Putting summative predicates into context

Kurt Erbach
Goethe-University Frankfurt am Main, Germany
& Saarland University, Saarbrücken, Germany
erbach@hhu.de

Stavroula Alexandropoulou
University College London
London, UK
stavroulaale@gmail.com

Richard Breheny
University College London
London, UK
r.breheny@ucl.ac.uk

Clemens Mayr
Georg-Austin-Universität Göttingen
Göttingen, Germany
clemens.steiner-mayr1@uni-goettingen.de

Jacopo Romoli
Heinrich-Heine-Universität
Düsseldorf, Germany
jacopo.romoli@gmail.com

Yasutada Sudo
University College London
London, UK
y.sudo@ucl.ac.uk

Abstract

The present paper investigates homogeneity and non-maximality effects of summative predicates like color terms (e.g., *blue*). It specifically tests theoretical accounts that make diverging predictions as to whether summative predicates in positive and negative sentences exhibit non-maximality effects to the same extent. To adjudicate between these accounts, we conducted experiments that tested how the availability of non-maximal readings is modulated by the conversational context. Our results were largely symmetric: Non-maximality readings of summative predicates show similar context-sensitivity in simple positive ('The flag is blue') and negative sentences ('The flag is not blue'), and under the quantifiers 'every' ('Every flag is blue') and 'no' ('No flag is blue'). We discuss the theoretical consequences of these findings in light of Romoli et al.'s (2024) observation that the non-maximal readings of plural definites are asymmetric with greater context sensitivity in the scope of 'every' than in the scope of 'no'.

1 Introduction

Summative predicates like color terms give rise to homogeneity effects (Krifka 1990; Löbner 2000; Paillé 2022; 2023a; 2023b). For example the positive sentence in (1) is true iff ⟦blue⟧ is true of every part of the flag under discussion, and its negative counterpart in (2) is true iff ⟦blue⟧ is true of no part of the flag, which is a stronger meaning than the logical negation of (1) ('⟦blue⟧ is true of not every part of the flag').

(1) The flag is blue.

(2) The flag is not blue.

It is also known that summative predication sometimes allows for *non-maximality effects*. That is, for (1) to be judged as true, it is often not absolutely necessary that the entire flag be blue. For instance, in many actual conversational contexts, (3) is judged as true of the EU flag (🇪🇺), despite its yellow stars.

In recent years, homogeneity and non-maximality effects have primarily been discussed in connection with plural definites (e.g. Malamud 2012; Križ 2016; Križ and Spector 2021; Bar-Lev 2018; Bar-Lev 2021), but in a series of recent works (Paillé 2022; Paillé 2023a; Paillé 2023b) Mathieu Paillé has turned the spotlight on summative predicates, discussing multiple ways of accounting for their homogeneity and non-maximality effects in terms of exhaustification. The

different theories he considers make diverging predictions as to whether summative predicates in positive and negative sentences exhibit non-maximality effects to the same extent. To adjudicate between them, we ran experiments inspired by Romoli et al.'s (2024) study on plural definites to test how the availability of non-maximal readings is modulated by the conversational context. Our results were largely symmetric: Non-maximality readings of summative predicates show similar context-sensitivity in the simple sentences (1) and (2), and under the quantifiers *every* and *no*. We discuss the theoretical consequences of these findings in light of Romoli et al.'s (2024)'s observation that the non-maximal readings of plural definites are asymmetric with greater context sensitivity in the scope of *every* than in the scope of *no*.

2 Theoretical background

Paillé 2022 proposes to account for homogeneity effects of summative predicates in terms of exhaustification, and discusses versions of his account that make different predictions for non-maximality effects (see also Paillé 2023a; Paillé 2023b). They can be said to be either *asymmetric* in that non-maximality readings are predicted to be more robustly available in positive than negative sentences, or *symmetric* in that polarity should not affect the availability of non-maximal readings. Let us look into them in some detail.

Both versions of Paillé's theory posit an existential meaning for predicates like *blue*—e.g. $[\![\text{blue}]\!] = \lambda y.\ \exists x \sqsubseteq y[\mathsf{blue}(x)]$ (where $\sqsubseteq$ is the mereological part-of relation). The truth-conditions this analysis derives straightforwardly account for the negative sentence in (2), but are too weak for the positive sentence in (1). Paillé proposes that sentences like (1) are strengthened via exhaustification with respect to alternative sentences containing other color terms, and discusses different ways of cashing out this idea that make different predictions for non-maximality.

What we call the asymmetric version of Paillé's theory assumes the standard bivalent exhaustification mechanism Exh (Chierchia, Fox, and Spector 2012; Fox 2007 among others).

