Disentangling Semantic Empty-Set Effects in Quantifier Comprehension From Simple Associations: Processing Evidence From Exceptive-Additives

Nadine Balbach[1]*, Fabian Schlotterbeck[1]*, Hening Wang[1], and Oliver Bott[2]

[1] University of Tübingen, Tübingen, Germany
{nadine.balbach,fabian.schlotterbeck,hening.wang}@uni-tuebingen.de
[2] University of Bielefeld, Bielefeld, Germany
oliver.bott@uni-bielefeld.de
*Co-first authors.

Abstract

According to a recent proposal, so-called empty-set (ES) quantifiers are extraordinarily difficult to interpret because their verification algorithm involves one specific rule intended to handle ES situations. The rule relies on the explicit representation of objects in the restrictor set that are not in the scope set. This type of encoding is assumed to be cognitively costly because the cognitive system is tuned to encoding the presence of properties rather than their absence, a strategy viable for non-ES quantifiers but doom to failure for ES quantifiers. While previous data support this assumption, they do not 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. We report on a picture verification experiment in which we combined ES quantifiers with German *außer*- ('besides')-phrases to investigate how the additive inference, triggered by *außer*, interacts with ES effects. Our results indicate that ES effects are in fact based on compositional meaning and cannot be explained based on superficial features of the stimuli alone.

1 Introduction

Recently, [5] proposed an algorithmic theory of quantifier interpretation in natural language based on witness sets in order to account for a number of psycholinguistic findings. One specific distinguishing feature of this theory, e.g. in comparison to Generalized Quantifier Theory [20] and algorithmic implementations thereof in terms of semantic automata [4, 24], 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 none of the restrictor elements is also in the scope set (ES situation)), to be more difficult to interpret than quantifiers that do not have the ES among their witnesses. Since all downward entailing quantifiers are ES quantifiers, the proposal of [5] explains the well-known processing difficulty associated with downward entailing quantifiers (e.g. [14, 18, 7, 2]). 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 series of combined self-paced reading plus picture verification experiments, [5] confirmed these predictions across a range of different types of quantifiers. Moreover, they teased apart the difficulty of ES quantifiers from effects of monotonicity by establishing ES effects for the non-monotone empty-set quantifier *none or three*. In the current study, we combined ES quantifiers with the German so-called exceptive-additive *außer* ('besides', cf. [27, 28]) to investigate how the additive inference, triggered by *außer*, interacts with ES effects. This allowed us to further test the algorithmic theory of [5] by ruling out an alternative explanation of ES effects that is based on a superficial mismatch between linguistic and visual features.

2 ES effects and their explanation

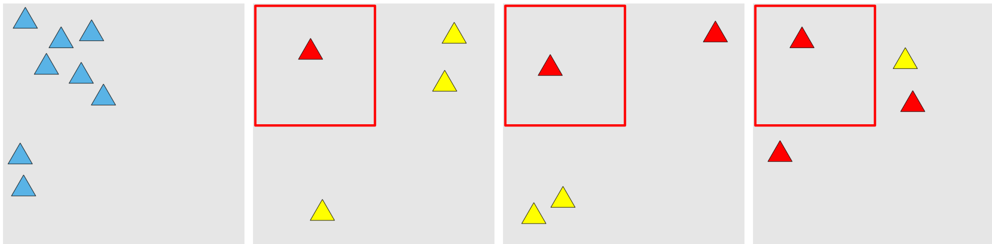
According to [5], ES quantifiers are difficult because their verification algorithm involves one specific rule intended to handle ES situations. This rule relies on the explicit representation of objects in the restrictor set that are not in the scope. [5] argue that this type of encoding is cognitively 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 doomed to failure for ES quantifiers. While the data of [5] support the assumption that the verification algorithm of ES quantifiers is costlier than the one of non-ES quantifiers, they do not 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 1a (below). The sentence is true although none of the objects has the mentioned color (i.e. none is red). The lexical meaning of the adjective *red* does not match any of the color features of the given visual context. 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 ES effects.

