Zero and L-triviality

Andreas Haida
Hebrew University of Jerusalem
andreas.haida@mail.huji.ac.il

Tue Trinh
Leibniz-Zentrum Allgemeine Sprachwissenschaft
tuetrinh@alum.mit.edu

1 A puzzle

Bylinina & Nouwen (2018) assume that numerals have the weak, 'at least' meaning as basic and the strong, 'exactly' meaning as derived by way of exhaustification, and propose that **zero** be considered no different from other numerals in this respect. See (1b) for illustration.[1]

(1) a. zero students smoked $\Leftrightarrow$
$\left| [\![\text{students}]\!] \cap [\![\text{smoked}]\!] \right| \geq 0$
'the number of students who smoked is zero or greater'

b. exh [zero students smoked] $\Leftrightarrow$
$\left| [\![\text{students}]\!] \cap [\![\text{smoked}]\!] \right| \geq 0 \wedge \left| [\![\text{students}]\!] \cap [\![\text{smoked}]\!] \right| \ngeq 1$
'the number of students who smoked is zero and not greater'

A consequence of this semantics is that a non-exhaustified **zero**-sentence is trivial. Thus, (1a) is true in all situations. In light of this fact, Bylinina and Nouwen claim that **zero**, unlike other numerals, requires non-vacuous exhaustification, as trivial sentences are "semantically defective."[2] This claim may be said to receive support-

[1] Assuming that **zero** alternates with other numerals, i.e. that the set of alternatives in this case is {**one student smoked**, **two students smoked**, ...}.

[2] A semantic representation of **zero students smoked** which would more immediately reflect Bylinina and Nouwen's analysis is (i).

(i) $\exists x([\![\text{zero}]\!](x) = {}^{\times}[\![\text{students}]\!](x) = {}^{\times}[\![\text{smoked}]\!](x) = 1)$

These authors assume that the linguistic ontology, i.e. D_e, contains a "bottom element" $\perp$ such that $[\![\textbf{zero}]\!](\perp) = 1$ and ${}^{\times}[\![\alpha]\!](\perp) = 1$ for each expression α of type $\langle e,t\rangle$. We prefer the equivalent but ontologically less controversial representation in (1a), as the discussion below does not hinge on the existence of $\perp$. Note, also, that in both representations, **zero** ends up being trivially downward and upward entailing in its NP as well as in its VP argument. This means that both representations are compatible with Bylinina and Nouwen's explanation for **zero**'s inability to license NPIs.

ing evidence from the deviance of (2a), which can be explained as resulting from exhaustification being made vacuous by the superlative adverb **at least**.[3]

(2) a. #at least zero students smoked
b. exh(2a) $\Leftrightarrow$ (2a) $\Leftrightarrow$ $\left| [\![\text{students}]\!] \cap [\![\text{smoked}]\!] \right| \geq 0$

Suppose Bylinina and Nouwen's claim is correct, a question that arises is (3).

(3) Is the requirement for non-vacuous exhaustification invoked by **zero** pragmatic or grammatical?

The answer given by Bylinina and Nouwen is that it is pragmatic. To quote from Bylinina & Nouwen (2018: 10): "Unlike other numerals, **zero** invokes exhaustification obligatorily. This is for purely pragmatic reasons."

We believe this answer is problematic: if the requirement were purely pragmatic, all syntactic forms expressing the weak meaning should be equally deviant, but this is not the case, as evidenced by the acceptability of (4), which is semantically equivalent to (2a).

(4) zero or more students smoked

Both (2a) and (4) are trivial and thus equally uninformative, but only (2a) is deviant. This suggests that the contrast between these two sentences is structural in nature. Let us consider their respective logical form, at the level of detail which is sufficient for the purpose at hand.[4]

(5) a. [$_S$ at least zero students smoked]
b. [$_S$ zero students smoked] or [$_S$ more than zero students smoked]

We distinguish between non-logical and logical terms in (5) by underlining the former.[5] This distinction serves to bring out a crucial difference between (5a) and (5b): although both sentences are trivial, the first is "L-trival" while the second is not. Let us turn to a brief discussion of L-triviality.

