Age stratification in collective readings of Present Day English *every*-DPs

Fabian Schlotterbeck
University of Tübingen
Tübingen, Germany
fabian.schlotterbeck@uni-tuebingen.de

Polina Berezovskaya
University of Tübingen
Tübingen, Germany
polina.berezovskaya@uni-tuebingen.de

Sigrid Beck
University of Tübingen
Tübingen, Germany
sigrid.beck@uni-tuebingen.de

Abstract

We present an experimental study on English determiner phrases (DPs) containing *every* in combination with collective sentence predicates. The experiment is based on the hypothesis that *every*-DPs are undergoing semantic change towards a referential plural interpretation. The results support this hypothesis. In particular, change in progress is visible from variable acceptability depending on the type of predicate, and an age gradation effect. In addition, the results indicate that there is an unexpected asymmetry between collective interpretations with *every*-DPs in subject vs. object position.

1 Introduction

English DPs with the determiner *every* can be used in sentences like (1). This is unexpected given the standard Generalized Quantifier semantics (cf. Barwise and Cooper 1981) of *everyone* in (2-a), predicting that the DP cannot be combined with a collective predicate like *gather*, which requires a plural subject, cf. (3). Our study tests the hypothesis that Present Day English *every* is in the process of acquiring a referential plural reading, (2-b), where *everyone* denotes the group containing all (relevant) persons (cf. Champollion 2010).

(1) a. Everyone gathered in the hallway.

(2) a. $[\![everyone]\!] = \lambda Q.\forall x[person(x) \rightarrow Q(x)]$ b. $[\![everyone]\!] = max(^*\lambda x.person(x))$

(3) a. The students / Amy, Bill and Cyril gathered in the hallway.
b. *The student / Amy gathered in the hallway.

Collective predicates can be subdivided into the classes in (4) (Brisson 2003; Dowty 1987). The corpus study and the informant judgments reported in Beck (in prep.) led us to believe that the use of *every*-DPs in sentence contexts like (1) is an innovation, and that the type of collective predicate matters for the acceptability of *every*-DPs. Beck compared a historical corpus with data up until 1914 to a modern corpus and found that plural *every*-DPs are more common today than they used to be. They occur in the modern corpus with subdistributive predicates. We conjectured that plain collective predicates, as in (4-b/c), are not (yet) acceptable with *every*-DPs today, cf. (5), which contrasts with the subdistributive predicate in (1) above. We further conjectured that the prospective change is more advanced in younger speakers than in older speakers (the age gradation or Apparent Time Hypothesis according to which such effects may be indicative of change in progress; see Labov 1963; 1966; 1994; see also e.g. D'Arcy 2005 for a more recent example study and discussion). These conjectures are visualized in Fig. 1. Finally, we test if the change is more advanced in *everyone*/*everybody* than in *every*+NP, compare (5-c).

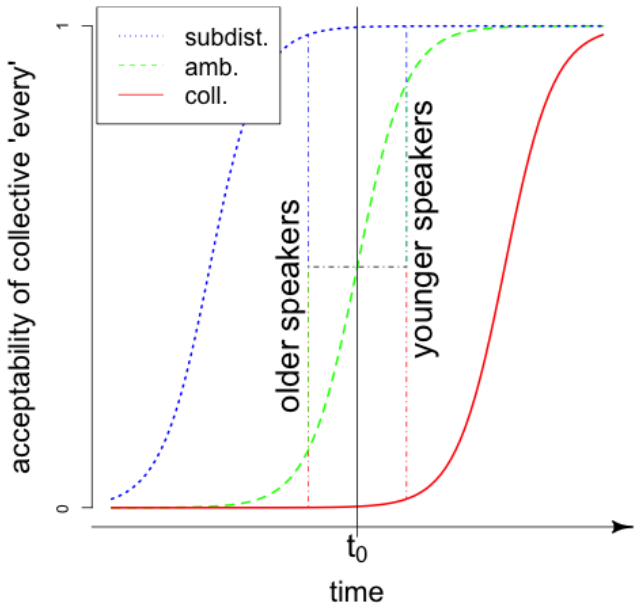


Figure 1: Inconic sketch of hypothesized stepwise change in progress for different types of predicates. coll.: plain collective; amb.: ambiguous collective; subdist.: subdistributive predicates; t_0 stands for the present day.

