

ORIGINAL ARTICLE

# Gradience in subject–verb number agreement: Can bilinguals tune in?

Anna Jessen<sup>1\*</sup> , Lara Schwarz<sup>2,3</sup> and Claudia Felser<sup>1</sup>

<sup>1</sup>University of Potsdam, Potsdam, Germany, <sup>2</sup>The Pennsylvania State University, University Park, PA, USA  
and <sup>3</sup>University of Dortmund, Dortmund, Germany

\*Corresponding author. Email: [anna.jessen@ugent.be](mailto:anna.jessen@ugent.be)

(Received 18 December 2019; revised 14 June 2021; accepted 21 June 2021; first published online 03 September 2021)

## Abstract

This study investigates native German speakers' and bilingual Turkish/German speakers' sensitivity to constraints on verbal agreement with pseudo-partitive subjects such as *eine Packung Tabletten* ("a pack of pills"). Although number agreement with the first noun phrase (headed by a container noun) is considered to be the norm, agreement with the second (containee) noun phrase is also possible. We combined scalar acceptability ratings with a stochastic constraint-based grammatical framework to model the relative strength of the constraints that determine speakers' agreement preferences and subsequently tested whether these models could correctly predict speakers' verb choices in a production task. For both participant groups, number match between the container noun phrase and the verb was the strongest determinant of both acceptability and production choices. The relative ranking of the constraints that we identified was the same for both groups, and the lack of age-of-acquisition effects suggests that constraints on variable subject–verb agreement, and their relative strength, are acquirable by both early and later learners of German. Group differences were seen in the absolute constraint weightings, however, with the bilinguals' agreement preferences being more strongly influenced by number match with the containee phrase, indicating a comparatively greater reliance on surface-level cues to agreement (such as noun proximity) among the bilingual group.

**Keywords:** subject–verb agreement; pseudo-partitives; German; bilingualism; Gradient Symbolic Computation

---

## Author Note

We have no known conflict of interest to disclose.

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number: 317633480 – SFB 1287 Project B04.

Supplementary materials such as critical stimuli, additional nonsignificant results, and data collected for the experiments can be found on OSF: <https://osf.io/ns7gm/>

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

In subject–verb agreement marking languages such as German, binominal subjects sometimes allow for variable number agreement with the verb, which is illustrated by the examples in (1a–d) containing pseudo-partitive subjects.

- (1) a. *Ein Kilo Linsen kostet/kosten drei Euro*  
 a kilogram lentils costs/cost three Euro  
 “A kilogram of lentils costs three Euros.”
- b. *Eine Menge Menschen verlor/verloren ihre Arbeit.*  
 a lot people lost<sub>SG</sub>/lost<sub>PL</sub> their work  
 “A lot of people lost their job.”
- c. *Ein Haufen Akten wurde/wurden vernichtet.*  
 a heap files was/were destroyed  
 “A heap of files was destroyed.”
- d. *Zwei Eimer Wasser ist/sind genug.*  
 two buckets water is/are enough  
 “Two buckets of water are sufficient.”

In pseudo-partitives the first noun phrase (NP1) measures a proportion of the second one (NP2), which is indefinite and non-referential. German pseudo-partitives differ from partitives such as *ein Kilo von den Linsen* (“one kilogram of the lentils”) or *eine Menge dieser Menschen* (“a lot of these people”) in that NP2 is a bare noun phrase and not linked to NP1 by a preposition or genitive determiner.<sup>1</sup> Although agreement with NP1 is usually considered to be the prescriptive norm in present-day German, agreement with NP2 is also possible. The *Duden* German reference grammar describes this kind (and other cases) of variable agreement as *Zweifelsfälle* (“cases of doubt”) (Hennig *et al.*, 2016), acknowledging that subject–verb number agreement in pseudo-partitives is not determined by any categorical rules (see also Jaeger, 1992; Wegerer, 2012).

Linguistic phenomena that display gradience or variation can often be captured rather well within constraint-based frameworks that allow individual constraints to be weighted and different types of constraints to interact (e.g., Haskell & MacDonald, 2003). Cross-linguistic evidence shows that several factors can influence number agreement with binominal subjects. These include a noun’s proximity to the verb (Haskell & MacDonald, 2005), morpho-phonological factors (Vigliocco *et al.*, 1995), plausibility (Thornton & MacDonald, 2003), and the subject phrase’s notional plurality (Humphreys & Bock, 2005; Smith *et al.*, 2018). For pseudo-partitive subjects as in (1a–d), the choice of agreement controller has been linked to whether the first noun is interpreted as referential or as an abstract unit of measurement (e.g., Selkirk, 1977). However, the factors that determine speakers’ number agreement preferences with German pseudo-partitives are as yet poorly understood.

