Variable agreement with conjoined subjects in German: Testing the grammatical lacuna hypothesis

Eva Neu, UMass Amherst*

September 14, 2023

1 Introduction

Subject-verb agreement is known to sometimes give rise to semantically vacuous free variation, such that one and the same subject can trigger more than one possible agreement form on its verbs. This is commonly regarded as an instance of a grammar producing multiple, equally acceptable outputs from a single input (e.g., Adger, 2006, Adger and Smith, 2010). However, a different interpretation of variable agreement has recently been pitched by Foppolo and Staub (2020), namely that in some cases, the grammar might not produce any output at all. Investigating number agreement with disjoint subjects such as *the dogs or the cat* in English and Italian, Foppolo and Staub found that speakers were largely indifferent between singular and plural agreement but also judged neither form fully acceptable. Moreover, ratings were highly inconsistent over the course of the experiment. To account for these results, Foppolo and Staub proposed that number agreement with disjoint subjects is, in the languages investigated, a grammatical lacuna, meaning that the grammar does not contain any rules for resolving agreement in such contexts. As a result, speakers rely on 'patch-up devices,' strategies which are ad-hoc and probabilistic rather than grounded in firm grammatical knowledge and which result in erratic response behavior and degraded judgments. To my knowledge, this hypothesis, while intriguing, has so far not been subjected to further scrutiny.

In this paper, I test the grammatical lacuna hypothesis against a similar instance of variable agreement with conjoined subjects which has been attested in German (Driemel, 2018, Timmermanns et al., 2004). If the subject conjoins second and third person features, the verb can surface either with second or with third person plural agreement (1):[1]

(1)   Du   und der Hund seid   / sind   zu   schnell für mich.
     you.SG and the dog   be.2PL / be.3PL too fast   for   me
     'You and the dog are too fast for me.'

This variation is limited to combinations of second and third person features. First-and-second as well as first-and-third person subjects consistently result in first person agreement, in line with Corbett's (2006) cross-linguistically well-attested resolution rules.

---

[1]First and third person agreement are syncretic in German; however, anaphors can distinguish between the two:

(i)   Du   und Kim können   ?sich   / *uns   auf das Sofa setzen.
     you and Kim can.1PL/3PL themselves / ourselves on the sofa sit
     'You and Kim can sit on the sofa.'

Judgments for third – as well as second – person anaphors are subtle and might be slightly degraded for some speakers, a fact I must sidestep in this paper. The first person anaphor, on the other hand, is categorically unavailable, confirming that the syncretic verb form realizes third instead of first person features. Note that variable agreement with second-and-third subjects is equally attested in Dutch (Timmermanns et al., 2004), which does not have a 1/3 syncretism.

Some preliminary evidence that these data might constitute an instance of a grammatical lacuna comes from a previous production study (Timmermanns et al., 2004) which presented speakers with second-and-third person subjects and asked them to fill in the verb form. They found that a considerable number of speakers violated the instructions by adding an appositive second person plural pronoun to the sentence, resulting in forms such as *you and the dog, you...* These findings indicate that speakers actively avoid agreement with second-and-third subjects – perhaps because they sense the presence of a grammatical gap.

Thus, this paper seeks to confirm whether variable second-and-third person agreement in German is due to an absence of grammatical knowledge or should rather be regarded as reflecting multiple grammatical outputs. To this end, I conducted an empirical study which I describe below.

2 Methods

In order to detect the presence of a grammatical lacuna, I investigated how acceptable speakers find the two variants and how stable their judgments are. As outlined above, the grammatical lacuna hypothesis predicts similar, and similarly degraded ratings for the two agreement forms and inconsistent response behavior. In addition, I tested whether second-and-third person agreement in German is subject to linearity effects which would suggest that the verb agrees with the closer conjunct only, as is attested in many languages (Nevins and Weisser, 2019).

