

How Surprising Is Unacceptable? Evidence from Polish Genitive of Negation

Berke Şensekerçi (b.sensekerçi@uw.edu.pl)¹, Monika Gurak²

¹Faculty of Philosophy, University of Warsaw, ²Faculty of Psychology, University of Warsaw

Introduction Frequency and acceptability are related, but not in a straightforward way: while unacceptable sentences tend to be unlikely, unlikely sentences are not always judged to be unacceptable [1]. Large language models (LLMs) offer a new way to study this relationship using token-level surprisal. In this study, we compare LLM surprisal with (non-binary) human acceptability judgments for a relatively understudied phenomenon in Polish: non-local genitive of negation.

Experiments In Polish, genitive of negation is generally obligatory (cf. (1a)–(1b)), including in non-local configurations where negation on a higher verb licenses genitive case on an embedded object instead of the usual accusative (see (1c)). We conducted three acceptability judgment experiments with native speakers to test whether this licensing depends on the type of verb. Experiment 1 (N = 87) compared four verb classes that select infinitival complements: (i) regular verbs, (ii) modal verbs, (iii) verbs combining with infinitival subjects, and (iv) quasi-verbs (forms homonymous with nominative singular nouns but inflecting for tense and mood). Experiment 2 (N = 73) focused on four quasi-verbs that are fairly rare in this construction (19 attested occurrences in total in the (balanced) National Corpus of Polish with $\sim 250\text{M}$ words). Experiment 3 (N = 57) tested four infinitival-subject verbs with frequencies ranging from extremely low (*przyjąć się*, 1 occurrence) to relatively high (*udać się*, 1000+ occurrences). Participants rated sentences on a 7-point Likert scale. In all experiments, verb (type) was crossed with the case of the object (accusative vs. genitive) (see p. 3 for details).

Surprisal We computed surprisal for all stimuli using two recent LLMs trained on Polish text: PLLuM 8B-base and PLLuM 12B-base. In addition to total surprisal ($S_{total} = \sum_{i=1}^n \log_2 P(w_i | w_1, \dots, w_{i-1})$), we used a standard length-normalized measure to control for differences in item length (and possible tokenizer variation): PenSurprisal ($S_{pen} = \frac{S_{total}}{((5+n_{tokens})/(5+1))^{0.8}}$).

Results We first assessed whether LLM-derived surprisal and human judgments lead to similar conclusions about the constructions under investigation. We fitted linear mixed-effects models separately to min-max scaled human judgments and surprisal measures obtained from the two models, and compared accusative vs. genitive conditions across verb (types). The direction and significance of effects were highly consistent across surprisal (from both models) and human data: lower acceptability ratings generally corresponded to higher values in surprisal metrics. However, effect sizes differed substantially, with both surprisal metrics generally underestimating the magnitude of contrasts (see ‘difference’ column in Table 1). Model diagnostics further indicated that item-level differences accounted for a large portion of the variance in surprisal, suggesting strong sensitivity to lexical content. Subsequent correlation analyses revealed (generally) moderate and highly significant negative correlations between surprisal and acceptability across items (see Figure 1).

Conclusion LLM-derived surprisal seems to align well with human acceptability judgments in terms of directionality, which supports previous findings [2]. The results further show that LLMs can capture complex syntactic configurations, even when these can be rare in language use. At the same time, the systematic mismatch in effect sizes and only moderate correlations indicate that frequency (as approximated by basic LLM surprisal) and acceptability may not be reducible to a single dimension.

References

- [1] Gerard Kempen and Karin Harbusch. Comparing linguistic judgments and corpus frequencies as windows on grammatical competence: A study of argument linearization in German clauses. In Anita Steube, editor, *The Discourse Potential of Underspecified Structures*, pages 179–192. de Gruyter, Berlin, 2008.
- [2] Jiayi Lu, Jonathan Merchan, Lian Wang, and Judith Degen. Can syntactic log-odds ratio predict acceptability and satiation? In Richard Futrell, Connor Mayer, and Noga Zaslavsky, editors, *Proceedings of the Society for Computation in Linguistics 2024*, pages 10–19, Irvine, CA, June 2024. Association for Computational Linguistics. URL <https://aclanthology.org/2024.scil-1.2/>.