(4) a. The students *gathered*. (subdistributive)
b. The students are *numerous*. (plain collective (subject))
c. The workstations *outnumber* the programmers. (plain collective (sub., obj.))
d. Irene *weighed* the kittens. (ambiguous (obj.))
e. The children *bought* the house. (ambiguous (sub.))

(5) a. *Every student is numerous.
b. *The workstations outnumber every programmer.
c. ?Every programmer gathered in the hallway.

In summary the hypotheses we aimed to test are (i) that Present Day English *every* is in the process of acquiring a referential plural reading; (ii) that the type of collective predicate matters for the acceptability of *every*-DPs; and (iii) that the prospective change is more advanced in younger speakers than in older speakers.

2 Methods

To test our hypotheses, we conducted a web-based experiment in which participants read contexts and then judged the acceptability of a phrase that was targeting a crucial (in some conditions disambiguating) bit of the context (e.g. the weighing event in the example item in (6) below). Acceptability was operationalized as a binary judgment that participants provided by pressing one of two response keys. Response key assignments were counterbalanced between participants. We tested the different types of predicates in (4). In particular, we tested three predicates of each predicate type (subdistributive: *be spread out*, *gather in the hallway*, *crowd around the monitor*; collective subject: *be numerous*, *be a good team*, *be politically homogeneous*; collective object: *the workstations outnumber* _, *the committee is formed by* _, *Alex compared* _ ; collective subject of ambiguous predicate: *carry the piano upstairs*, *own the house*, *write the play*; collective object of ambiguous predicate: *I played against* _, *I weighed* _, *the lead ball outweighs* _). We embedded the 15 predicates in sentence constructions in which a collective interpretation of *everyone* (as

in (6-ii)) and *every*+NP (as in (6-iii)) was made salient by the context.[1] These constructions were compared to a clearly acceptable control with a definite plural subject (e.g., *weigh the guests* in (6-i)) and a clearly unacceptable control (e.g. *weigh the explosion* in (6-iv)). The controls served the purpose of providing a baseline of clearly acceptable and clearly unacceptable verb-argument combinations against which responses in the target conditions could be compared. We generated 15 items (3 per PREDICATE TYPE) and collected data from 199 participants in five AGE GROUPs ($18-30:40$; participants $31-40:39$; $41-50:40$; $51-60:40$; $61-75:40$) via the platform `prolific.co`. We had originally aimed for 40 participants per AGE GROUP, but one submission from the $31-40$ range was not returned. Altogether this yielded a mixed $4 \times 5 \times 5$ design in which we manipulated the three factors CONSTRUCTION (cf. (6), manipulated within items; 4 levels: *acceptable control*, *everyone*, *every+NP*, *unacceptable control*), PREDICATE TYPE (cf. (4) manipulated between items; 5 levels: collective (amb, obj), collective (amb, subj), subdistributive, collective (obj) and collective (subj)) and AGE GROUP (manipulated between participants; see above). Using a Latin square, the items were distributed across four lists in such a way that each participant saw each item (i.e. each predicate; e.g. *weigh*) in only one CONSTRUCTION. Each participant thus saw one fourth of the entire set of the materials. A set of 17 filler items from an unrelated sub-experiment were added to each list. Lists were randomly assigned to participants.

(6) Example item (with *collect. amb. obj* predicate)

Context: You own a small boat on which you take people fishing. A family of three has hired you and your boat by phone for the afternoon. When you see them at the harbour, you get worried because they are big people. Your boat can carry 320kg and you weigh 60kg. Fortunately, the harbour owns a scale. You put the three people on the scale and read off the weight. The scale says 250kg. You decide to go ahead with the trip. As you board, your cousin Erica passes by. She looks worried and asks you if you aren't taking on a bit too much.

Target: Can you say the following?