Functional morphology, including subject–verb agreement marking, has long been known as a source of difficulty in non-native language acquisition (Slabakova, 2019). Most previous studies have investigated bilinguals’ ability to compute grammatically conditioned (i.e., categorical) agreement, however, and we still know very little about whether bilinguals can acquire the constraints that determine variable or optional agreement. From the perspective of language acquisition, linguistic

phenomena that do not follow any categorical grammatical rules are interesting because they cannot easily be taught, so that their target-like acquisition is likely to depend on learners' ability to extract relevant constraints and constraint weightings from the input through experience (compare e.g., Haskell et al., 2010). As bilingual speakers, especially those who acquired their second language (L2) not from birth but later in life, typically receive less (and qualitatively different) input compared to monolingually raised individuals, we might expect this to affect their performance in domains that exhibit within-language variation. Comparing bilingual and monolingually raised speakers' sensitivity to different kinds of constraints should thus allow us to determine which constraints are affected by suboptimal input conditions and which are more robust.

Focusing on binominal subjects with container nouns (as in 1d above), the present study asks to what extent German speakers accept (Experiment 1) and produce (Experiment 2) singular versus plural verbs in cases where there is a number conflict between NP1 and NP2. Secondly, we ask whether Turkish/German bilingual speakers are guided by the same constraints as non-bilingual German speakers, and how their performance might be affected by individual factors such as the age at which they started acquiring German. Note that in Turkish, no number conflicts exist in these structures, as is illustrated by the examples in (2).

- (2) a. *üç kutu vitamin*  
 three box vitamin  
 "three boxes of vitamins"  
 b. *iki kilo elma*  
 two kilo apple  
 "two kilograms of  
 apples"

In (2a,b), both the first and the second noun lack a plural marker, and plurality is indicated only by the numeral. Since subject–verb agreement is determined by the head noun and not by the numeral in Turkish (e.g., Sağ, 2018), a singular (or unmarked) verb would be chosen even in the presence of a plural numeral.

To identify the constraints that determine German speakers' agreement preferences in cases such as (1d), and the relative weighting of these constraints, we carried out a scalar acceptability judgment task and subsequently modeled the results using Gradient Symbolic Computation (GSC) (Smolensky et al., 2014). GSC is a stochastic, constraint-based harmonic framework similar to stochastic Optimality Theory (OT) or Harmonic Grammar (HG), and it has been shown to be capable of modeling rare occurrences of variation in bilingual language production such as structural blends or portmanteau constructions (Goldrick et al., 2016). In the second experiment, we tested whether the constraint weightings thus obtained can successfully predict speakers' choices in a timed production task.

### Agreement with pseudo-partitive subjects in German

German is a non *pro*-drop language in which finite verbs are required to agree with the subject in person and number. The choice of agreement controller is not always

obvious for pseudo-partitive subjects as in (1) above, however. When NP1 and NP2 carry conflicting number features, the conditions under which agreement may be triggered by NP2 rather than NP1 are not yet fully understood. It has often been noted that variable agreement with pseudo-partitives is linked to the ambiguity of NP1 (Grestenberger, 2015; Scontras, 2014; Selkirk, 1977; Stickney, 2009, among others). In complex noun phrases such as *zwei Eimer Wasser* (“two buckets of water”), the first noun has two different readings: a referential (or container) reading and a quantity reading, where *Eimer* (“bucket”) is interpreted as an abstract unit of measurement rather than as an actual object.

One way of accounting for variable number agreement formally is to assume that the subject phrase is structurally ambiguous, such that either the first or the second noun can project and thus determine the overall NP’s grammatical number (Selkirk, 1977). This is indicated schematically in (2a,b).

- (3) a. [<sub>NP1-PL</sub> *zwei Eimer* [<sub>NP2-SG</sub> *Wasser* ]]  
 b. [<sub>NP2-SG</sub> [<sub>NP1-PL</sub> *zwei Eimer* ] *Wasser* ]

In (3a), the plural noun phrase *zwei Eimer* (“two buckets”) projects, rendering the subject phrase plural, whereas in (3b) the first NP is merged as the specifier of the second one, an analysis which corresponds to the quantity reading of NP1. If the grammatically singular substance noun heads the overall NP as in (3b), singular agreement will be the grammatically appropriate option.

Exactly how the two readings are represented structurally has been a matter of some debate, however. Grestenberger (2015), for example, proposes that the quantity reading results from analyzing the ambiguous noun as a functional NUMBER head which measures a portion of the substance noun, rather than as a lexical noun, as shown in simplified form in (4b) versus (4a) (see Scontras, 2014, for a similar proposal).