3 Testing empty-set effects in *außer* ('besides') sentences

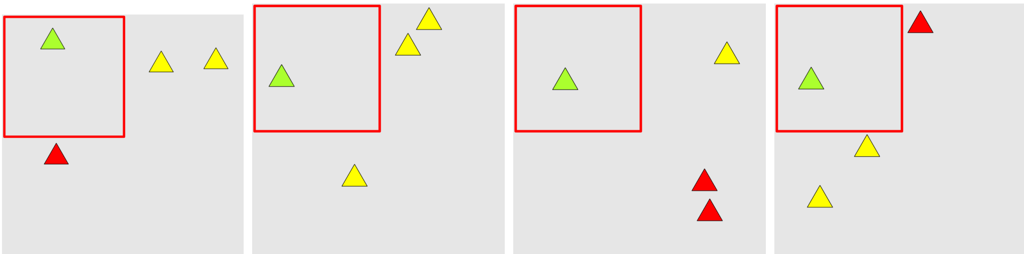
Modifying ES quantifiers with the so-called exceptive-additive *außer* ('besides') (cf. [27, 28]), allows for testing ES effects in the absence of the type of superficial mismatch just described.This is because an *außer*-phrase can explicitly mark the properties of some objects whose presence are a prerequisite for computing the truth-value but are not necessarily part of the asserted content. As a consequence, it is possible to create ES-situations without a visual mismatch between the predication and the picture. For a sentence such as (1), this is illustrated by the model in Fig. 1b, which constitutes an empty-set situation even though there are red triangles in the picture. The truth-value is determined by evaluating the quantifier, Q, with respect to all triangles except the marked one and the prerequisite is that the marked triangle is in the scope set of the quantifier (i.e. the red objects). In the present paper, we call the former meaning component the *quantificational claim* and the latter the *additive inference* to set them apart from [19]'s generality and exception claims.

(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.'

Quantifiers combined with *außer*-phrases provide also an interesting test case for the present follow-up study in terms of their inherent constructional quantificational complexity relative to the cases already investigated in [5]. Bott et al. compared ES effects in simply quantified sentences with those in doubly quantified sentences forming iterations of quantifiers [22] and observed ES effects of different strengths in picture verification. ES effects varied between less than 30% errors for ES quantifiers in simply quantified sentences as compared to more than 50% errors for ES quantifiers in doubly quantified sentences. ES quantifiers in combination with exceptive-additive phrases are therefore of additional interest from a theoretical perspective because they involve relations between three sets of entities, i.e. they are akin to quantifiers of type $\langle 1,1,1\rangle$ (modulo non-assertive status of some aspects of their meaning, cf. [19]). Considering the semantic types of the relations involved, they thus fall in between sentences with a single type $\langle 1,1\rangle$ quantifier and iterations of such quantifiers, resulting in type $\langle 1,1,2\rangle$.



(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 1: Example visual stimuli

Außer ('besides') constitutes a prime example for the postulated class of exceptive-additives. In combination with a restricted class of quantifiers (called generalizers in [19]), it can nevertheless also receive the better studied 'pure exceptive reading' (for combinatory restrictions and the analysis of 'pure exceptives', see e.g. [19, 12, 10, 9, 25, 15, 13]). In combination with the quantifiers tested in the present study (see below), *kein* ('no') is the only one that allows for a 'pure exceptive reading'. However, the additive inference is assumed to remain the same even in these cases. The only difference may consist in the fact that [27, 28] considers the inference to be a presupposition whereas others remain agnostic in this regard (see e.g. [19]). The present study, therefore, included a control experiment in addition to the main experiment testing for the additive inferences and their semantic status, that is whether the quantifiers under investigation give rise to an additive inference and whether this inference should be considered as asserted or presupposed. Last but not least, neither 'pure' exceptives nor exceptive-additives have received much attention in experimental work (for exceptives see [16, 17]) and are therefore interesting in their own right.