(i)	I weighed the guests.	*acceptable control*
(ii)	I weighed everyone.	*everyone*
(iii)	I weighed every guest.	*every+NP*
(iv)	I weighed the explosion.	*unacceptable control*

To analyze the binary acceptability judgments statistically and specifically test for age effects, we used a form of *penalized logistic regression* (cf. Firth 1993). In particular, we computed so-called Bayesian generalized linear mixed-effects models as implemented in the `blme` package (Chung et al. 2013). This type of analysis is similar to analyses based on logit mixed-effects models, but, in addition, allows the specification of prior distributions for Bayesian maximum a posteriori inference on fixed effects. These prior distributions can be used to penalize large absolute parameter estimates that may result from separation in the data (e.g. if there is no variation in responses in some cells of the experimental design; see Clark et al. 2023 for discussion of linguistic applications in studies involving quasi-complete separation). We included fixed effects of the factors PREDICATE TYPE, AGE GROUP and CONSTRUCTION and their interactions as well as random intercepts of items and participants (given their inclusion allowed for convergence). The fixed effect of PREDICATE TYPE was dummy coded with *subdist* as reference level, i.e. all other PREDICATE TYPEs were compared to *subdist*. For the other two factors we specified the following contrasts. For CONSTRUCTION, we compared *unacceptable controls* to the mean ratings of the other CONSTRUCTIONs (contrast *c1*), the *acceptable controls* to *everyone* (contrast *c2*), and the *acceptable controls* to *every*+NP (contrast *c3*). For AGE GROUP, we compared mean ratings in the $18-30$ and $31-40$ groups to the other groups (contrast *ag1*), the $18-30$ group

[1] Materials, data and analyses are made available at https://osf.io/gnjqt.

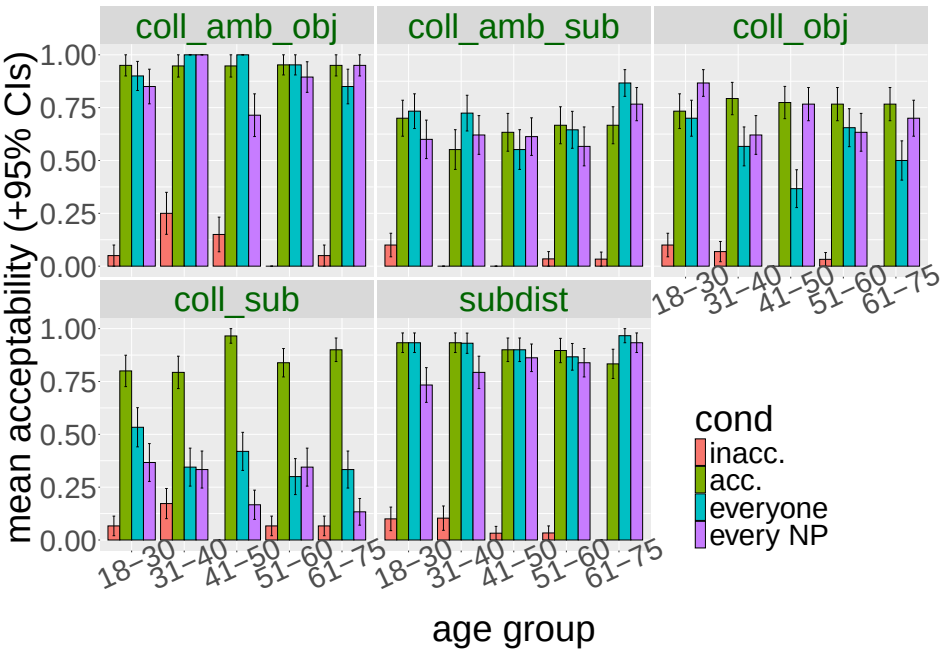
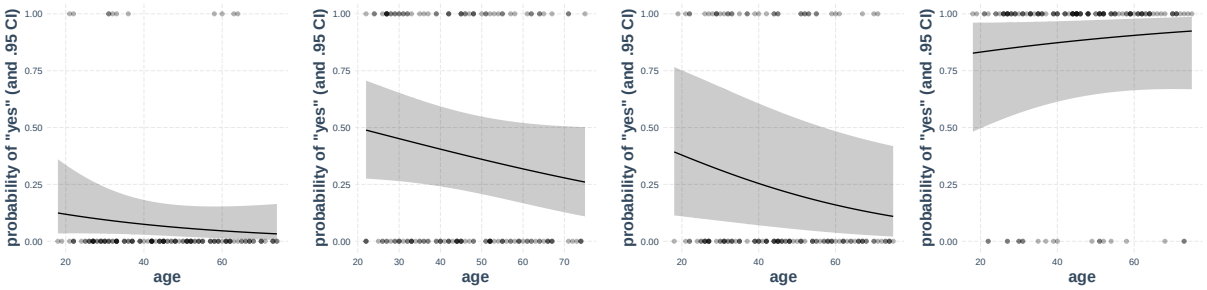