- (4) a. [<sub>NumP</sub> *zwei* [<sub>Num'</sub> [<sub>Num</sub> Ø ] [<sub>NP1</sub> [<sub>N1</sub> *Eimer* ] [<sub>NP2</sub> *Wasser* ]]]] (container reading)  
 b. [<sub>NumP</sub> *zwei* [<sub>Num'</sub> [<sub>Num</sub> *Eimer* ] [<sub>NP</sub> *Wasser* ]]] (quantity reading)

Grestenberger further states that “number agreement in numeral constructions in a language like German will always be determined by the denotation of the numeral, independent of the number marking on the substance noun” (Grestenberger, 2015: 107). However, this statement seems questionable given that examples such as (5a) and (5b), both of which involve an agreement mismatch with the numeral, may be considered acceptable to some degree.

- (5) a. *1 kg Linsen sind genug.*  
 1 kilogram lentils are enough  
 b. *2 kg Mehl ist genug.*  
 2 kilograms flour is enough

Noun types differ with regard to how readily they allow for a referential versus quantity reading. For English pseudo-partitives, Smith *et al.* (2018) showed that those with container nouns (e.g., *a pot of flowers*) elicited fewer plural verb forms

compared to those containing a collection noun (e.g., *a heap of blankets*) or a measure noun (e.g., *a lot of postcards*). For German, variable agreement with pseudo-partitives has not previously been examined systematically in controlled experimental settings. The results from a cross-linguistic questionnaire study by Berg (1998) indicate differences between German and English native speakers. Testing a variety of different subject types, Berg observed that agreement choices were strongly influenced by semantics in English while being influenced primarily by syntactic factors in German. In the present study, we focus on grammatical factors rather than on semantic ones.

Most experimental studies of German subject–verb agreement have examined agreement attraction errors (e.g., Hartsuiker et al., 2003; Häussler, 2009; Hölscher & Hemforth, 2000; Lago & Felser, 2018; Reifegerste et al., 2020). One frequently reported finding in the agreement attraction literature that is potentially relevant for the current study is a singular/plural asymmetry. That is, plural-marked second nouns in subject phrases such as *The key to the cabinets . . .* are more likely to elicit agreement errors (e.g., choice of *are* rather than *is*) than singular second nouns as in *The keys to the cabinet . . .* (e.g., Bock & Miller, 1991). This asymmetry is usually ascribed to plural features being marked, or plural forms being more salient in memory than singular forms, making them more likely to be mis-selected as agreement controllers than the latter (compare e.g., Eberhard, 1997; Staub, 2009). While findings from attraction studies have contributed crucially to the development and evaluation of theoretical models of agreement computation, the current study was not designed for this purpose. What sets our study apart from attraction studies is the fact that we examine permissible variation arising from the subject phrase’s structural ambiguity rather than agreement errors. Being unable to tell whether participants’ preference for a singular or plural verb reflects the way the subject phrase was syntactically encoded, or erroneous agreement computation, makes our data of limited relevance to theoretical debates about the mental mechanisms involved in agreement computation. Rather than seeking to evaluate formal approaches to pseudo-partitives or to test psycholinguistic models of agreement computation, our study aims to assess German speakers’ number agreement preferences in controlled experimental settings and to quantify the constraints that determine these preferences.

An earlier study that did look at permissible variation in German subject–verb agreement is a questionnaire study by Wegerer (2012), who tested more than 2000 participants on over 80 different sentences listed under “cases of doubt” in the *Duden*. Her materials also included some pseudo-partitive items such as *eine Menge Kohlen wird/werden geliefert* (“a lot of coals is/are delivered”). Participants were asked to choose either a singular or plural verb as they deemed appropriate. Her results do not present a clear picture of German native speakers’ preferences for these constructions, however. The above-mentioned example with *eine Menge* (“a lot”) elicited different proportions of singular verbs depending on the post-modifying noun: with *Kohlen* (“coals”) singular and plural answers were equally distributed, but with *Bücher* (“books”) there was a clear preference for plural verbs, for example. Wegerer concluded that a variety of factors may influence German speakers’ choice of verb number marking, including semantic and pragmatic factors, a noun’s proximity to the verb, as well as individual-level factors such

as sociodemographic and geographic differences between participants. Given that her analysis shows only proportions of choices, and the fact that her experimental sentences were not constructed as minimal pairs but differed from one another both structurally and lexically (which is also the case for Berg's, 1998, materials), we cannot draw any reliable conclusions regarding number agreement with pseudo-partitives despite the large number of people tested.

### The present study

This study examines number agreement with German pseudo-partitive subjects with container nouns as in (1d). Number marking on both NP1 and NP2 was systematically manipulated so as to allow us to assess which of the two was preferentially taken to be the agreement controller. Our aim was to identify some of the constraints that determine German speakers' agreement preferences, and the relative weighting of these constraints. Moreover, comparing monolingual and bilingual speakers allows us to investigate whether and how these constraint weightings are affected by language status (native vs. non-native), by the age at which German was acquired or by other bilingualism variables. Given the dearth of experimental work on pseudo-partitives and the fact that most bilingualism research on agreement has examined categorical rather than variable agreement, the present study is largely exploratory.