(3) $[\![\text{Exh } \phi]\!] = 1$ iff $[\![\phi]\!] = 1$ and for each innocently excludable alternative ψ to ϕ, $[\![\psi]\!] = 0$.

For (1), it delivers the following meaning by negating alternative sentences formed with other color terms: $\exists x \sqsubseteq f[\mathsf{blue}(x)] \wedge \neg\exists x \sqsubseteq f[\mathsf{yellow}(x)] \wedge \neg\exists x \sqsubseteq f[\mathsf{red}(x)] \wedge \cdots$ (where f is the relevant flag). This means that the flag only has blue parts.

This theory accounts for non-maximality effects in terms of pruning of alternatives (see Bar-Lev 2018; Bar-Lev 2021 for plural definites). If *yellow* is excluded from the set of alternatives of *blue*, for example, the exhaustified meaning will no longer entail that there are no yellow parts, allowing the flag to have some yellow parts, as in the case of the EU flag (although, as Paillé himself points out, the resulting non-maximal meaning is too weak, failing to require the flag to be mostly blue, rather than partly blue; cf. the Ukranian flag 🇺🇦). This theory is *asymmetric* because pruning of alternatives, and hence a non-maximal reading, should be a possibility only when exhaustification takes place, and by assumption, it does not take place in negative sentences like (2).

The symmetric version of Paillé's theory uses a different exhaustification mechanism that gives rise to trivalent truth-conditions, called Pexh and defined as in (4) (after Bassi, Del Pinal, and Sauerland 2021).

(4) $$[\![\text{Pexh } \phi]\!] = \begin{cases} 1 & \text{if } [\![\text{Exh } \phi]\!] = 1 \\ 0 & \text{if } [\![\phi]\!] = 0 \\ \# & \text{otherwise} \end{cases}$$

Crucially, this version of the theory always works with the full set of alternatives. With Pexh,

View	Statement		Context	
	Experiment 1	Experiment 2	Existential	Universal
Asymmetric	The flag is blue.	Every flag is blue.	✓	×
	The flag is not blue.	No flag is blue.	×	×
Symmetric	The flag is blue.	Every flag is blue.	✓	×
	The flag is not blue.	No flag is blue.	×	✓

Table 1: Predictions of two views for target sentences with respect to non-maximality.

(1) will entail that the flag only has blue parts, as before. Its non-maximality reading is explained, not in terms of pruning of alternatives, but in terms of the pragmatics of #: It is assumed a sentence denoting # may nonetheless be perceived as true depending on contextual considerations. For instance, with *the flag* referring to the EU flag, (1) semantically denotes #, but in a context where all that is relevant is whether or not the flag has any blue at all ('existential context'), it is judged as practically true, while in a context where all that is relevant is whether or not the flag is entirely blue ('universal context'), it is judged to be untrue. Turning now to the negative sentence in (2), negation flips the truth and falsity conditions and Pexh is assumed to be obligatorily present under the negation, giving rise to a trivalent meaning (we ignore the version of this account where Pexh is optional to save space). It is then predicted that a non-maximal reading should be available, except that the judgments should be reversed: (2) should be perceived as true in the universal context and untrue in the existential context.

3 Experimental background

Romoli et al. 2024 investigated definite plurals and experimentally evaluated the predictions of symmetric and asymmetric theories of non-maximality effects. Building on Križ and Chemla 2015, who investigated non-maximal readings of plural definites in positive and negative statements, they conducted a picture-sentence verification experiment manipulating both the conversational context and the polarity of the target statement (positive/negative). Their study included an existential context ("Opening the presents is prohibited") and a universal context ("Opening the presents is required"). Their findings showed that context differentially affected non-maximality interpretations of positive sentences with a plural definite in the scope of *every* (e.g., *Every girl opened her presents*) and of genuine negative sentences with a plural definite in the scope of *no* (e.g., *No girl opened her presents*). More specifically, positive conditions were more context-sensitive than negative conditions in non-homogeneous situations where, for two out of four depicted girls, some but not all of their presents were opened.

4 Experiments

We conducted experiments that are similar in design to the experiments of Romoli et al. 2024 in order to investigate how context influences the interpretation of the positive and negative sentences in (1) and (2) (Experiment 1) as well as the quantified sentences in (5) and (6) (Experiment 2), in light of the predictions of the symmetric and asymmetric views of non-maximality, as summarized in Table 1 (We assume strong Kleene quantifiers for the trivalent theory).

(5) Every flag is blue.

(6) No flag is blue.