Figure 2: Mean ratings and 95% CIs by PREDICATE, AGE GROUP and CONDITION. CIs were approximated using the `summarySE` function in the `Rmisc` package (Hope 2022). acc: acceptable; inacc: inacceptable.

to the $31-40$ group (contrast *ag2*), the $41-50$ and $51-60$ groups to the $61-75$ group (contrast *ag3*) and the $41-50$ to the $51-60$ group (contrast *ag4*). In addition, we conducted two post-hoc analyses. The first post-hoc analysis focussed on the one PREDICATE TYPE were we found an indication of the age gradation effect (see results below). In this PREDICATE TYPE, we analyzed the four CONSTRUCTIONS separately and regressed response probabilities against the AGE of each individual participant (except for one who did not provide the information) as a (quasi) continuous predictor. The purpose was to further test for age gradation in acceptability. The second post-hoc analysis compared *every*+NP to *everyone* across PREDICATE TYPEs and AGE GROUPS. For this analysis, we do not report p-values because they are not independent of the contrasts tested in the main analysis.

3 Results

Mean ratings are shown in Fig. 2. There was substantial variation between predicates. The *every*-conditions patterned with the *acceptable controls* in subdistributive (88% vs. 90% on average, resp.) and ambiguous predicates (*coll amb obj.*: 91% vs. 95%; *coll amb sub.*: 67% vs. 64%;), but led to substantially lower ratings than the controls in collective predicates (*coll obj.*: 64% vs. 77%; *coll sub.*: 33% vs. 86%;). The statistical analysis revealed a marginally significant three-way interaction (model comparison: $\chi^2(48) = 61.57, p = .09$) which was driven by significant interaction terms in the fitted model that involved AGE GROUPs and CONSTRUCTIONs wrt. collective predicates (e.g. (4-b) and (4-c)). To resolve the three-way interaction, we analyzed the different PREDICATE TYPES separately. We found that *unacceptable controls* received significantly lower ratings than the other CONSTRUCTIONs across PREDICATE TYPEs (all $|z| > 7$). In addition, we observed differences between the *acceptable controls* and the universal conditions in all PREDICATEs except *coll. amb. sub.* In particular, the condition with *everyone* received lower ratings than the *acceptable controls* in *coll. sub.* ($c2 : z = 4.20, p < .001$) and *coll. obj.* ($z = 4.07, p < .001$). The condition with *every*+NP received lower ratings than



(a) *unaccaptable control* (b) *everyone* (c) *every*+NP (d) *acceptable control*

Figure 3: Results of analysis regressing response probabilities against AGE in the four CONSTRUCTIONs involving *coll. sub.* PREDICATE TYPEs. The lines show model predictions. The shaded area represents 95% confidence intervals around these predictions, which were calculated based on the fixed effects in the logit mixed effects models. The dots above and below the regression line represent the given responses.

.

the *acceptable control* in *coll. sub.* ($c3 : z = 6.88, p < .001$), *coll. amb. obj* ($z = 1.98, p = .048$) and *subdist.* ($z = 2.42, p = .015$). Moreover, there were effects of AGE GROUP in the two ambiguous PREDICATE TYPES, with relatively high ratings across constructions for the AGE GROUP $31 - 40$ in *coll. amb. obj.* ($z = 2.56, p = .011$) and for the AGE GROUP $61 - 75$ in *coll. amb. sub.* ($z = -2.64, p = .008$). Crucially, the separate analysis of PREDICATE TYPEs revealed an interaction between AGE GROUP and CONSTRUCTION but only in the *coll. sub.* PREDICATE TYPE ($\chi^2(12) = 22.95, p = .028$), where we found an increase in ratings of the *every*+NP condition as compared to the *acceptable control* for the youngest two AGE GROUPs ($c3 \times ag1 : z = -1.92, p = .055$) and an increase for the $41 - 50$ vs. the 51-60 AGE GROUP ($c3 \times ag4 : z = -2.31, p = .021$).