[3] Assuming that **at least** alternates with **exactly** and **more than**, i.e. that the set of alternatives in this case is {**exactly zero students smoked**, **more than zero students smoked**} (Kennedy 2015; Buccola & Haida 2018). We presuppose the definition of **exh** proposed in Fox (2007).

[4] We assume that (4) results from the binary disjunction (5b) by way of PF-deletion.

[5] For this discussion, we assume the following criterion which, we believe, underlies all characterizations of the distinction between non-logical and logical terms: non-logical terms have world-dependent while logical terms have world-independent denotation. It is obvious that the denotation of **students** and **smoked** in a certain world depends on what the facts are, specifically who is a student and who smoked, in that world. In contrast, it is not obvious that the denotation of **at least**, **zero**, and **more than** depends on what the facts are. We will assume that it does not, and thus, that these words belong to the logical vocabulary of English.

An L-trivial sentence is one which is trivial by virtue of its logical skeleton. Specifically, if we replace every occurence of every non-logical term in the LF of a sentence ϕ with a different variable and the result is trivial under all variable assignments, then ϕ is L-trivial (cf. Gajewski 2003). Take (5a), for example. Replacing every occurence of every non-logical term in this structure with a different variable results in (6).

(6) $[_S$ at least zero P_1 $P_2]$

As we can see, no matter which predicates are assigned to P_1 and P_2 in (6), the result will still be trivial. Thus, suppose P_1 is mapped to 'teachers' and P_2 to 'sang,' the result will be 'at least zero teachers sang,' which is trivial. The situation is different, however, in the case of (5b). Replacing every occurence of every non-logical term in this structure with a different variable yields (7).

(7) $[_S$ zero P_1 $P_2]$ or $[_S$ more than zero P_3 $P_4]$

This structure will not be trival under the assignment $\{\langle P_1, \text{'students'}\rangle, \langle P_2, \text{'smoked'}\rangle, \langle P_3, \text{'students'}\rangle, \langle P_4, \text{'sang'}\rangle\}$, as (8) is not trivial: it is false if some students smoked and no students sang.[6]

(8) $[_S$ zero students smoked] or $[_S$ more than zero students sang]

L-triviality has been argued to cause deviance (Barwise & Cooper 1981; Fintel 1993; Gajewski 2003; Abrusán 2007; Gajewski 2009). The fact that (5a) is L-trivial while (5b) is not, then, can be taken to explain the contrast between these two structures, i.e. between the two sentences (2a) and (4).[7]

But there is a problem. Note that L-triviality is a formal concept: a sentence containing an L-trivial constituent is expected to be grammatically deviant, whether or not that sentence, as a whole, is uninformative and thus pragmatically deviant.

[6] Assuming that the first disjunct **zero students smoked** is understood in its exhaustified meaning (cf. Hurford 1974; Chierchia et al. 2012).

[7] The following contrast might raise doubt about the deviance of **#at least zero** being due to L-triviality, as it is implausible that **Celsius** and **Kelvin** are part of the logical vocabulary.

(i) a. The temperature is at least zero degrees Celsius
b. #The temperature is at least zero degrees Kelvin

However, we do not need to assume that **Celsius** and **Kelvin** are logical terms to explain this contrast. What we can say is that the presence of **Celsius** vs. **Kelvin** affects whether **zero** denotes the lower endpoint of the scale or not. Our discussion on **zero** is premised on the understanding that it does denote such a point. Non-standard readings in which it doesn't, as in (ia), are not relevant.

This is why (9a) is unacceptable, even though exhaustification, due to the presence of the universal modal, would make it contingent.[8]

(9) a. #John is required to talk to at least zero students
b. exh(9a) ⇔ required(at least zero) ∧ ¬required(exactly zero) ∧ ¬required(more than zero) ⇔ John is not required to talk to exactly zero students ∧ John is not required to talk to more than zero students

Further evidence for L-triviality being a formal requirement is the fact that (10a) is deviant, even though exhaustification would make it contingent, assuming that **every** alternates with **some**.[9]

(10) a. #there is every student
b. exh(10a) ⇔ there is every student ∧ ¬there is some student ⇔ there is no student

But if unexhaustified **zero** causes deviance by way of the formal constraint against L-triviality, we would expect that exhaustification cannot alleviate this deviance. In other words, we would make the wrong prediction that (1a) is as bad as (9a) and (10a).