4 Methods

We adopted the methods from [5] and conducted a combined self-paced reading plus picture verification experiment (80 native German speakers, all self-reported students, 41 female, 39 male, recruited via Prolific). 4 of these participants were excluded from the data analysis due to fails in the attention checks or due to too many timeouts.

In the web-based experiment participants first received German sentences as in (1) employing a self-paced reading paradigm with moving window presentation (phrase by phrase, as indicated by vertical lines). After having read the final region of the sentence, the sentence disappeared and instead, a picture as in Fig. 1b-1d was shown in the middle of the screen for which a truth-value judgment had to be provided within a time limit of 10 s. In the verification stage, both judgments and judgment RTs were recorded. Reading times and picture verification latencies were preprocessed by adopting the methods from [5, Exp. 2].

As for the sentence materials, quantifiers were manipulated 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 an otherwise similar non-ES quantifier. The four sentences with quantifiers were accompanied each by a picture showing 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. 1e-1h. These control conditions violate the additive inference of *außer*. In the models of all of these four conditions the additive inference was falsified because the marked object was never in the mentioned color whereas the rest of the objects were colored in such a way to make the quantificational claim true. In total, 64 items were created in the 16 conditions each and were distributed together with 84 fillers across 16 lists using a Latin Square design.

5 Results

Prior to further analysis, reading times below 200 ms and above 2.5 standard deviations above a participant's mean RT were removed (removal of 0.2% of the data).[1] Furthermore, judgment RTs were corrected for differences between response types (expected answer: TRUE vs. FALSE) 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, [3]) on a set of designated filler items and subtracting these effects from the corresponding RT in the main experiment. Below, the resulting residual judgment RTs are reported. Throughout the study, we analyzed log judgment RTs using LMER and the accuracy of judgments using generalized (logit) mixed effects regression (GLMER, [3]) modeling. In the statistical analysis of the accuracy data, we aggregated responses across 1- and 2-models. The reason was that a comparison between 1- and 2-models is problematic because correct responses in these conditions may differ between quantifiers and, in contrast to RTs, responses themselves cannot easily be corrected for response bias, etc. For the random effects structure, we started with including by-participants and by-items random intercepts and then included random slopes for all the relevant fixed factors until the models did not converge or no longer improved model fit.

Judgment RTs and accuracy across the twelve target conditions are shown in Fig. 2. In the target conditions, log-likelihood ratio tests revealed a significant three-way interaction between QUANTIFIER TYPE, ES PROPERTY and MODEL ($\chi^2(2) = 6.31, p = .042$). This was due to qualitatively similar but quantitatively more pronounced effects in *degree* as compared to *non-degree* quantifiers: In both QUANTIFIER TYPES, we found the predicted two-way interaction between ES PROPERTY and MODEL (*degree*: $\chi^2(2) = 53.09, p < .001$; *non-degree*: $\chi^2(2) = 18.22, p < .001$). The reason for these interactions was that, for non-ES quantifiers, additional objects in the mentioned color (cf. Fig. 1b-1d) led to an increase in RT whereas for ES quantifiers the ES situation caused deviation from this pattern and led to the longest RTs (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 *degree* ES quantifier (*at most one*), which led to only 54.7% correct responses. All the other quantifiers had accuracy above 90% in all models.

[1] Data and analysis scripts are provided at https://osf.io/w8d95/.

