Plurals under quantification: a comparison of English and Mandarin

Claire RONG, based on joint work with Benjamin SPECTOR

Institut Jean Nicod (CNRS, ENS-PSL, EHESS), Department of Cognitive Studies
École Normale Supérieure, Paris

Overview

- This project investigates **plural interpretation** under **universal quantification** in **mixed scenarios**, e.g. the interpretation of "Each box contains books" in scenarios where all boxes contain ≥ 1 book and some but not all boxes contain ≥ 2 books.
- Empirically, truth-value judgments reveal **gradient effects incompatible with categorical predictions** of existing theories.
- Using **statistical model comparisons**, we determine the **available readings** with the best-fitting combination of factors coding for readings.
- Readings available in the English comprehension task appear to be **different from those available in Mandarin**, a language with **optional number marking**.
- Our findings **challenge current theories of plural interpretation** and highlight **cross-linguistic variation** and **methodological challenges** in detecting semantic readings.

Plurals under universal quantification: theoretical predictions

Plural expressions **do not always denote strict plurality** (≥ 2).
They generally trigger a **multiplicity inference** in upward-entailing (UE) environments, but **the multiplicity inference disappears** in downward-entailing (DE) environments. Examples:

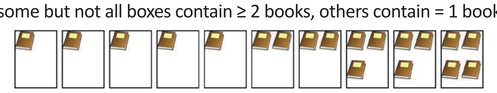
(1) The box contains books. → the bare plural suggests ≥ 2 books
(2) The box doesn't contain books. → the bare plural suggests ≥ 1 book

Puzzle of a **logical gap:** the meaning of bare plurals in UE environments is **not** the negation of their meaning in DE environments.

Two classes of **theoretical approaches** to this phenomenon:
- **Bivalent approaches** (e.g., Spector 2007, Zweig 2009):
  bare plurals have a ≥ 1 denotation, which gets pragmatically strengthened to ≥ 2.
- **Trivalent approaches** (e.g., Kriz 2017, Bassi et al. 2021):
  bare plurals have truth conditions (≥ 2), falsity conditions (zero) and are undefined for = 1.

These theories make different predictions on the truth conditions, in **mixed situations**, of bare plurals in the scope of a **universal quantifier**.
(3) Each box contains books.

Mixed situations: some but not all boxes contain ≥ 2 books, others contain = 1 book (e.g., picture below).



Three possible readings of (3):
1. **Literal** reading: each box contains ≥ 1 book.
2. **Weak** reading: some but not all boxes contain ≥ 2 books, others contain = 1 book.
3. **Strong** reading: each box contains ≥ 2 books.

Logical strengths: **strong > weak > literal**.
Methodological consequence: **some combinations of readings cannot be tested.**

Sets of readings predicted: Spector 2007 → literal, weak, strong
Zweig 2009 → strong
Kriz 2017, Bassi et al. 2021 → strong (+ literal reading is undefined)

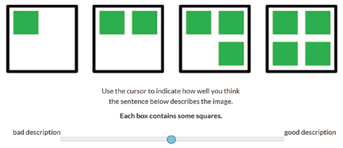
Research questions

- **Empirically, what are the available readings in English?**
- **Are available readings the same in Mandarin, a language with optional number marking?**

Experiment in English

Procedure & Stimuli

Picture paired w/ sentence **"Each box contains [some NPs]"** + continuous cursor response.



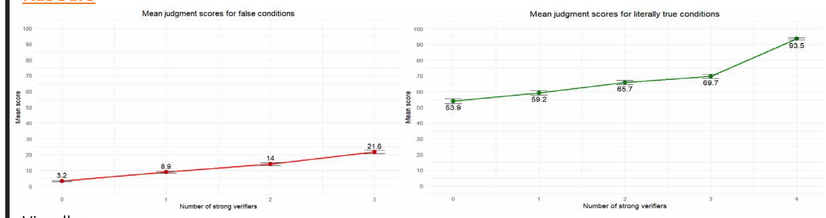
Theories make the same predictions for bare plurals and [some NPs].

A **strong verifier** = a box with multiple shapes.
Conditions labels:

FALSE	at least one box is empty;	no reading is true.
LITERAL	all boxes contain ≥ 1 shape;	only literal reading is true.
WEAK	all boxes have ≥ 1 shape, not all have ≥2;	literal + weak readings true.
STRONG	all boxes have ≥ 2 shapes;	all readings are true.

Participants: after exclusions, 200 native English speakers recruited via Prolific.
Each saw all 9 conditions 3 times, with different shape/color combinations.

Results



Visually:
- **Gradience** within a same level of reading (**FALSE** and **WEAK**) → **not predicted** by any theory.
- Qualitative shifts from **FALSE** to **LITERAL**, from **WEAK** to **STRONG**, but not from **LITERAL** to **WEAK.**
  → suggesting that **only literal and strong** readings are accessed.
  → confirmed by statistical analyses (see below).

Analyses

Predictors:

c_{vrf}	number of strong verifiers	c_{lit}	literal reading true? (binary 0/1)
c_{weak}	weak reading true? (binary 0/1)	c_{str}	strong reading true? (binary 0/1)

Linear mixed-effects model **response ~ c_{vrf} + (1 | participant)** fitted to:
- subset of literally true conditions. Compared to a null model (interception only) through LRT (likelihood ratio test): $\chi^2(1) = 1052.9$, $p < 10^{-15}$.
- subset of **WEAK** conditions alone. Also compared to a null model: $\chi^2(1) = 65.19$, $p < 10^{-15}$.

→ confirms that **c_{vrf} significantly improves model fit.**

BIC (Bayesian Information Criterion) and AIC (Akaike Information Criterion) comparison of 16 models with all possible combinations of predictors:
Best fitting-model across all 9 conditions is **response ~ c_{vrf} + c_{lit} + c_{str} + (1 | participant).**
Second-best model contains all 4 predictors. ΔBIC = 9 → strong evidence for the best model.