Since the two experiments have similar materials, design, procedure, and methods, we provide a joint description in what follows.

Figure 1: Example stimuli of Experiment 1. The flag is (not) blue. (A) is True Control for positive sentences and False Control for negative sentences, (B) is False Control for positive sentences and True Control for negative sentences. (C) is the Target for both positive and negative sentences.

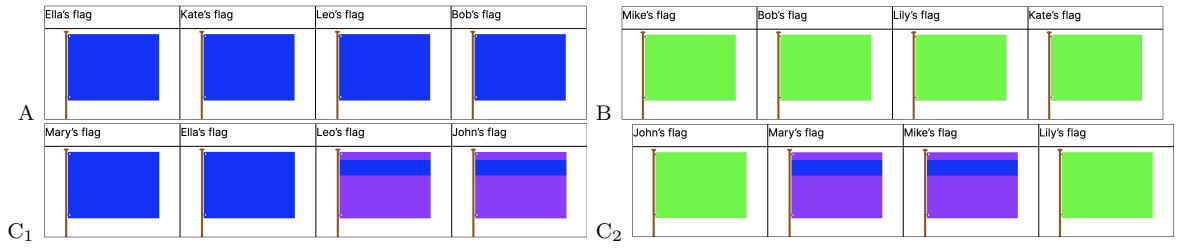

Figure 2: Example stimuli of Experiment 2. Every/No flag is blue. (A) is True Control for positive sentences and False Control for negative sentences, (B) is False Control for positive sentences and True Control for negative sentences. (C_1) is the Target for positive sentences and (C_2) is the Target for negative sentences.

4.1 Methods

Building on Romoli et al.'s (2024) study on non-maximality of plural definites, we conducted a picture-sentence verification task with a (7-point) Likert scale. In each experiment, participants judged how true the sentences were in three different pictorial situations: (i) True Control, (ii) False Control, and (iii) Target. See Figures 1 and 2 for example visual stimuli paired with the sentences in (1)/(2) and (5)/(6) in Experiments 1 and 2, respectively. True and False Control pictures depicted homogeneous situations, so the corresponding target sentences were expected to receive high and low ratings, respectively, on the response scale.

The Target pictures depicted non-homogeneous situations (see Pictures C in Figures 1 and 2). In Experiment 1, the Target picture shows a flag that is only partially blue (less than 50%), while in Experiment 2, the Target pictures (C_1/C_2) depict a situation where half of the flags are partially blue, as in Experiment 1, and the rest are uniformly one color (homogeneous flags).

It was explained in the instructions at the beginning of each experiment that the flags were drawn and colored by students in an art class, with all students being assigned the same drawing task each day by the teacher. The context, our main manipulation, was also introduced in the instructions. Similarly to Romoli et al. 2024, the context was manipulated between participants (existential vs. universal) in both experiments: In the existential context, students were instructed to create original flags, using any color except one specific color (e.g., blue), which was the same for all students. In the universal context, students were instructed to create original flags using only one specific color (blue), and no other. The teacher's instructions (i.e., the contextual manipulation) were also presented at the top of every experimental trial as a single sentence to remind participants of the context they were in (e.g., existential: "Teacher's instructions on Tuesday: Use only orange for your flag.", universal: "Teacher's instructions on Tuesday: Use any color but orange for your flag."). In both experiments, each experimental trial consisted of the teacher's instructions at the top of the display, followed by the pictorial situation and the critical statement to be verified against the depicted situation. Before making their judgement, participants were required to complete a secondary task: a yes/no question asking whether the instructions had been followed given the depicted situation. This helped ensure that participants understood the context. See supplementary materials on OSF presents example trial displays for each experiment.

Participants were randomly assigned to one of the two Context conditions, and received

the corresponding instructions. After reading the instructions, participants were presented with two practice items to familiarize themselves with the task. Both experiments had three manipulations: the Polarity of the statement (Positive/Negative), the Condition type (True Control/False Control/Target), and Context (Existential/Universal). Unlike Context, Polarity and Condition type were within-participant factors.

For each experiment, we constructed 24 experimental items: 3 flag shapes (rectangular, square, triangular) × 8 colors (blue, red, green, purple, orange, yellow, black, pink). Experimental items appeared in all 6 conditions (2 Polarity × 3 Condition type) and were rotated through six lists in a Latin square design. Every list had four items per condition. Half of the lists displayed the response button labels in one sequence (see supplementary materials on OSF), whereas the other half presented them in the reverse order. Each experiment had 12 lists in total: 6 lists with existential instructions and 6 lists with universal instructions. In addition, each list had the same 32 filler trials: 8 with color terms and 24 with other predicates such as *triangular, 5-sided, rectangular, square, circular, spotted, striped, chequered.* Both filler types were balanced for polarity (positive/negative) and truth value (true/false). These filler trials were also designed to balance the yes/no responses to the secondary question, resulting in as even a distribution as possible for both context types. Both experiments were programmed in Gorilla and administered online.