In our first experiment, we gathered scalar acceptability ratings to gauge the extent to which singular and plural verb forms are accepted in situations where both options might be possible. We then modeled the judgment results using GSC so as to identify the constraints that govern participants' number agreement preferences, and the relative weightings of these constraints. To explore whether agreement conflicts are resolved differently in comprehension-based versus production tasks, we additionally carried out a binary forced-choice production experiment and compared the models of the two tasks.

### Predictions

Based on previous findings and anecdotal evidence, we predict that monolingually raised German speakers should prefer NP1 as the agreement controller. That is, NP1-match sentences should be judged as significantly more acceptable than NP1-mismatch sentences. However, given the additional possibility of NP2 controlling agreement, we expect that an NP2 that matches the verb in number should improve acceptability to some degree, especially where NP1 mismatches the verb's number. Participants' judgments may also show evidence of a singular/plural asymmetry as has been reported in the agreement attraction literature. If plural-marked NP2s are more likely to trigger verbal agreement than singular NP2s, then the presence of a plural NP2 should improve the acceptability of plural verbs if NP1 is singular more than in the reverse situation.

As for the bilingual speakers, several outcomes are conceivable. Current formal approaches to bilingualism do not necessarily have much to say about what might determine L2 speakers' choices in cases of permissible variation. The phenomenon under investigation is not explicitly taught, and agreement preferences must

therefore be acquired based on language experience. We might therefore expect any L1/L2 differences to be smaller in early than in late bilinguals. From the perspective of usage-based models of language acquisition and processing (e.g., MacWhinney, 2005), it is conceivable that due to reduced experience or suboptimal acquisition conditions L2 speakers adopt a general NP1 agreement strategy for German pseudo-partitives as this is the most common agreement pattern. Claims to the effect that L2 processing is more affected by cognitive resource limitations than L1 processing (e.g., McDonald, 2006) might lead to the opposite expectation that our bilingual participants should choose NP2 agreement more frequently compared to monolinguals, as this would seem to be the most memory-friendly option. Any such general tendency should be especially conspicuous in situations with added processing pressure (i.e., in our speeded binary choice task). Another possibility is that L2 speakers are more likely than monolingual native speakers to compute an underspecified or shallow syntactic representation of German pseudo-partitive subjects (e.g., Clahsen & Felser, 2006), leading to their agreement preferences being determined more strongly by surface-level factors such as a noun's proximity to the verb or by the relative greater salience of plural versus singular noun phrases, irrespective of the nature of the task. Although our study was not designed to systematically examine L1 transfer, if the lack of agreement with plural numerals in Turkish influences our bilingual participants' agreement preferences in German, they should show a general preference for singular over plural verbs and generally rate the singular verb conditions more highly compared to the monolinguals.

## **Experiment 1**

### *Method*

*Participants.* We recruited 40 German native speakers (mean age: 28.8 years; SD: 11.2) via the University's participant database, Facebook, and Email. None of them reported to have any developmental or language disorder or to have grown up bilingual. All of them learnt English at school, and the majority also had classes in at least one other foreign language (e.g., French, Spanish, and Latin), but none of them reported to have any knowledge of Turkish.<sup>2</sup>

Our bilingual group consisted of 41 Turkish–German bilinguals (mean age: 29.4 years, SD: 8.5) who had all learned Turkish from birth. One participant was later excluded due to their low German proficiency. The age of acquisition (AoA) of German of the remaining 40 participants ranged from 0 to 27 years (mean: 7.9, SD: 7.49), with a length of immersion to German between 2 and 40 years (mean: 19.5, SD: 9.97). The group's average weekly usage of German was 73% (SD: 17.5, range 10–90%). Approximately half of the group considered German as their dominant language.

The bilingual participants were recruited via Facebook, through Flyers on streets and metro stations, and word of mouth. All bilingual participants filled in an extensive questionnaire about their language biography, also including the amount of usage of the two languages in all four domains (reading, writing, listening, and speaking), self-rated language skills, their preferred language, language classes attended, their age of arrival in Germany (if not born there), among others. Additionally, their German proficiency was tested using a subset from the Telc

*Kompetenzcheck Deutsch Beruf-Test* (telc gGmbH, 2015), which consisted of 5 short cloze tests, each with a maximum of 10 points for correct answers. Participants who scored less than 30 points (out of 50) were excluded from further analysis ( $n=1$ , as mentioned above). The remaining participants' mean German proficiency score was 42.15 (SD: 5.2), which placed them within the B2–C2 range according to the Common European Framework of Reference (CEFR; Council of Europe, 2001).