The first post-hoc analysis, regressing response probabilities against AGE in the *coll. sub.* PREDICATEs, revealed that the probability of accepting the target sentence decreased significantly with AGE in the *every*+NP CONSTRUCTION (one-tailed test: $\beta = -.029, z = -2.16, p = .015$; see Fig. 3c) and marginally in the *everyone* CONSTRUCTION ($\beta = -.019, z = -1.56, p = .059$; see Fig. 3b). In the control conditions, no significant association with AGE was found (*unacceptable*: $\beta = -0.026, z = -1.2, p = .116$; see Fig. 3a; *acceptable*: $\beta = 0.016, z = 1.04, p = .15$; see Fig. 3d). The second post-hoc analysis comparing *every*+NP and *everyone* directly across AGE GROUPs and PREDICATE TYPEs revealed an indication that the difference between *every*+NP and *everyone* depended on the PREDICATE TYPE ($\chi^2(4) = 21.63$) but not on AGE GROUP ($\chi^2(4) = 0.96$). Separate analyses of the PREDICATE TYPEs showed that, in *subdist* and *coll. sub.* PREDICATE TYPEs, *everyone* was judged acceptable more often than *every*+NP ($z = -2.13$ and $z = -2.4$). There was no indication of significant differences in any of the other PREDICATE TYPEs, but a numerical trend in the opposite direction in *coll. obj.* ($z = 1.74$).

4 Discussion

We observed relatively high acceptance of collective readings of *every*-DPs across the board. This indicates that the change in progress that our study anticipated is relatively advanced; i.e. in the iconic Fig. 1, the solid black bar, i.e. t_0, should be shifted towards the right in our case study. The high acceptability was, however, accompanied by substantial variation between predicate types. The observation that not all sentence contexts show acceptability of the new meaning of *every*-DPs to the same degree supports our hypothesis of change in progress. In particular, we derived from the semantic analysis developed in (Beck in prep.) that the acceptability would be higher or equal in subdistributive vs. ambiguous and ambiguous vs. plain collective predicate

types (all relative to the corresponding control conditions). The change in progress hypothesis is further substantiated by the age gradient in the acceptability of *every*+NP in the *coll. sub.* predicate type (cf. (5-b)). We predicted the existence of age gradients based on the Apparent Time Hypothesis (cf. the classical work on language change by Labov 1963; 1966; 1994). Given the assumed succession of the change (cf. Fig. 1), it was not surprising to us that we found an age gradient only in collective, but not ambiguous or subdistributive predicate types. The idea is that the change is already too advanced in the latter two types of predicates for an age gradient to be detectable. In the following paragraph, we provide a brief description of key assumptions by Beck (in prep.) that led us to expect variation between predicates. Afterwards, we address an explorative aspect of our study and provide an explanation for the observed asymmetry between plural denoting DPs in subject vs. object position. What we did not find, despite small numerical differences, was reliable evidence for further development of *everyone* vs. *every*+NP.

At the outset of the change, the new meaning of *every*-DPs requires a distributive sentence context. With Beck (in prep.), we implement this as an uninterpretable feature [u∀] on the DP. A distributive operator DIST (e.g. Link 1983) carries the required [i∀] feature. The resulting meaning is indistinguishable from a Generalized Quantifier interpretation, as shown in (7). This requirement is dropped in a stepwise fashion. Subdistributive predicates contain a distributive subcomponent. Following Brisson (2003), we analyze them as containing DIST in a subordinate position, see (8). The semantic change becomes visible when such a subordinate DIST becomes an available licensing context for *everyone*[u∀]. Speakers generally accept the plural interpretation in (8) with this weakened requirement today. Some speakers, predominantly the youngest ones, accept *every*-DPs without any distributivity in the sentence context, i.e. with plain collective predicates. In those speakers, the change is the most advanced. There is no longer any distributivity requirement, i.e. the feature [u∀] is dropped and LFs without plural operators like DIST, cf. (9), become grammatical.