In each experiment, we recruited 288 native speakers of English raised with their native language only through Prolifc, and paid £1.50 for their participation. All participants gave their informed consent before participation. We excluded data from color-blind participants based on their response to a relevant question at the beginning of the experiment. We further excluded data from participants who had an accuracy rate of 75% or lower on the principal task in the Control items (with ratings higher than 3 for False Controls and ratings lower than 3 for True Controls being considered errors). This resulted in a final dataset of 276 participants in Experiment 1 and 278 participants in Experiment 2.

4.2 Results

We excluded trials where the secondary question was answered incorrectly from the analysis. Note that following Romoli et al. 2024, we recoded the Context factor into Lax (positive sentences in existential context, negative sentences in universal context) and Strict (positive sentences in universal context, negative sentences in existential context), with the former encouraging non-maximal interpretations and the latter discouraging them. Figure 3 shows mean verification ratings per condition in each experiment. In both experiments, the True Control conditions received high mean ratings overall, while False Control conditions received low ratings. This suggests the availability of homogeneity effects for both positive and negative sentences.

For the data analysis of both experiments, we fitted cumulative link mixed-effects models to analyze verification ratings for Target items. The models included Context (sum-coded), Polarity (treatment-coded, with negative as the reference level), and their interaction as fixed effects. They also included random intercepts for Items and for Participants, and random by-Participant slopes for Polarity. The analyses were conducted using the ordinal package (Christensen 2023) in R (R Core Team 2022).

In Experiment 1, the statistical analysis revealed a strong positive effect of Context ($\beta = 3.21$, $SE = 0.39$, $z = 8.25, p < 0.0001$), with Lax contexts receiving higher ratings than Strict ones overall. This contextual modulation was significantly greater for negative sentences compared to positive ones (significant negative Polarity×Context interaction: $\beta = -0.97$, $SE = 0.35$, $z = -2.79, p < 0.01$). In Experiment 2, the analysis revealed a positive effect of Polarity ($\beta = 0.63$, $SE = 0.30$, $z = 2.12, p < 0.05$) such that positive conditions received generally higher ratings than negative ones, as well as a robust positive effect of Context ($\beta = 3.17$, $SE = 0.38$,

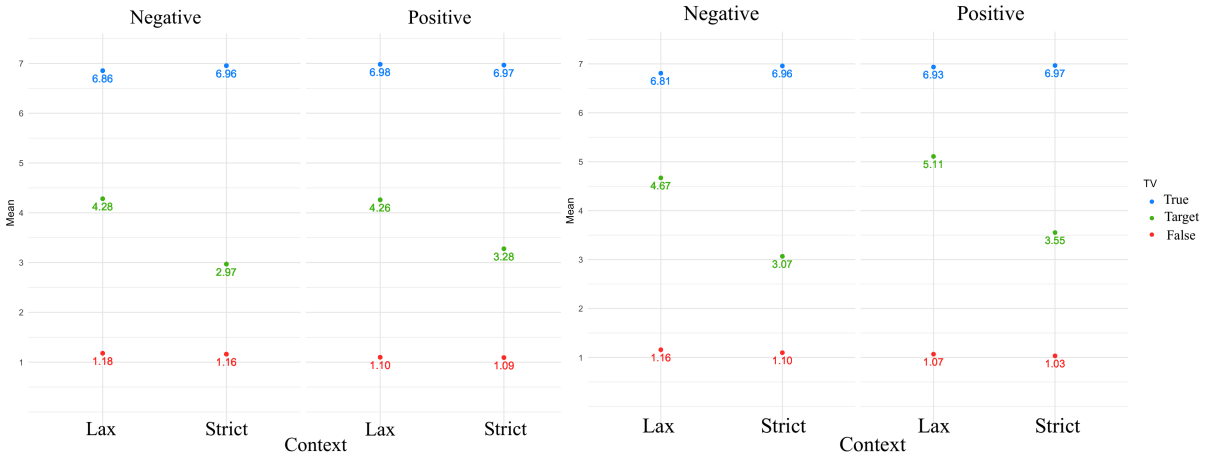

Figure 3: Mean verification ratings per condition in Experiments 1 (left) and 2 (right).