The puzzle, then, is this: if **zero** requires exhaustification to circumvent deviance, that requirement must be grammatical and not pragmatic, but if that is the case, exhaustification should not be able to circumvent the deviance.

2 Some quantitative data

2.1 Existential sentences

To corroborate the intuition that **zero**, unlike other numerals, cannot be modified by the adverb **at least**, we conducted an experiment, hosted on Amazon MTurk, in which participants were asked to rate the *naturalness* of eight English sentences. The sentences were derived from the sentence frames in (11) by replacing the placeholder *n* with the numeral **zero**, for a set of four deviant sentences, and the numeral **two**, for a set of four non-deviant sentences.

(11) a. there are at least *n* students in the seminar room

[8] Again, we assume that **at least** alternates with **exactly** and **more than** (see note 3).

[9] The triviality of (10a) is shown in (i), where ⟦**there**⟧ is taken to be the set E of all entities and thus $P \subseteq$ ⟦**there**⟧ for all predicates P (Barwise & Cooper 1981).

(i) (10a) ⇔ $\forall x(x \in$ ⟦**students**⟧ $\rightarrow x \in$ ⟦**there**⟧$)$

b. the drawer contains at least n towels
c. the company hired at least n employees
d. the bartender served at least n guests

We asked participants to rate the resulting sentences, presented in pseudo-randomized order, on the following four-point Likert scale: 4 (natural), 3 (relatively natural), 2 (relatively weird), 1 (weird). To illustrate, in advance of the trials, how these values may be associated with English sentences, we gave participants the four sentences in (12) and commented: "You may agree that (12a) is natural, (12b) is weird, while (12c) and (12d) are possibly somewhere in between."[10]

(12) a. Everyone but John came to the party
b. Someone but John came to the party
c. Everyone who is not John came to the party
d. Someone who is not John came to the party

We used an even-numbered scale to force participants to discriminate between sentences which they may be unable to judge as 4 ('natural') or 1 ('weird'), and avoided suggesting that there is agreement on the score of such sentences in order not to stifle participants' trust in their own judgments. We believe that the use of a four-point scale, excluding a neutral value, does not have an adverse effect on the interpretation of the expected findings of our study: even if there was perceived social or other pressure (not) to classify sentences as '(relatively) natural' or '(relatively) weird', potential skewing would not eliminate the expected difference in the score for sentences with **at least zero** and sentences with **at least two**.

We conducted our experiment with 32 Amazon MTurk workers who, after the trials, identified as being native speakers of English. Thus, overall we received 32 scores for each of our eight sentences and hence 128 scores per sentence type. Figure 1 shows that sentences with **at least two** received the highest score 4 ('natural') by $\geq 50\%$ of all subjects, while sentences with **at least zero** received the two lowest scores 2 ('relatively weird') and 1 ('weird') by $\geq 50\%$ of all subjects. The difference in the means of the scores (3.4 v 2.0), depicted in Figure 2, is highly significant ($p < 2.2^{-16}$).

[10] The acceptability contrast between (12a) and (12b) is discussed and explained in Fintel (1993).

Figure 1: Boxplot of **at least two** & **at least zero**.

