *at least one* (h) additivity control for *at most one*

Figure 2: Example visual stimuli

References:

1. Bott, O. et al. *Journal of Semantics* **36,** 99–163 (2019). 2. Peters, S. et al. *Quantifiers in Language and Logic* (Oxford University Press, New York, 2006). 3. Just, M. A. et al. *Journal of Verbal Learning and Verbal Behaviour* **10,** 244–253 (1971). 4. Vostrikova, E. *Phrasal and clausal exceptive-additive constructions crosslinguistically* PhD thesis (UMass Amherst, 2019). 5. Von Fintel, K. *Natural language semantics* **1,** 123–148 (1993).