Logicality within an empirically-grounded typology of unacceptability judgments

Introduction: It is commonly assumed that speakers' unacceptability judgments are categorized into different types. These include syntactic ungrammaticality and semantic anomaly ([3]), as well as semantic/pragmatic infelicity given a context ([6]). Although there is extensive work on the empirical validation of syntactic acceptability judgments (e.g., [7, 8]), fewer works investigate the extent to which the different types of unacceptability reported in the theoretical literature track speakers' intuitions (see [1] for a survey). Of particular interest is the so-called Logicality cases, such as unlicensed NPIs, where logical-semantic factors have been claimed to lead to ungrammaticality (e.g. [5, 2, 4]). The assumption that these cases are categorized as ungrammatical rather than semantic infelicity is not thoroughly empirically validated. In this paper, we report on an experiment that is aimed at providing an empirically grounded typology of unacceptability, using judgment data from an array of prompts. Our results clearly place Logicality in the category of syntactic phenomena, distinct from both semantics and pragmatics, in line with the theoretical literature.

Methods: 120 US-residing English monolinguals were recruited on Prolific and paid £1.75. Each participant answered one of the 6 prompts in (1) about 75 sentences, using a slider. Prompts and associated instructions were between-subjects, but items were the same for all subjects. The sentences spanned a range of phenomena argued to lead to unacceptability, from syntax (A and $\overline{\text{A}}$ violations, processing difficulties), semantics (contingent false sentences, contradictions, presupposition failures) and pragmatics (tautologies, Maximize Presupposition violations, Magri cases), as well as 10 controls designed to trigger a positive response to each prompt. We also included 4 phenomena claimed to involve Logicality (existential *there* with strong quantifiers, NPIs, anti-rogative predicates, weak islands). Since no context was provided, all sentences were designed so their truth could be evaluated against common knowledge. A PCIbex demo is available here (the last digit in the URL determines the prompt). Stimuli, data, and scripts are available on OSF.

(1)
a. Does the sentence sound natural? *Grammaticality*
b. Do you think the sentence is true? *Truth*
c. Do you think the sentence is false? *Falsity*
d. Do you understand what the speaker meant? *Recoverability*
e. Can you imagine someone genuinely believing this? *Contingency 1*
f. Could you imagine this sentence being true if the world was different? *Contingency 2*

Results: 24 participants were excluded for rejecting the controls. We fitted a linear mixed-effects model on response times with sentence length and prompt as predictors, and by-subject random slopes. The residuals from this model were z-scored and combined with z-scored slider responses, and the average by prompt and item was computed. This gave us 12 dimensionless features for each item, from which we computed a cosine similarity matrix, represented in Fig 1. We also computed the similarity between categories as the mean similarity between their respective items, which allowed us to cluster the categories themselves (see tree on the left of Fig 1).

We found that semantic violations were clearly distinguished from all other categories. Pragmatic violations were close to good controls, while syntactic violations and candidate logicality cases formed a separate branch, relatively close to good controls. In more detail, weak islands are closer to hard-to-process sentences than ungrammatical ones, while NPIs and anti-rogative verbs pattern with $\overline{\text{A}}$-violations. Existential *there* did not form a consistent category: it was treated as contradictory with *every*, acceptable with *the*, and as a syntactic violation otherwise (*most*, *both*, proper noun). We see a few more inconsistencies (some Magri cases are judged false, some presupposition failures are ignored…). Syntactic A-phenomena were also a mixed bag, with case and agreement violations tolerated, but unlicensed reflexives close to contradictions.

Figure 1: Cosine similarity between individual items, organized hierarchically by similarity within and between Categories. The tree indicates clustering on Categories by mean similarity.

A PCA on our 12 features showed that the top 3 components captured 82% of the variance. Keeping only these components returned the exact same clustering on categories. The first component mainly tracked truth, the second grammaticality, and the third one speed. Data from only the first two prompts was enough to replicate the clustering, but dropping RTs prevented it.

Discussion: Overall, the logicality cases are closer to syntactic violations than either semantic or pragmatic ones, confirming the assumption in the literature. This said, we have also found differences in behaviour across the subclasses of logicality, which cannot be straightforwardly accounted for within extant theories of Logicality, as far as we can see. From a methodological standpoint, we note that our design is not particularly sensitive to pragmatic violations (possible reasons: no prompt specifically targeting them; lack of conversational context). Yet, this does not affect our conclusion regarding logicality, and a closer look at the data suggests that the second component of the PCA does distinguish them from good controls (so the cosine similarity as the measure of distance may be to blame). We also showed that two prompts were sufficient to replicate the results, as long as RTs are measured. In short, we offered an easy and affordable method to distinguish semantic and syntactic phenomena, and even subcategories within each group.

References: [1] Abrusán, M. 2019. "Semantic anomaly, pragmatic infelicity, and ungrammaticality". *Ann. Rev. Ling.* 5. [2] Chierchia, G. 2013. *Logic in grammar.* OUP. [3] Chomsky, N. 1955/75. *The Logical Structure of Linguistic Theory*, Plenum. [4] Del Pinal, G. 2019. "The logicality of language". *Noûs* 53. [5] Gajewski, J.R. 2002. "On Analyticity in Natural Language". Ms., MIT. [6] Matthewson, L. 2004. "On the methodology of semantic fieldwork". *Int. j. Am. ling.* 70. [7] Sprouse, J. & D. Almeida. 2012. "Assessing the reliability of textbook data in syntax". *J. Ling.* 48. [8] Sprouse, J. & D. Almeida. 2013. "The empirical status of data in syntax". *Lang. & Cog. Processes* 28.