Discussion

- Evidence for **gradience as a factor of its own**, not merely due to more readings being satisfied.
- Model comparisons suggest the weak reading may not be accessed in comprehension
  → seems to support theories that **do not predict a weak reading**.
- This experiment was adapted into a version with a binary response option, using a logistic mixed-effects model for the analysis. The results were different: c_{vrf} is absent and the weak reading is detected in the best model when subsetting to the data of literally true conditions. However, this is subject to caution, in particular because of limitations of the logistic model.

References & Ackowledgments

Bassi, Itai, Guillermo Del Pinal, and Uli Sauerland. "Presuppositional exhaustification". Manuscript (2021).■ Cheng, Lisa Lai-Shen, and Rint Sybesma. "Bare and not-so-bare nouns and the structure of NP". Linguistic inquiry (1999).■ Chemla, Emmanuel, and Benjamin Spector. "Experimental evidence for embedded scalar implicatures". Journal of Semantics (2011).■ Chemla, Emmanuel, and Benjamin Spector. "Distinguishing typicality and ambiguities". Manuscript (2013).■ Križ, Manuel. "Homogeneity effects in natural language semantics". Language and linguistics compass (2019).■ Rong, Claire. "Plurals under quantification: new experimental perspectives". Master's thesis, supervised by Benjamin Spector, École Normale Supérieure (2025).■ Spector, Benjamin. "Aspects of the Pragmatics of Plural Morphology: On Higher-Order Implicatures". Presupposition and Implicature in Compositional Semantics (2007).■ Zhang, Niina Ning. "Expressing number productively in Mandarin Chinese". Linguistics (2014).■ Zweig, Eytan. "Number-neutral bare plurals and the multiplicity implicature". Linguistics and Philosophy (2007).

Thank you to Benjamin Spector! This project was supported by ANR grant ComCog-Mean (ANR-23-CE28-0016; PI: Benjamin Spector).

Experiment in Mandarin

Mandarin has a threefold number expression system:
- **Bare nouns: number-neutral** when used as indefinites (Zhang 2014, Cheng & Sysbesma 1999, a.o.). Bare nouns are under-specified for number and are widely **preferred in production**.
- **[one + CL + NP]: singular indefinites** (e.g., yi-gè 一个). CL = generic abbreviation for atomic classifiers. [one + CL + NP] triggers a uniqueness inference.
- **[one + *xie* + NP]: plural indefinites.** [one + *xie*] is the gloss for yi-xiē 一些, where xiē (些) is the plural classifier. [one + *xie* + NP] triggers a multiplicity inference.
  →[one + *xie* + NP] is a **weak plural**, like [some NPs] in English (shown by embedding in DE environments).
  Do Mandarin speakers access the same readings as English speakers for [some NPs]?

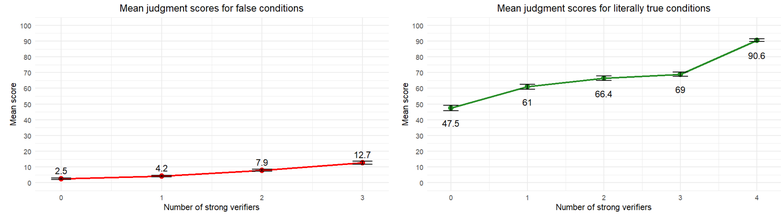
Procedure & Stimuli

Same design as the English experiment, with instructions translated into Mandarin and sentences as in (4).

(4) 每 个 盒子 里 都 有 一 些 *[NP]*
měi gè hé-zi lǐ dōu yǒu yī xiē
each CL box in DOU EXIST **one** **XIE** [NP]
Mandarin translation of "Each box contains [some NPs]".

Results & Discussion

Participants: after exclusions, 155 native Mandarin speakers recruited via Prolific and direct contact.



Same analyses as in the English experiment.
- c_{vrf} also found to have to have a significant effect.
- But model comparisons rank a different model as best: **response ~ c_{vrf} + c_{lit} + c_{weak} + c_{str} + (1 | participant)**

Why is the weak reading detected in Mandarin but not in English?
Maybe [one + *xie*] in Mandarin elicits more fine-grained levels of interpretation than English [some NPs].
Two possible explanations:
- In a threefold number marking system, competition with the bare noun may lead classifier-marked singulars and plurals to be used in more specific contexts.
- Differences in participant populations in the two languages.

Conclusion

Answers to our research questions

1. **Empirically, what are the available readings in English?**
   The best-fitting model for the English data contains the c_{lit} and c_{str} factors, and not the c_{weak} factor. However, we do not have conclusive evidence for or against the weak reading being accessed. Detection of the weak reading is methodologically challenging, because we need to disentangle readings from gradient effects.

2. **Are available readings the same in Mandarin, a language with optional number marking?**
   The best model fitting the Mandarin data showed support for the weak reading. Further theoretical work is necessary to understand the link between optional number marking systems and possible availability of the weak reading.

Perspectives

1. Investigate the **source of gradient effects** (*distance* to closest situation that makes a certain reading true? See Chemla & Spector 2013).
2. Link **continuous and binary** responses: is there a model predicting, from the continuous responses data, the responses of a binary version of the task? Does the continuous response option prompt the emergense of gradient effects? (preliminary results in Rong 2025)
3. Link **language production and language comprehension**: do speakers behave in a Bayesian way across production and comprehension? (preliminary results in Rong 2025)
4. Exploring **other lexical scales and quantifiers** (Chemla & Spector 2011).
5. Examining **context-sensitivity** through manipulation of the Question Under Discussion (QUD).