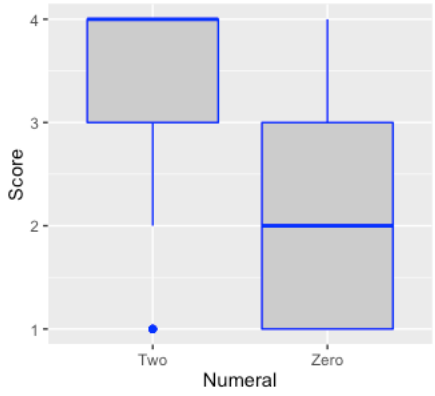
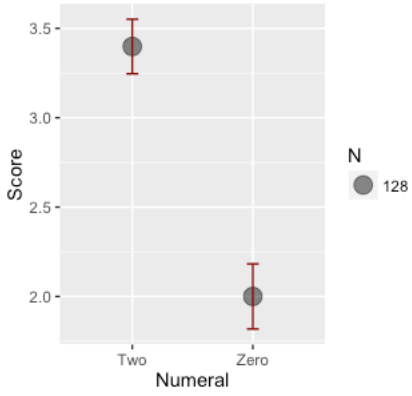


Figure 2: Means of **at least two** & **at least zero**.



This result supports our empirical claim that numerical statements containing **#at least zero**, unlike statements containing other superlative modified numerals, are deviant.

2.2 Universally quantified sentences

If the deviance induced by the triviality of **zero** is pragmatic, it should be alleviated by exhaustification, and **at least zero** should be acceptable under universal quantification, as such quantification renders exhaustification non-vacuous. Example (9a) suggests that the deviance of **#at least zero** persists under universal quantification. To corroborate this intuition, we conducted another experiment, again hosted on Amazon MTurk. This experiment targeted the contrast between **#every . . . at least zero** and its counterpart containing **zero or more** as well as the non-zero counterparts of the former two. We used the eight sentences in (13) as stimuli.

(13)	a.	(i)	#every human has at least zero children	(TRUE)
		(ii)	#every human has at least zero biological mothers	(FALSE)
	b.	(i)	every human has zero or more children	TRUE
		(ii)	every human has zero or more biological mothers	FALSE
	c.	(i)	every human has at least one biological mother	FALSE
		(ii)	every human has at least two relatives	TRUE
	d.	(i)	every human has one or more biological mothers	FALSE
		(ii)	every human has two or more relatives	TRUE

Two of these sentences, viz. the sentences in (13a), are deviant, while the others are non-deviant. Abstracting from the deviance of (13a-i) and (13a-ii),[11] we can characterize the meaning of the sentences in (13) in the following way. All sentences license a distributive inference.[12] For instance, the sentence in (13b-i) licenses the inference that some humans have (exactly) zero children and some humans have one or more children. The distributive inference renders (13b-i), as well as the other sentences (13a) and (13b), non-tautological. The distributive inference of (13b-i) doesn't, however, render this sentence false in the actual world. In this regard, it differs from e.g. the sentence in (13a-ii): in the actual world, it is false that some humans have (exactly) zero biological mothers. Overall, as indicated, if we take the distributive inference into account four of the eight sentences in (13) are true in the actual world, while the other four are false.[13]

We instructed participants to classify the sentences in (13) as either 'true', 'false', or 'weird'[14] by giving them the following instruction in advance of the trials: "Imagine you are helping an alien who wants to learn English and also learn about humans in general. Your job is to say for several English sentences whether they are true, false, or sound weird." The scenario we sketched in this instruction served to introduce an addressee without common knowledge. With this, we aimed for subjects to be less drawn to classify non-deviant sentences such as the sentences in (13b-i) or (13c-ii) as weird despite the fact that they express common knowledge with little informational content.[15] At the same time, the instruction aimed at making subjects more inclined to classify the deviant sentences in (13a) as weird by pointing out that the addressee also wants to learn English.

We expect that the proportion of 'weird' responses is higher for the sentences in (13a) than for the sentences in (13b), in accordance with our empirical claim that the former, in contrast to the latter, are deviant. Furthermore, we expect that the proportion of 'weird' responses is higher for the sentences in (13a) and (13b) than for the sentences in (13c) and (13d). This expectation is based on two factors: (i) the assumption that tautological sentences are more prone to being judged as weird than contingent sentences, and (ii) the experimental finding that there tends to be a substantial subpopulation of subjects that do not to compute distributive inferences

[11] That is, we consider these sentences as if language was not sensitive to L-triviality.