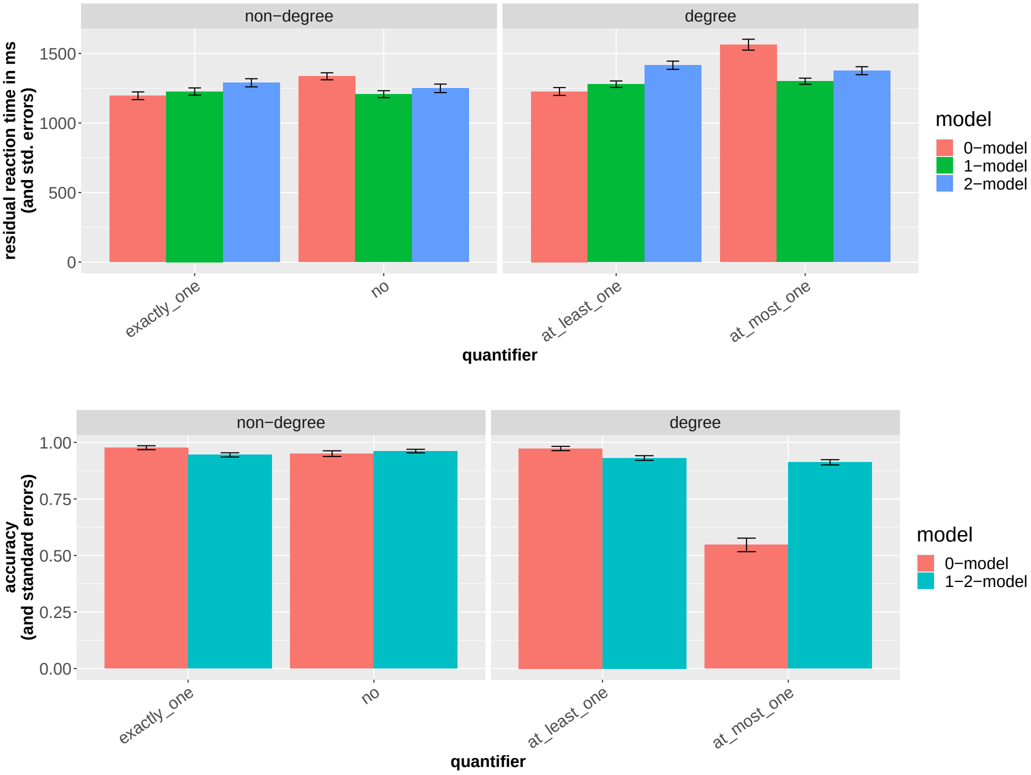


Figure 2: Verification data from target conditions in the experiment. Top: RT; bottom: accuracy.

This pattern in accuracy also led to a significant three-way interaction in a GLMER analysis with the same fixed effects and the same random effects structure as for the LMER analysis ($\chi^2(1) = 8.25, p = .004$). The three-way interaction was due to the fact that accuracy was significantly lower in the 0-model compared to the other two types of models for *at most one* ($\beta = -1.32, z = -10.6, p < .001$), not significantly different for *no* ($\beta = -0.15, z = -0.84, p = .4$), but significantly higher for the other two quantifiers ($\beta = 0.47, z = 2.43, p = .015; \beta = 0.43, z = 2.11, p = .035$). Taken together, *at most one* thus led to rather robust empty-set effects both in judgments and judgment RTs, whereas ES effects for *no* turned out to be weaker affecting only judgment RTs but not the judgments themselves.

The violation of the additive inference in the controls slowed down the verification process and decreased accuracy: Residual RTs for control conditions were between approximately 1,500 and 1,650 ms, comparable to the longest RTs among the target conditions, but there were no significant differences within the control conditions ($\chi^2(3) = 6.13, p = .105$). Similarly, accuracy was between 47% and 61 % and thus in the range of the lowest accuracy among the target conditions. The evaluation of sentences with *kein* was easier than with other quantifiers: A GLMER analysis with *kein* as reference level revealed significant differences between *kein* and all the other quantifiers (*at most one*: $\beta = -1.16, z = -3.87, p < .001$; *at least one*: $\beta = -0.87, z = -3.39, p < .001$; *exactly one*: $\beta = -0.70, z = -2.16, p = .031$).