I used a simple non-speeded acceptability judgment task in which subjects were asked to rate sentences on a 5-point Likert scale. Items were constructed using a 2x3 Latin square design. The two factors were a) agreement marking on the verb – 2PL, 3PL and an ungrammatical 1SG control condition – and b) order of conjuncts, the latter to detect linearity effects. All individual conjuncts used were singular. Table 1 gives an overview over the resulting six conditions:

(2)

2&3_1SG	3&2_1SG
You and the dog be.1SG...	The dog and you be.1SG...
2&3_2PL	3&2_2PL
You and the dog be.2PL...	The dog and you be.2PL...
2&3_3PL	3&2_3PL
You and the dog be.3PL...	The dog and you be.3PL...

Table 1: Experimental conditions

I recruited 80 participants via Prolific, a number which was determined by a power analysis based on an earlier pilot study. Each subject saw 60 test items as well as 120 fillers, the latter of which consisted of 40 grammatical items, 40 ungrammatical items and 40 items of intermediate acceptability. Responses to these fillers were used to evaluate whether participants followed the instructions and paid sufficient attention to the task; the data of 2 subjects were discarded on these grounds. Subjects were given examples of grammatical, ungrammatical and intermediate sentences at the beginning of the study. At the end, they were asked to describe their dialect and to share their thoughts and comments on the study. The experiment was implemented in PCIbex and analyzed in R.

3 Results

Figure 1 shows the mean ratings for each of the six conditions. As expected, the two ungrammatical control conditions score poorly. The linear order of conjuncts has no effect on agreement ratings, confirming that the verb does not agree with one of the conjuncts only. The four test conditions all receive highly uniform scores around 4, a rating which is only slightly degraded (grammatical fillers scored around 4.7). Figure 2 contains the histograms of the test conditions, showing unimodal distributions which are, again, highly similar to each other.
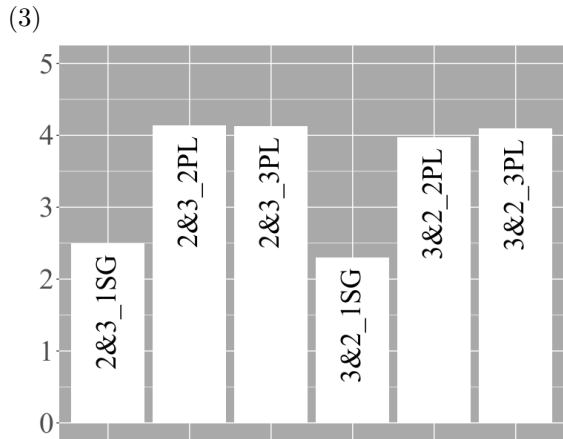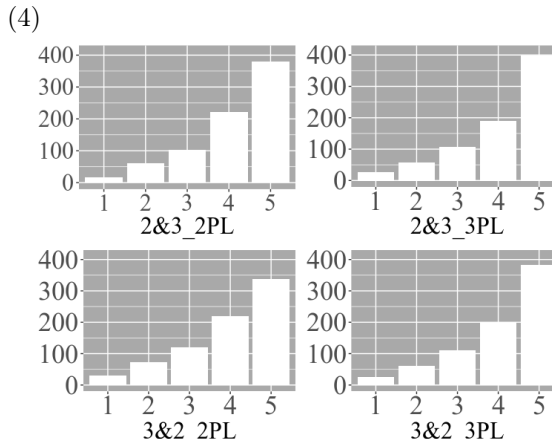
(3)

Figure 1: Mean ratings

(4)

Figure 2: Histograms of test conditions

I further analyzed the data by building a Bayesian mixed-effects ordinal regression model (4 chains, iter=4000), using the `brms` package in R, with subject and item as random effects. The three fixed effects were a) test vs. control conditions, b) second vs. third person agreement and c) order of conjuncts. As expected, the contrast between test and control conditions emerged as a strong effect (Est. -0.67, 95%CI -0.76–-0.58) whereas no effect was found for the contrast between second and third person agreement (Est. -0.05, 95%CI -0.19–0.08). There was a small effect of order of conjuncts (Est. 0.08, 95%CI 0.04–0.12): independent of agreement, speakers showed a slight preference for placing the second before the third person, arguably due to prosodic or pragmatic reasons.