*Design and materials.* The experiment had a 2x2x2 design with the factors NP1 number (singular vs. plural), NP2 number (singular vs. plural), and verb number (singular vs. plural). We created 24 experimental sentence sets which contained pseudo-partitive subjects in 8 conditions.<sup>3</sup> The first noun was always a container noun, as in *eine Packung Tabletten* ("one pack of pills"). In half of the sentences, the substance noun was a pluralized count noun (such as *Tabletten* "pills") and in the other half we used a non-countable noun (such as *Zucker* "sugar"). This was necessary for us to be able to manipulate the second noun's grammatical number, because singular count nouns cannot serve as substance nouns in German pseudo-partitives. All critical sentences contained predicative adjectives followed by the third person present tense singular or plural form of the auxiliary *sein* ("to be"), that is, by either *ist* ("is") or *sind* ("are"). All experimental sentences started with a main clause of the form "X said" or "X believed," and our pseudo-partitive subjects always appeared in a finite complement clause introduced by *dass* ("that"). The fact that German embedded clauses are verb-final allowed us to make sure that the critical auxiliary was always the final word. Our eight experimental conditions are illustrated in (6a–h) below.

- (6) a. NP1 PLURAL - NP2 PLURAL - VERB PLURAL (PPP)  
 Nina glaubt, dass drei Packungen Tabletten vorhanden sind.  
*Nina believes that three packs pills at hand are.*  
 "Nina believes that three packs of pills are at hand."
- b. NP1 PLURAL - NP2 PLURAL - VERB SINGULAR (PPS)  
 Nina glaubt, dass drei Packungen Tabletten vorhanden ist.  
*Nina believes that three packs pills at hand is.*  
 "Nina believes that three packs of pills is at hand."
- c. NP1 PLURAL - NP2 SINGULAR - VERB PLURAL (PSP)  
 Nina glaubt, dass drei Packungen Zucker vorhanden sind.  
*Nina believes that three packs sugar at hand are.*  
 "Nina believes that three packs of sugar are at hand."
- d. NP1 PLURAL - NP2 SINGULAR - VERB SINGULAR (PSS)  
 Nina glaubt, dass drei Packungen Zucker vorhanden ist.  
*Nina believes that three packs sugar at hand is.*  
 "Nina believes that three packs of sugar is at hand."
- e. NP1 SINGULAR - NP2 PLURAL - VERB PLURAL (SPP)  
 Nina glaubt, dass eine Packung Tabletten vorhanden sind.  
*Nina believes that one pack pills at hand are.*  
 "Nina believes that one pack of pills are at hand."
- f. NP1 SINGULAR - NP2 PLURAL - VERB SINGULAR (SPS)  
 Nina glaubt, dass eine Packung Tabletten vorhanden ist.  
*Nina believes that one pack pills at hand is.*  
 "Nina believes that one pack of pills is at hand."

- g. NP1 SINGULAR - NP2 SINGULAR - VERB PLURAL (SSP)  
 Nina glaubt, dass eine Packung Zucker vorhanden sind.  
*Nina believes that one pack sugar at hand are.*  
 “Nina believes that one pack of sugar are at hand.”
- h. NP1 SINGULAR - NP2 SINGULAR - VERB SINGULAR (SSS)  
 Nina glaubt, dass eine Packung Zucker vorhanden ist.  
*Nina believes that one pack sugar at hand is.*  
 “Nina believes that one pack of sugar is at hand.”

Conditions (6a) and (6h) served as grammatical baselines because both NPs match the verb's number. In conditions (6b) and (6g), on the other hand, neither NP1 nor NP2 agrees with the verb, so that these conditions served as ungrammatical baselines. Of primary interest are conditions (6c–f), where NP1 and NP2 differ in their grammatical number.

The plausibility of the pseudo-partitive noun phrases (e.g., *drei Packungen Tabletten* “three packs of pills”) was rated by 34 German native speakers (who did not participate any further in this study) on a scale from 1 (“very plausible”) to 5 (“completely implausible”). There was no significant difference in the plausibility rating between the pseudo-partitive constructions containing mass nouns compared to count nouns ( $t = 0.15$ ,  $df = 23$ ,  $p\text{-value} = 0.88$ ). Mass nouns and count nouns were also matched for frequency based on total counts of lemma frequency from the dLex Corpus (<http://www.dlexdb.de/Qgmilhc>) ( $t = 0.04$ ,  $df = 23$ ,  $p\text{-value} = 0.97$ ) but could not be matched for length in letters ( $t = 3.1$ ,  $df = 23$ ,  $p\text{-value} = 0.005$ ), since the count nouns carried a plural affix which made them almost always longer than the mass nouns.

The experimental items were distributed over eight lists in a Latin square design. Each list contained 24 critical sentences, plus 22 sentences from a different experiment which served as fillers. This experiment investigated morphosyntactic variation in a different phenomenon and used stimulus sentences that were structurally different from our experimental sentences. We also added 10 additional fillers, some of which were similar to our critical sentences but contained clear grammatical errors, and no pseudo-partitives. This resulted in a total of 66 sentences per list.