[12] This means that the universal quantifier **every** behaves like the universal modal **require** in licensing this inference. See the discussion of example (9a) for how it is derived. Furthermore, as pointed out in Büring (2008), **at least** *n* licenses the same inferences as *n* **or more**.

[13] Again, we are abstracting from the deviance of the sentences in (13a).

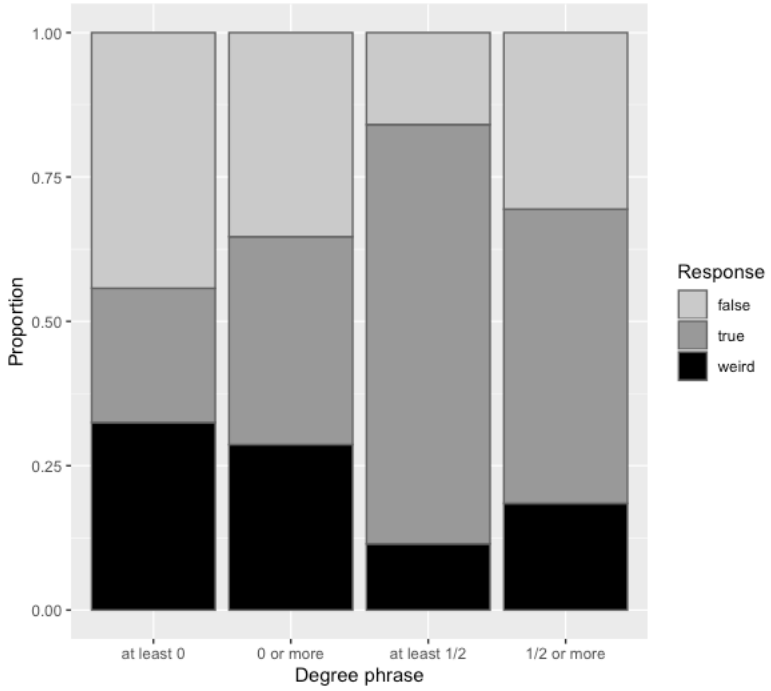
[14] A pilot study using the four-point Likert scale in § 2.1 failed to elicit discriminative judgments for the contrast between (13a-i) and (13b-i). That is, both sentences received equally low scores. We believe that different factors are at play for the two sentences to be judged as (relatively) weird on the four-point scale we used. See the discussion below.

[15] Recall that (13b-i) has a distributive inference which renders this sentence non-tautological.

(Crnič et al. 2015). Such a subpolulation derives tautological truth conditions for the sentences in (13a) and (13b), while the sentences in (13c) and (13d) are contingent with or without the distributive inference.

We conducted our experiment with 157 Amazon MTurk workers who, after the trials, identified as being native speakers of English. Thus, we received 157 scores for each of our eight sentences (1256 observations in total). Figure 3 shows the proportion of 'true', 'false', and 'weird' responses for the degree phrases in (13a), (13b), (13c), and (13d), respectively.

Figure 3: Proportion of True-False-Weird judgments.



In the subsequent statistical analysis, we consider as dependent variable the binary distinction between 'weird' responses and 'true' or 'false' responses ('non-weird' responses). We used binomial logistic regression to analyze the relationship between this binary *response* variable and the following four independent variables: *modifier* with values 'at least' and 'or more', *numeral* with values 'zero' and 'non-zero', *at least zero* with values '+' and '-', and *truth value* with values 'true' or 'false'.[16] All four variables turn out to be significant predictors of the *response* variable with p values of 0.01449, 0.00273, 0.01984, and 0.00995, respectively.