(7) a. Everyone smiled.
b. everyone [u∀] [DIST[i∀] smiled]
c. $\forall x[x \leq max(*person) \rightarrow smiled(x)]$

(8) a. Everyone gathered.
b. [IP everyone [u∀] [1 [[DIST[i∀] t_1 DO] [PRO_1 be-gathered]]]]
c. $\forall x[x \leq max(*person) \rightarrow x$ did something that lead to $max(*person)$ being together.]

(9) a. ??Everyone is numerous.
b. [IP everyone [VP is numerous]]

Apart from variation between predicates, we also found differences between cases where the collective *every*-DP appears in subject vs. object position, i.e. we observe an asymmetry between subjects and objects. In particular, collective *every*-DPs seem more acceptable in object vs. subject position. This finding was not predicted. We think an explanation may be provided based on syntactic agreement. The new meaning of *every*-DP leads to a mismatch between syntax and semantics. The NP sister of *every* has to be semantically plural on the new interpretation of *every* as *max*, but it is syntactically singular. When the *every*-DP is in subject position, because of subject-verb agreement, the finite verb should also be marked as plural, but it is syntactically singular. When the *every*-DP is in object position, this second mismatch does not arise. We conjecture that the overall higher acceptability of *every*-DPs in collective object positions is due to this difference between subject and object in English. In sum, our paper presents an experimental study that applies classical methods from diachronic linguistics to a semantic phenomenon and presents a formal semantic analysis of the change in progress. Strikingly, the change targets one of the archetypal Aristotelian quantifiers in natural language, English *every*.

Acknowledgements. We thank Robin Edds for help with the English experimental items.

References

Barwise, John and Robin Cooper (1981). "Generalized quantifiers and natural language". In: *Linguistics and Philosophy* 4.2, pp. 159–219. DOI: 10.1007/bf00350139.

Beck, Sigrid (in prep.). *The Universal Semantic Cycle*. Ms., University of Tübingen.

Brisson, Christine (2003). "Plurals, all, and the non-uniformity of collective predication". In: *Linguistics and Philosophy* 26, pp. 129–184. DOI: 10.1023/A:1022771705575.

Champollion, Lucas (2010). "Cumulative readings of *every* do not provide evidence for events and thematic roles". In: *Proceedings of the 17th Amsterdam Colloquium*. Vol. 4. 2. Springer, pp. 213–222. DOI: 10.1007/978-3-642-14287-1_22.

Chung, Yeojin et al. (2013). "A nondegenerate penalized likelihood estimator for variance parameters in multilevel models". In: *Psychometrika* 78.4, pp. 685–709. DOI: 10.1007/s11336-013-9328-2.

Clark, Robert G. et al. (2023). "Dealing with complete separation and quasi-complete separation in logistic regression for linguistic data". In: *Research Methods in Applied Linguistics* 2.1, pp. 1–11. DOI: 10.1016/j.rmal.2023.100044.

D'Arcy, Alexandra (2005). "Like: syntax and development". PhD thesis. University of Toronto.

Dowty, David (1987). "Collective predicates, distributive predicates, and all". In: *Proceedings of the 3rd ESCOL*. Ed. by F. Marshall. Vol. 4. Ohio State University.

Firth, David (1993). "Bias reduction of maximum likelihood estimates". In: *Biometrika* 80.1, pp. 27–38. DOI: 10.2307/2336755.

Hope, Ryan M. (2022). *Rmisc: Ryan Miscellaneous*. R package version 1.5.1. URL: https://CRAN.R-project.org/package=Rmisc.

Labov, William (1963). "The social motivation of a sound change". In: *Word* 19.3, pp. 273–309. DOI: 10.1080/00437956.1963.11659799.

— (1966). *The Social Stratification of English in New York City*. 2nd ed. Cambridge University Press. ISBN: 978-0-521-82122-3.

— (1994). *Principles of Linguistic Change*. Oxford, UK: Blackwell.

Link, Godehard (1983). "The logical analysis of plurals and mass terms: a lattice-theoretical approach". In: *Meaning, Use, and Interpretation of Language*. Ed. by Rainer Bäuerle, Christoph Schwarze, and Arnim von Stechow. Berlin, Boston: De Gruyter, pp. 302–323. DOI: doi:10.1515/9783110852820.302.