To investigate how consistent speakers were in their response behavior, I determined the test-retest reliability using the model-based method from Staub (2021). For each of the three fixed effects, a new Bayesian model was built which had two fixed effects, namely the original fixed effect in the even and the odd items, respectively. This was achieved by coding the effect as usual in the even but as zero in the odd items, or vice versa. I added random subject-level slopes and intercepts to the model and extracted the correlation between the subject-level random slopes for the effect in even and the effect in odd items. This correlation thus serves as a measure of how predictive the behavior of a subject in one half of the experiment is of their behavior in the other half with respect to a given fixed effect.

I found extremely high test-retest reliability for the main effect of interest, second vs. third person agreement (Est. 0.92, 95% CI 0.75–0.99). These results thus indicate that subjects adopted a consistent strategy for agreement with second-and-third subjects instead of responding in a random and erratic manner. For test vs. control, test-retest reliability was equally high (Est. 0.96, 95% CI 0.87–1.00), whereas for order of conjuncts, reliability was very low (Est. -0.18, 95% CI -0.92–0.81). The latter finding can be attributed to the fact that subjects did not vary much with respect to this effect in the first place; the standard deviation of the subject-level effect of order in the original model was 0.06 (95% CI 0.00–0.12).

To additionally determine how much individual subjects differed from each other in their response to the contrast between second and third person agreement, Figure 3 shows the subject-level effects for second vs. third person from the original model. Bars showing a positive value correspond to speakers with a preference for second over third person, and vice versa for bars with a negative value; a small number of speakers in the middle of the plot are truly indifferent. Values on the y-axis correspond to points on the Likert scale; e.g., for the subject corresponding to the bar on the very left side of the plot, the average score for second person agreement was approximately 1.4 points higher than that for third person agreement. The plot thus demonstrates that while on a population level, second and third person are treated equally – with a fixed effect close to 0 –, most individual speakers consistently prefer one variant over the other to a gradient degree.

By way of contrast, Figure 4 shows the subject-level effects for test vs. control condition, which equally had extremely high test-retest reliability. We see that while all speakers prefer the test over the control conditions, they do so to varying degrees. Importantly, what emerges from the juxtaposition of Figure 3 and Figure 4 is that the range of variation between individual subject-level effects is larger for second vs. third person (1.4–-1) than for test vs. control condition (-0.8–-2.2). Since the latter is a contrast between grammatical

and ungrammatical conditions, any variation in this respect should be interpreted as idiosyncratic differences in individual response behavior, such as how subjects interpreted the scale (Sprouse and Almeida, 2017). The fact that inter-subject variation with respect to second vs. third person agreement goes beyond this baseline indicates that speakers differ from each other to a non-negligible extent in their intuitions about agreement with second-and-third subjects.
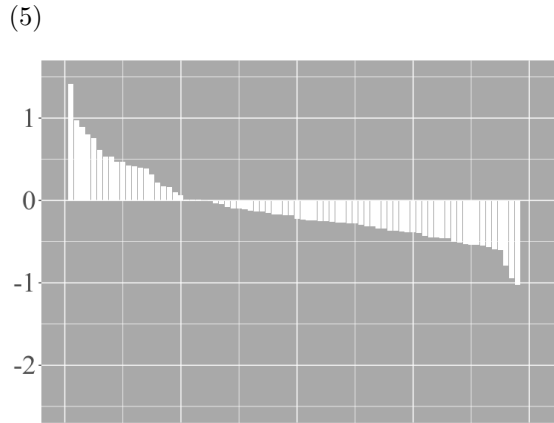
(5)

(6)