*Procedures.* Eight versions of our web-based questionnaire were created using Google Forms, which were counterbalanced between participants. Our monolingual participants received a link to the experiment and could complete it at home, and the bilingual group was tested under supervision in a laboratory room in Berlin. The current experiment was part of a larger test battery for this group of participants, and the order of experiments and test languages was counterbalanced between participants. The judgment task could be completed within 20–30 min.

Participants saw one sentence at a time and could continue to the next item via button press after rating it. Sentences were rated on a scale from 1 (“highly acceptable”) to 5 (“absolutely unacceptable”).<sup>4</sup> Participants were allowed as much time as they needed to provide their judgments. They were instructed to follow their intuition regarding the sentences’ acceptability and that the task was not about giving “right” or “wrong” answers. Participants received either course credit or a small fee for participation.

**Table 1.** Mean acceptability ratings (and standard deviations) by condition, Experiment 1

| Condition | Monolinguals | Bilinguals  |
|-----------|--------------|-------------|
| PPP       | 1.43 (0.87)  | 1.78 (1.2)  |
| PPS       | 4.49 (1.1)   | 4.03 (1.29) |
| PSP       | 1.18 (0.56)  | 1.73 (1.07) |
| PSS       | 4.66 (0.8)   | 4.10 (1.23) |
| SPP       | 4.14 (1.2)   | 3.28 (1.62) |
| SPS       | 1.50 (1.03)  | 2.58 (1.56) |
| SSP       | 4.60 (0.75)  | 3.95 (1.3)  |
| SSS       | 1.29 (0.62)  | 1.82 (1.15) |

For statistical analysis, we used linear mixed-effects models on the (log-transformed) ratings. Fixed factors were number of verb (singular vs. plural, “Verb” in future reference), NP1 match or mismatch with verb (“NP1” in future reference), NP2 match or mismatch with verb (“NP2” in future reference), and Group, with crossed random effects for participants and items (see, e.g., Baayen *et al.*, 2008). For the random effects structure of the model, we followed current guidelines in psycholinguistics, and we initially constructed a maximal model that included random intercepts and slopes for all fixed effects and their interactions (Barr *et al.*, 2013). When this maximal model failed to converge, we gradually simplified the random effects structure following the suggestions by Barr *et al.* (2013). The remaining random slope structure of the omnibus model over all factors and both groups included one interaction by participant (Verb x NP1), and two interactions by item (NP1 x Verb, NP1 x Group). For main effects and overall interactions, factors employed sum-coded contrasts (−0.5, 0.5).<sup>5</sup> For separate group analysis, we conducted models on each of the two groups’ data sets.

### Results

The two participant groups’ mean acceptability ratings and standard deviations are shown in Table 1.

As can be seen in Table 1, both groups rated the grammatical baseline conditions (PPP and SSS) as highly acceptable and the ungrammatical conditions (PPS and SSP) as rather unacceptable. When NP1 and NP2 carried conflicting number features, agreement with NP1 was generally rated more favorably than agreement with NP2. However, if NP1 mismatched the verb’s number, a matching NP2 improved acceptability compared to the corresponding ungrammatical baseline, at least when NP2 was plural (SPP > SSP). Conversely, the presence of a plural NP2 reduced the acceptability of grammatical sentences with singular agreement (SPS < SSS).

All significant results of the overall analysis with sum-coded contrasts are shown in Table 2. The analysis revealed a main effect of NP1, with NP1-match sentences rated significantly better than NP1-mismatch sentences across both groups. There was also a significant main effect of NP2, indicating that matching NP2s increased

**Table 2.** Main effects and interactions, Experiment 1

|                    | Estimate | SE   | t     |
|--------------------|----------|------|-------|
| Group              | -0.04    | 0.04 | -1.03 |
| NP1                | 2.5      | 0.1  | 21.98 |
| NP2                | -0.2     | 0.04 | -4.75 |
| Verb               | 0.28     | 0.07 | 4.45  |
| NP1xNP2            | -0.02    | 0.03 | -0.51 |
| NP1xVerb           | -0.01    | 0.05 | -0.24 |
| NP2 x Verb         | -0.008   | 0.03 | -0.22 |
| NP1 x Group        | 1.12     | 0.2  | 6.03  |
| NP2 x Group        | 0.27     | 0.08 | 3.17  |
| Verb x Group       | -0.3     | 0.13 | -2.32 |
| NP1 x NP2 x Verb   | 1.3      | 0.17 | 7.96  |
| NP1xNP2 x Group    | -0.09    | 0.07 | -1.44 |
| NP1xVerb x Group   | 0.01     | 0.09 | 0.14  |
| NP2 x Verb x Group | 0.002    | 0.07 | 0.03  |
| NP1xNP2xVerbxGroup | -0.15    | 0.14 | -1.1  |