Figures and Tables

verb (type)	difference	p.value	experiment	respondent
infinitival	0.044	.066	exp1	S_{total}
infinitival	0.041	.130	exp1	S_{pen}
infinitival	-0.136	.034	exp1	humans
modal	0.156	< .001	exp1	S_{total}
modal	0.164	< .001	exp1	S_{pen}
modal	-0.603	< .001	exp1	humans
quasi	0.011	.629	exp1	S_{total}
quasi	0.032	.227	exp1	S_{pen}
quasi	0.042	.502	exp1	humans
regular	0.124	< .001	exp1	S_{total}
regular	0.117	< .001	exp1	S_{pen}
regular	-0.403	< .001	exp1	humans
grzech	-0.118	.005	exp2	S_{total}
grzech	-0.085	.057	exp2	S_{pen}
grzech	0.250	< .001	exp2	humans
strach	-0.129	.003	exp2	S_{total}
strach	-0.093	.037	exp2	S_{pen}
strach	0.104	< .001	exp2	humans
szkoda	0.072	.079	exp2	S_{total}
szkoda	0.132	.004	exp2	S_{pen}
szkoda	-0.055	.064	exp2	humans
żal	0.048	.228	exp2	S_{total}
żal	0.036	.410	exp2	S_{pen}
żal	-0.008	.776	exp2	humans
opłacać	0.102	< .001	exp3	S_{total}
opłacać	0.126	< .001	exp3	S_{pen}
opłacać	-0.595	< .001	exp3	humans
przyjąć	0.038	.080	exp3	S_{total}
przyjąć	0.052	.086	exp3	S_{pen}
przyjąć	-0.400	< .001	exp3	humans
udać	0.122	< .001	exp3	S_{total}
udać	0.128	< .001	exp3	S_{pen}
udać	-0.689	< .001	exp3	humans
wystarczyć	0.002	.928	exp3	S_{total}
wystarczyć	0.006	.829	exp3	S_{pen}
wystarczyć	-0.288	< .001	exp3	humans

Table 1: Estimated ACC–GEN differences in min–max scaled PLLuM–12B surprisal and human judgments.