[16] These variables describe whether the stimulus sentence contains the modifier **at least** or the modifier **or more**, whether the stimulus sentence contains the numeral **zero** or the numeral **one** or **two**,

Specifically, the modifier **at least** and the truth value 'true' *de*crease the odds of the 'weird' response (reduction of the odds by 43% and 28%, respectively), while the numeral **zero** and the degree phrase **at least zero** *in*crease the odds of the 'weird' response (increase of the odds by 78%, and 110%, respectively). The results of the ANOVA between this model and the null model show that, as suggested by the p values of the logistic regression, the *numeral* variable is by far the biggest contributor to the fit of the model (reduction of deviance by 44.4 from 1347.7 null deviance, $p = 2.686^{-11}$), followed by the *at least zero* variable (reduction of 6.7, $p = 0.0094$), and *truth value* (reduction of 5.2, $p = 0.0221$), while the *modifier* variable does not contribute significantly to the fit of the model ($p = 0.5$).
These results confirm the expectation that the proportion of 'weird' responses is higher for the sentences in (13a) and (13b) than for the sentences in (13c) and (13d), and higher for the sentences in (13a) than for the sentences in (13b). Thus, overall the results are compatible with our empirical claim that the sentences in (13a) are deviant, while those in (13b), (13c), and (13d) are non-deviant.[17] Questions remain about the influence of the truth or falsity of the stimulus sentence on the odds of the 'weird' response and about the reason for the overall tendency that for modifier **or more** to raise the odds of the 'weird' response compared to the modifier **at least**.[18] We hope to address these questions in future research.

3 Conclusion

Bylinina & Nouwen (2018) propose the 'at least' meaning as basic for **zero** to account for a number of intuitions about this numeral. However, there is a fact about **zero** which seems puzzling for the proposal: **zero** is incompatible with **at least**. At first, it seems that this incompatibility can be made to follow from Bylinina and Nouwen's claim about exhaustification, namely that it can alleviate deviance induced by triviality. That claim turns out to be problematic upon closer inspection: the kind of triviality that causes deviance cannot be circumvented by exhaustification. The task remains, then, of formulating an adequate semantics for **zero**. It is expected that the accomplishment of this task will also inform our understand-

whether the stimulus sentence contains the degree phrase **at least zero** or any other degree phrase, and whether the stimulus sentence is true or false in the actual world, respectively.

[17] For further corroboration, we note that an informal Google search on September 14, 2019, shows a clear difference in the production of the expressions under consideration: a search of, e.g., the phrase **zero or more times** gives 237,000 results, while a search for **at least zero times** only gives 453 results.

[18] Recall that, importantly, our results also show that the odds are reversed for **at least zero** relative to other degree phrases.

ing of exhaustification and triviality, and more generally, of the interplay between grammar and pragmatics.

References

Abrusán, Martha. 2007. *Contradiction and Grammar: the Case of Weak Islands*: MIT dissertation.

Barwise, Jon & Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4(2). 159–219.

Buccola, Brian & Andreas Haida. 2018. Obligatory irrelevance and the computation of ignorance inferences. To appear in *Journal of Semantics*.

Büring, Daniel. 2008. The least *at least* can do. In C.B. Chang & H.J. Haynie (eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 114–120. Somerville: Cascadilla Press.

Bylinina, Lisa & Rick Nouwen. 2018. On "zero" and semantic plurality. *Glossa: a journal of general linguistics* 3(1). 1–23.

Chierchia, Gennaro, Danny Fox & Benjamin Spector. 2012. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In Paul Portner, Claudia Maienborn & Klaus von Heusinger (eds.), *Semantics: An International Handbook of Natural Language Meaning*, De Gruyter.

Crnič, Luka, Emmanuel Chemla & Danny Fox. 2015. Scalar implicatures of embedded disjunction. *Natural Language Semantics* 23. 271–305.

Fintel, Kai von. 1993. Exceptive constructions. *Natural Language Semantics* 1(2). 123–148.

Fox, Danny. 2007. Free choice disjunction and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*, 71–120. Palgrave-Macmillan.

Gajewski, Jon. 2003. L-analyticity in natural language. Unpublished manuscript.

Gajewski, Jon. 2009. L-triviality and grammar. Unpublished manuscript.

Hurford, James R. 1974. Exclusive or inclusive disjunction. *Foundations of Language* 11. 409–411.

Kennedy, Christopher. 2015. A "de-Fregean" semantics (and neo-Gricean pragmatics) for modified and unmodified numerals. *Semantics and Pragmatics* 8(10). 1–44.