$\text{lmer}(\log(\text{Rating}) \sim \text{NP1} * \text{NP2} * \text{Verb} * \text{Group} + (1 + \text{NP1} * \text{Verb}) | \text{Participant}) + (1 + \text{NP1} * \text{Verb} + \text{NP1} * \text{Group}) | \text{Item}, \text{data})$

acceptability compared to NP2s whose number feature mismatched that of the verb. The main effect of verb number reflects the fact that overall, sentences with a plural auxiliary were rated as more acceptable than sentences with a singular auxiliary. We also found a significant three-way interaction between the factors NP1, NP2, and verb number, indicating that participants' judgments of sentences with singular versus plural auxiliaries were affected by both NPs' number but in a nonuniform way. Each of the three within-participants' factors interacted with the factor Group, but there were no significant three-way interactions involving Group and no four-way interaction.

To further explore the between-group differences that gave rise to the observed two-way interactions, additional analyses were carried out. The NP1 x Group interaction reflects the fact that although both groups preferred NP1-match sentences over NP1-mismatch sentences, this difference was larger in the monolingual group (ML: 1.35 vs. 4.48; BL: 1.98 vs. 3.84). The observed Verb x Group interaction is due to the bilingual group showing overall higher ratings for plural than for singular verbs, compared to the monolingual group (ML: 2.84 vs. 2.98; BL: 2.69 vs. 3.13).

Of particular interest to us is the interaction of NP2 with Group, which reflects the fact that the bilinguals rated NP2-match sentences significantly better than NP2-mismatch sentences (2.75 vs. 3.07,  $\beta = 0.32$ , SE = 0.07,  $t = 4.38$ ), whereas monolinguals did not show a significant difference (2.88 vs. 2.95,  $\beta = 0.07$ , SE = 0.06,  $t = 1.04$ ). For both groups, the effect of NP2 seems to be stronger when NP2 is plural, which likely gave rise to the three-way interaction reported in Table 2 above. We

therefore conducted further analyses on NP2 effects, splitting the data by NP2 number. The results revealed a significant NP2 Match  $\times$  Group interaction for sentences with plural NP2s ( $\beta = 0.30$ ,  $SE = 0.09$ ,  $t = 3.22$ ) which was not significant for sentences with singular NP2s ( $\beta = 0.04$ ,  $SE = 0.06$ ,  $t = 0.61$ ). The interaction indicates that a matching versus mismatching plural NP2 affected the two groups' judgments differently. While neither group showed any effects of a singular NP2 on acceptability, a plural NP2 increased the acceptability of sentences with plural verbs significantly for both groups if NP1 mismatched the verb in number (SPP > SSP: ML:  $\beta = -0.45$ ,  $SE = 0.12$ ,  $t = -3.98$ ; BL:  $\beta = -0.65$ ,  $SE = 0.14$ ,  $t = -4.73$ ). Conversely, in NP1-match sentences, a mismatching plural NP2 decreased acceptability for both groups (SPS < SSS: ML:  $\beta = 0.23$ ,  $SE = 0.1$ ,  $t = 2.25$ ; BL:  $\beta = 0.76$ ,  $SE = 0.14$ ,  $t = 5.49$ ). For the monolingual group only, the presence of two matching plural NPs decreased acceptability relative to sentences in which only NP1 matched the verb's plural feature (PPP < PSP:  $\beta = 0.25$ ,  $SE = 0.1$ ,  $t = 2.5$ ).

To examine whether our bilingual speakers' AoA of German, their length of exposure to German, language dominance, usage of German, or their proficiency affected their performance, we ran additional analyses for this group only, including all fixed factors as described above, except Group (i.e., NP1 match/mismatch, NP2 match/mismatch, and verb number). We did not find any effects of, or interactions with, any of these factors besides participants' German proficiency. Proficiency interacted with NP1 ( $\beta = 0.098$ ,  $SE = 0.03$ ,  $t = 3.12$ ) such that with increasing proficiency, NP1-match sentences were rated more favorably and NP1-mismatch sentences less favorably. There was also a two-way interaction with NP2 ( $\beta = -0.03$ ,  $SE = 0.01$ ,  $t = -2.1$ ), due to NP2-mismatch conditions being rated worse with increasing proficiency. A three-way interaction between NP1, verb number and proficiency ( $\beta = 0.07$ ,  $SE = 0.04$ ,  $t = 2.0$ ) reflects the fact that the effect of proficiency on NP1 match/mismatch was stronger for sentences with a singular than a plural auxiliary.