6 Discussion

We take the long RT and neither clearly true nor false judgments from the four control conditions as evidence that the tested *außer*-sentences do, as intended, trigger an additive 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. [27, 28]). In the twelve experimental conditions, we found ES effects for both *degree* and *non-degree* quantifiers. Because the marked object always had the color mentioned in the sentence, these results indicate that ES effects are, in fact, 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 [5], who also found less pronounced ES effects for the *non-degree* quantifier *kein* ('no') as compared to modified numerals. This may be related to the fact that *kein* is the simplest possible ES quantifier in the sense that its verification needs fewer operational steps than other ES quantifiers like, e.g., the superlative *degree* quantifier *höchstens ein* ('at most one') (see also [11] on the complexity of *at most*). A direct comparison of effects across experiments is, however, problematic, especially since [5] tested *kein* in a Boolean combination like *none or three*. Nevertheless, another parallel between the present study and [5] is noteworthy: The extremely high error rate of almost 50% for the superlative ES quantifier *höchstens ein* that we observed in the present experiment resembles an accuracy even below the 50% chance level that [5] found in doubly quantified sentences containing *at most*. This parallel may point to an interaction between compositional structure of a sentence as encoded in semantic types and the cognitive complexity of the individual quantifiers it contains, specifically regarding the ES property. As pointed out in section 3, the complexity of *außer*-sentences in terms of their semantic type lies between the complexity of sentences with one single quantifier and of iterations of two quantifiers in a doubly quantified sentence. Further experiments are required to directly test for this type of interaction between different sources of complexity.

Apart from indicating an additive inference, data from the control conditions indicate that participants were uncertain how to judge violations of this inference. Assuming a presuppositional status of the inference (as, e.g., in [27, 28]), these results are consistent with the assumption that a sentence whose presupposition is not met is neither TRUE nor FALSE (see e.g. [23, 8, 26]) and with previous experimental studies that looked at falsified presuppositions (see [29, 6, 21] but also [1]). The relatively high accuracy of *kein* ('no') as compared to the other three controls may be related to the fact that this was the only quantifier among those we tested that can uncontroversially also be interpreted as a 'true' exceptive (cf. [19]), albeit with the same truth-conditional import. Alternatively, it may be explained by the fact that this was the only condition among the controls for which the quantificational claim involved a potential ES effect.

In our view, exhausting the whole compositional potential of combining domain subtraction with the ES property is theoretically challenging and empirically promising at the same time. If we put ES quantifiers inside the *außer*-phrase as in *all but at most five*, we can push the distinction between lexical features and truth-values to the extreme and study 'embedded ES effects' in situations where all elements from the restriction are also in the scope set. At the same time, *all but at most five* is an upward entailing ES quantifier that allows for 'global ES effects' in situations where the restrictor set contains fewer than five elements. In particular, this is the case if none of those elements is in the scope set. ES effects of that kind would allow us to follow up on one of the goals of [5] by completely separating monotonicity from the ES property.

Acknowledgments

We thank our student assistants Sarah Ackerschewski, Julia Nieslony, Tim Tichy, and Margarethe van Liempt for helping with the implementation of the experiment. Furthermore, we would like to thank the anonymous reviewers. A special thanks goes to Ekaterina Vostrikova for valuable feedback and discussion on exceptive-additives. Nadine Balbach received funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 833 – Project ID 75650358.

References

[1] Márta Abrusán and Kriszta Szendrői. Experimenting with the king of France: Topics, verifiability and definite descriptions. *Semantics & Pragmatics*, 6:1–43, 2013.

[2] Petra Augurzky, Fabian Schlotterbeck, and Rolf Ulrich. Most (but not all) quantifiers are interpreted immediately in visual context. *Language, Cognition and Neuroscience*, 35(9):1203–1222, 2020.

[3] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[4] Johan van Benthem. *Essays in Logical Semantics*, volume 29 of *SLAP*. Springer, Dordrecht, 1986.