$z = 8.41, p < 0.0001$). Crucially, we found no evidence of greater context-sensitivity for non-maximal interpretations in positive sentences compared to negative ones (Polarity×Context interaction: $p = 0.23$), suggesting that the effect of Context is comparable for positive and negative sentences.

5 General Discussion

Our overall results are symmetric and hence more in line with the predictions of the symmetric view. Note in particular that the interaction effect observed in Experiment 1 was in the opposite direction of the prediction of the asymmetric view, and its effect size is relatively small.

It should also be mentioned that it is unlikely that Experiment 2 was underpowered, given that using a parallel experimental method and design Romoli et al. 2024 observed a robust interaction effect with definite plurals under *every* and *no* such that definite plurals exhibited a larger context effect under *every* (they did not test definite plurals in simple positive and negative sentences because of potential confounds with scope; cf. Križ and Chemla 2015). In order to reinforce this conclusion further, we plan to compare plural definites and summative predicates directly in a single experiment.

Assuming that the difference between summative predicates and definite plurals is real, it can be taken as suggesting that different mechanisms for non-maximality need to be postulated for summative predicates and plural definites. One possible way of theoretically implementing it is, for example, that non-maximality of summative predicates is due to Pexh, thereby predicting the symmetry, while non-maximality of plural definites is due to alternative-pruning, thereby predicting the asymmetry. More generally, our results point towards the conclusion that the class of homogeneity phenomena in natural language is heterogeneous, calling for further investigations that compare different homogeneity phenomena.

Acknowledgements. The research in this paper was supported by an AHRC-DFG grant (AHRC: AH/V003526/1; DFG: EB 523/2-1, STE 958/12-1, STE 2555/3-1).

References

Bar-Lev, Moshe (2018). "Free choice, homogeneity and innocent inclusion". PhD thesis. The Hebrew University of Jerusalem.

Bar-Lev, Moshe E. (2021). "An Implicature Account of Homogeneity and Non-Maximality". In: *Linguistics and Philosophy* 44.5, pp. 1045–1097. DOI: 10.1007/s10988-020-09308-5.

Bassi, Itai, Guillermo Del Pinal, and Uli Sauerland (2021). "Presuppositional exhaustification". In: *Semantics and Pragmatics* 14.11.

Chierchia, Gennaro, Danny Fox, and Benjamin Spector (2012). "Scalar implicature as a grammatical phenomenon". In: *Semantics: An International Handbook of Natural Language Meaning*. Ed. by C. Maienborn, K. von Heusinger, and P. Portner. Vol. 3. Mouton de Gruyter, pp. 2297–2331.

Christensen, Rune H. B. (2023). *ordinal—Regression Models for Ordinal Data*. R package version 2023.12-4.1. URL: https://CRAN.R-project.org/package=ordinal.

Fox, Danny (2007). "Free choice and the theory of scalar implicatures". In: *Presupposition and Implicature in Compositional Semantics*. Ed. by Uli Sauerland and Penka Stateva. London: Palgrave Macmillan UK, pp. 71–120.

Krifka, Manfred (1990). "Boolean and non-boolean 'and'". In: *Papers from the second symposium on logic and language*. Akadémiai Kiadó Budapest, pp. 161–188.

Križ, Manuel (2016). "Homogeneity, maximality, and *all*". In: *Journal of Semantics* 33, pp. 493–539.

Križ, Manuel and Emmanuel Chemla (2015). "Two methods to find truth value gaps and their application to the projection problem of homogeneity". In: *Natural Language Semantics* 23, pp. 205–248. DOI: 10.1007/s11050-015-9114-z.

Križ, Manuel and Benjamin Spector (2021). "Interpreting plural predication: homogeneity and non-maximality". In: *Linguistics & Philosophy* 44.5, pp. 1131–1178. DOI: 10.1007/s11050-015-9114-z.

Löbner, Sebastian (2000). "Polarity in natural language: Predication, quantification and negation in particular and characterizing sentences". In: *Linguistics and Philosophy* 23.3, pp. 213–308.

Malamud, Sophia A. (2012). "The meaning of plural definites: A decision-theoretic approach". In: *Semantics and Pragmatics* 5, pp. 1–58.

Paillé, Mathieu (2022). *Strengthening predicates*. McGill University (Canada).

— (2023a). "Alternatives and jurisdiction in predication". In: *Proceedings of Sinn und Bedeutung*. Vol. 27, pp. 483–499.

— (2023b). "Trivalent exh and summative predicates". In: *Semantics and Linguistic Theory*, pp. 421–438.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Romoli, Jacopo et al. (2024). "Putting plural definites into context". submitted.