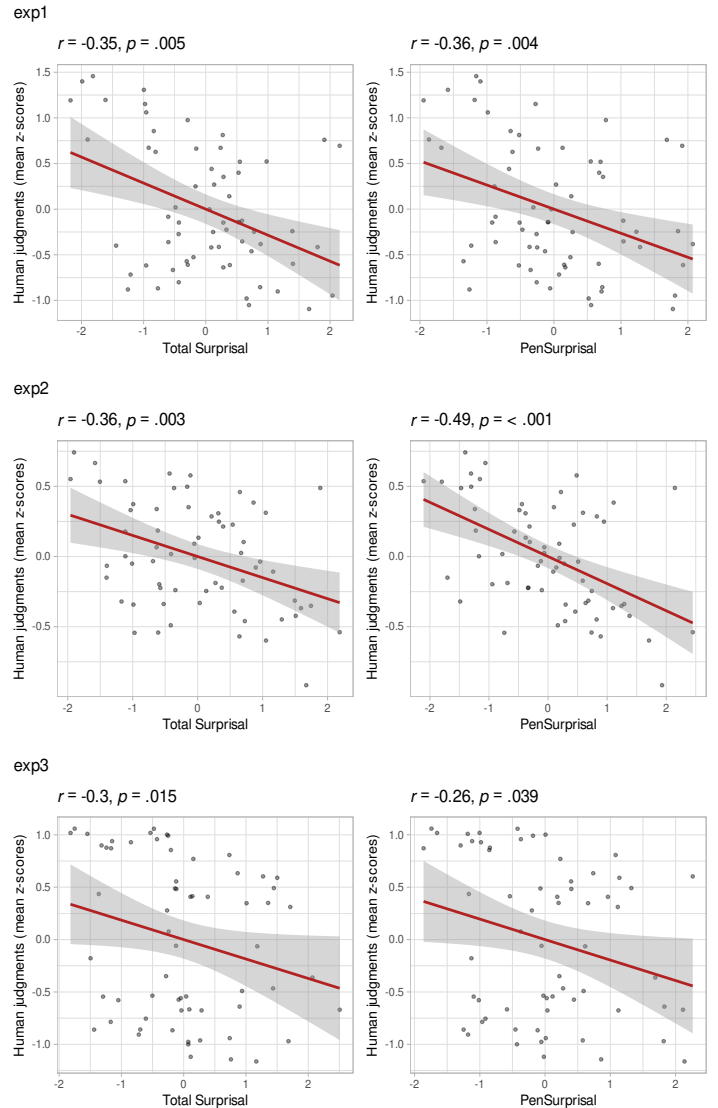


Figure 1: Spearman correlations between z-scored human judgments and PLLuM–12B surprisal for each stimulus.

Non-local Genitive of Negation in Polish

In Polish, verbs that assign the accusative case to their direct objects, when negated, require their direct objects to appear in genitive case (cf. (1a)–(1b)), which is the so-called *genitive of negation*. It is obligatory in Polish and it is also almost always obligatory in non-local contexts, where negation on a higher verb passes its genitive requirement to the direct object of an embedded infinitive (1c).

- | | | |
|---|---|--|
| (1a) <i>Jem gruszkę.</i>
eat.1SG pear.ACC

'I am eating a pear.' | (1b) <i>Nie jem gruszki.</i>
NEG eat.1SG pear.GEN

'I am not eating a pear.' | (1c) <i>Nie planuję jeść gruszki.</i>
NEG plan.1SG eat.INF pear.GEN

'I am not planning to eat a pear.' |
|---|---|--|

Quasi-verbs

Quasi-verbs are used predicatively but inflect only analytically and only for tense and mood (cf. (2a)–(2b)), unlike regular verbs. Some quasi-verbs have a modal character (e.g., *trzeba* 'one should') and some are homonymous with singular nouns (e.g., *wstyd* 'shame'); the grammatical status of the latter is controversial, as they might be parsed as nouns or adverbs in elliptical constructions.

- | | |
|---|---|
| (2a) <i>Wstyd mi pisać taką wiadomość.</i>
shame me.DAT write.INF this.ACC message.ACC

'I feel embarrassed writing this message.' | (2b) <i>Wstyd mi było pisać taką wiadomość.</i>
shame me.DAT was.PST.3.N.SG write.INF this.ACC message.ACC

'I felt embarrassed writing this message.' |
|---|---|

Infinitival subjects

Some verbs combine with an infinitive that alternates with a nominative subject. For example, *Wenecja* is a nominative subject in (3a) and the infinitival phrase *zwiedzić Wenecję* in (3b) occupies the same syntactic position. Since the genitive of negation targets direct objects, we asked whether it propagates as non-local genitive of negation when the embedded infinitive occupies the subject rather than the object position.

- | | |
|--|--|
| (3a) <i>Marzy mi się Wenecja.</i>
Dreams me.DAT REFL Venice.NOM

'I dream of Venice.' | (3b) <i>Marzy mi się zwiedzić Wenecję.</i>
Dreams me.DAT REFL visit.INF Venice.ACC

'I dream of visiting Venice.' |
|--|--|

Example experimental items

- | | |
|--|---|
| (4) <i>Nie warto czytać [tego listu / ten list] od naszych sąsiadów.</i>
NEG worth read.INF this.GEN letter.GEN / this.ACC letter.ACC from our neighbors

'It is not worth reading this letter from our neighbors.' | (Experiment 1; modal verb) |
| (5) <i>Nie żal wyrzucić [starego pamiętnika / stary pamiętnik] z młodości.</i>
NEG pity throw.away.INF old.GEN diary.GEN / old.ACC diary.ACC from youth

'It's not a pity to throw away an old diary from one's youth.' | (Experiment 2; <i>żał</i> 'pity') |
| (6) <i>Nie wystarczyło przygotować [krótkiego eseju / krótki esej] na zaliczenie.</i>
NEG suffice.PST.3.N.SG prepare.INF short.GEN essay.GEN / short.ACC essay.ACC for passing

'Preparing a short essay was not enough to pass.' | (Experiment 3; <i>wystarczyć</i> 'suffice') |