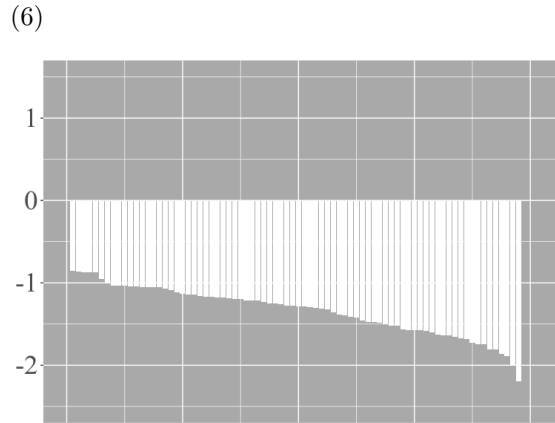Figure 3: Subject-level effects for second vs. third person

Figure 4: Subject-level effects for test vs. control condition

4   Discussion

To recapitulate the predictions of the grammatical lacuna hypothesis, ratings for the two variants were expected to be similar and somewhat degraded, and subjects were expected to behave inconsistently. The latter prediction was not borne out by the data, the former only partially. While the two variants received, on a population level, highly similar judgments, the ratings were only marginally lower than those of the grammatical fillers. What is more, using test-retest reliability I confirmed that subjects were in fact highly consistent in their responses. These results might be taken as clear evidence against the idea that variable second-and-third person agreement in German is an instance of a grammatical lacuna, rather supporting a traditional analysis as multiple grammatical outputs. However, several caveats apply.

First, when invited to comment on the experiment at the end of the study, several participants described feeling unsure about which verb form to use with second-and-third person subjects. One participant explicitly reported mentally replacing such subjects by a second person pronoun, in line with the findings by Timmer-manns et al. (2004). Thus, the claim that the grammar provides no clear guidance for second-and-third person agreement does arguably correspond to how speakers themselves intuitively experience such sentences, which I could also confirm anecdotally with several native speakers in the aftermath of the experiment. If both agreement variants are indeed perfectly grammatical, these findings require an explanation.

Secondly, a potential confound to consider is that the high consistency in subjects' responses could be an artificial effect of the study. That is, subjects might have adopted a rule for resolving agreement over the course of the experiment and then followed it consistently. It is not unlikely that subjects would attempt to make their responses regular and rule-governed, consciously or not, given the task-based nature of the study and the high number of relevant items they were exposed to. However, if speakers can form a rule for second-and-third person agreement over the course of a 20-minutes experiment, one might wonder if it is plausible that they have not already done so over the course of their 20+-years life, even taking into account that the relevant sentences are arguably not frequent. I argue that in fact, we do not face a binary choice here. While the notion of a grammatical lacuna suggests a dichotomy between the presence and the absence of knowledge, general rules are arguably induced from specific items in a probabilistic fashion. Thus, a reasonable scenario is that (some) speakers entered the experiment already in possession of a weakly held and loosely followed rule but that the latter was subsequently reinforced by the experiment. What makes this and similar hypotheses difficult to confirm or reject is the problem of the observer's paradox: in order

to determine whether speakers' behavior is rule-based, we must expose them to a sufficient number of items, but prolonged exposure runs the risk of giving rise to such a rule in the first place. More sophisticated methodologies would be needed to circumvent this conundrum.

Finally, while we have observed that speakers' judgments are stable, this is not necessarily evidence for a grammatical rule. The concept of a grammatical lacuna relies on a particular architecture of linguistic knowledge: it presupposes that a speaker's grammar – involving, e.g., operations such as Merge, Move and Agree – is only one factor which influences the linguistic object a speaker produces or comprehends, besides performance factors such as processing, frequency effects, etc.[2] It is from the latter factors that Foppolo and Staub's patch-up devices are recruited. However, if such a distinction between grammatical and extragrammatical knowledge is assumed, it is not obvious that consistency is exclusively a hallmark of the former. That is, speakers might rely on stable and rule-based strategies which are nevertheless not part of the grammar in the narrow sense. In this way, German second-and-third agreement might resemble what is known as the depth charge illusion, observed in sentences such as (7) (Wason and Reich, 1979):

(7)   No head injury is too trivial to be ignored.