[5] Oliver Bott, Fabian Schlotterbeck, and Udo Klein. Empty-set effects in quantifier interpretation. *Journal of Semantics*, 36(1):99–163, 2019.

[6] Emmanuel Chemla and Lewis Bott. Processing presuppositions: Dynamic semantics vs pragmatic enrichment. *Language and cognitive processes*, 28(3):241–260, 2013.

[7] Isabelle Deschamps, Galit Agmon, Yonatan Loewenstein, and Yosef Grodzinsky. The processing of polar quantifiers and numerosity perception. *Cognition*, 143:115–128, 2015.

[8] Gottlob Frege. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100(1):25–50, 1892.

[9] Jon Gajewski. Npi any and connected exceptive phrases. *Natural Language Semantics*, 16(1):69–110, 2008.

[10] Iván García Álvarez. *Generality and exception: A study in the semantics of exceptives*. PhD thesis, PhD dissertation, Stanford University, 2009.

[11] Bart Geurts, Napoleon Katsos, Chris Cummins, Jonas Moons, and Leo Noordman. Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes*, 25:130–148, 2010.

[12] Aron Hirsch. An unexceptional semantics for expressions of exception. *University of Pennsylvania Working Papers in Linguistics*, 22(1):16, 2016.

[13] Jack Hoeksema. The semantics of exception phrases. *Quantifiers, logic, and language*, pages 145–177, 1995.

[14] Marcel Adam Just and Patricia Ann Carpenter. Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behaviour*, 10(3):244–253, 1971.

[15] Friederike Moltmann. Exception sentences and polyadic quantification. *Linguistics and Philosophy*, 18(3):223–280, 1995.

[16] Prerna Nadathur and Dan Lassiter. An experimental look at the negative implications of exceptives. In Sherry Hucklebridge and Max Nelson, editors, *Proceedings of NELS 28*, vol. 2, pages 201–210. GLSA, 2018.

[17] Prerna Nadathur and Daniel Lassiter. Unless: An experimental approach. In Eva Csipak and Hedde Zeijlstra, editors, *Proceedings of Sinn und Bedeutung 19*, page 446–464, 2015.

[18] Mante S. Nieuwland. Quantification, prediction, and the online impact of sentence truth-value: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory,*

and Cognition, 42(2), 2016.

[19] Stanley Peters and Dag Westerståhl. The semantics of exceptives. *Linguistics and Philosophy*, Published Online, 07/08/2022.

[20] Stanley Peters and Dag Westerståhl. *Quantifiers in Language and Logic*. Oxford University Press, New York, 2006.

[21] Florian Schwarz. False but slow: Evaluating statements with non-referring definites. *Journal of Semantics*, 33:177–214, 2016.

[22] Shane Steinert-Threlkeld and Thomas F. Icard III. Iterating semantic automata. *Linguistics and Philosophy*, 36:151–173, 2013.

[23] Peter F. Strawson. On referring. *Mind*, 59(235):320–344, 1950.

[24] Jakub Szymanik. *Quantifiers and Cognition: Logical and Computational Perspectives*, volume 96 of *SLAP*. Springer, Dordrecht, 2016.

[25] Kai Von Fintel. Exceptive constructions. *Natural language semantics*, 1(2):123–148, 1993.

[26] Kai von Fintel. Would you believe it? The king of France is back! (Presuppositions and truth-value intuitions). In Marga Reimer and Anne Bezuidenhout, editors, *Descriptions and Beyond*. Oxford University Press, 2004.

[27] Ekaterina Vostrikova. Besides exceptives. In Duygu Göoksu Maggie Baird and Max Nelson, editors, *Proceedings of NELS 49*, pages 265–274, 2019.

[28] Ekaterina Vostrikova. *Phrasal and clausal exceptive-additive constructions crosslinguistically*. PhD thesis, UMass Amherst, 2019.

[29] Jérémy Zehr. *Vagueness, presupposition and truth-value judgments*. PhD thesis, École Normale Supérieure de Paris, 2014.