While (7) can be shown to be grammatically ill-formed in terms of semantic compositionality, the sentence is nevertheless consistently judged acceptable by speakers and assigned a coherent interpretation. By the same token, second-and-third person agreement might not fall out from the grammatical machinery itself but still be routinely understood and produced by speakers in a rule-based fashion. Overall, the notion of a grammatical lacuna thus raises thorny questions about the nature of the grammar, the extent to which it can be neatly separated from other forms of linguistic knowledge, and the distinction between competence and performance.

To conclude, this study has attempted to probe into what is and what is not in speakers' grammars by testing the consistency of their judgments. This diagnostic has yielded a clear result, namely that speakers are highly stable in their behavior during the experiment, but it has done little to elucidate the original question. Consistency does not necessarily correlate with confidence in grammatical intuitions, it is difficult to test due to the observer's paradox, and if a distinction between grammatical and extragrammatical factors is assumed, consistency might just as well be due to the latter as to the former. Hence, while I could neither clearly confirm nor disconfirm the grammatical lacuna hypothesis, this study suggests that the question ought to be framed in a more nuanced way. Whether speakers feel certain about their judgments or sense the absence of grammatical guidance, whether they follow a consistent strategy or improvise, and whether their behavior should be attributed to domain-specific linguistic knowledge or more general cognitive mechanisms might in fact be three orthogonal questions.

References

Adger, D. (2006). Combinatorial variability. *Journal of linguistics*, *42*(3), 503–530. https://doi.org/10.1017/s002222670600418x

Adger, D., & Smith, J. (2010). Variation in agreement: A lexical feature-based approach. *Lingua*, *120*(5), 1109–1134. https://doi.org/10.1016/j.lingua.2008.05.007

Corbett, G. G. (2006). *Agreement*. Cambridge University Press.

Driemel, I. (2018). Person mismatch agreement. In S. Hucklebridge & M. Nelson (Eds.), *Proceedings of NELS 48*. GLSA.

Foppolo, F., & Staub, A. (2020). The puzzle of number agreement with disjunction. *Cognition*, *198*. https://doi.org/10.1016/j.cognition.2019.104161

Nevins, A., & Weisser, P. (2019). Closest conjunct agreement. *Annual Review of Linguistics*, *5*(1), 219–241. https://doi.org/10.1146/annurev-linguistics-011718-012708

---

[2]This is different from, e.g., constraint-based frameworks such as OT, in which grammatical factors in the narrow sense and performance-based effects are subsumed in a single system termed the grammar. Confronting linguists working on constraint-based grammars with the idea of a grammatical lacuna results, in my experience, in a reaction ranging between hostility and utter confusion, and understandably so. If the grammar is regarded as the set of all factors determining a linguistic output, a grammatical lacuna should simply result in no output at all.

Sprouse, J., & Almeida, D. (2017). Design sensitivity and statistical power in acceptability judgment experiments. *Glossa: a journal of general linguistics*, *2*(1). https://doi.org/10.5334/gjgl.236

Staub, A. (2021). How reliable are individual differences in eye movements in reading? *Journal of Memory and Language*, *116*, 104190. https://doi.org/10.1016/j.jml.2020.104190

Timmermanns, M., Schriefers, H., Dijkstra, T., & Haverkort, M. (2004). Disagreement on agreement: Person agreement between coordinated subjects and verbs in Dutch and German. *Linguistics*, *42*, 905–929.

Wason, P. C., & Reich, S. S. (1979). A verbal illusion. *Quarterly Journal of Experimental Psychology*, *31*(4), 591–597. https://doi.org/10.1080/14640747908400750