In sum, our results show that both groups prefer the verb to agree with NP1, with this preference being stronger in monolinguals than in bilinguals. Additionally, both groups showed effects of a plural NP2, such that in sentences with a singular NP1 and a plural NP2, plural verbs were rated better compared to sentences with two singular nouns, while in cases of NP1-match with the verb, a mismatching plural NP2 led to a lower acceptability. This effect was stronger in the bilinguals than in the monolinguals, and not attested for plural NP1 followed by a singular verb, where the number of NP2 did not have any effect. The bilinguals' judgment patterns were found to be modulated by their German proficiency only.

### *Modeling*

The primary goal in developing models of subject-verb agreement for German pseudo-partitives is to illustrate what factors speakers rely upon most when selecting between singular or plural verbs. To this end, we take a stochastic constraint-based approach following stochastic or noisy HG (Boersma & Pater, 2016; Hayes, 2017) or GSC (Smolensky et al., 2014). Following this approach, candidates are evaluated against a set of weighted constraints to generate a Harmony value (H, the sum of all constraint violations multiplied by their respective weights). No candidate

is chosen as the single optimal candidate, as is the case in similar frameworks such as OT or classic HG. Instead, the probability of each candidate is calculated based on the distribution of the Harmony across the set. The closer to 0 the Harmony is, the more likely the candidate is to occur. This stochastic approach lends itself to modeling variation, such as cases of doubt, because all candidates with a nonzero probability are potential outputs.

Based on the research into number agreement and pseudo-partitives reviewed above, we postulated a set of constraints. First, we know that the verb must agree with one of the noun phrases, which motivated constraints (7a) and (7b). Second, we developed a constraint to capture the proximity effects found in the literature. Constraint (7c) evaluates the relationship between NP2 and the verb. Our constraint set comprises the following three constraints:

(7) a. AGRNP1: Agreement with NP1

Assign 1 violation to the candidate when the verb does not agree with NP1.

b. AGRNP2: Agreement with NP2

Assign 1 violation to the candidate when the verb does not agree with NP2:

c. NP2-V(SP > PS): Minimally maintain number features from NP2 to Verb.

Assign 1 violation to the candidate when the number feature of the verb is *less than* the number feature of NP2. E.g., a singular NP2 followed by a singular or plural verb receives no violation ( $\text{Num}_V \geq \text{Num}_{\text{NP2}}$ ), but a plural NP2 followed by a singular verb receives 1 violation ( $\text{Num}_V < \text{Num}_{\text{NP2}}$ ).

In this initial step of developing a model, the goal is to fit the model to the existing data. In other words, existing data are used to inform the constraint weights. Tools such as MaxEnt Grammar Tool (Wilson, 2006) are available to calculate constraint weights based on corpus data, on the basis of potential candidates and their frequency in the corpus, the constraints and the violations for each candidate of those constraints. With no corpus data, we look instead to the results of the judgment task to represent the frequency of the given candidates. This meant converting the speaker ratings into speaker preferences. One immediate question that arises is whether or not judgments reflect what speakers would actually produce. However, there is no perfect data source, as was pointed out by Poplack and Torres Cacoullos (2016). They argued that the frequency of phenomena such as blended structures in a corpus does not accurately reflect speaker production. If a structure does not appear in a corpus, that does not automatically mean that the structure cannot occur. Similarly, if a structure does appear in a corpus, it may be over- or under-represented depending on the context. Since acceptability judgment tasks are often used to tap into native speaker intuitions and uncover allowances the grammar makes that may not appear in production, we hope that converting speaker ratings to speaker preferences will reflect the flexibility of the grammar that is not necessarily present in corpus data.

The experimental conditions were paired by noun combination, differing only by number on the verb, for example, PPP and PPS. The assumption is that the noun phrases are the input to the grammar, and the output is the noun phrases with either a singular or plural verb. For each of these pairs, we tallied the favorable and unfavorable responses. Ratings of 1 and 2 were considered in favor of the candidate and

**Table 3.** Converted speaker judgments, Experiment 1

| Input | Output | Monolingual |                           | Bilingual  |                           |
|-------|--------|-------------|---------------------------|------------|---------------------------|
|       |        | Preference  | N of removed ratings of 3 | Preference | N of removed ratings of 3 |
| PP    | PPP    | 0.93        | 6                         | 0.82       | 4                         |
|       | PPS    | 0.07        | 4                         | 0.18       | 14                        |
| PS    | PSP    | 0.97        | 4                         | 0.86       | 9                         |
|       | PSS    | 0.03        | 3                         | 0.14       | 7                         |
| SP    | SPP    | 0.11        | 7                         | 0.44       | 11                        |
|       | SPS    | 0.89        | 6                         | 0.56       | 11                        |
| SS    | SSP    | 0.03        | 4                         | 0.20       | 7                         |
|       | SSS    | 0.97        | 5                         | 0.80       | 4                         |

responses of 4 and 5 were considered in favor of the complementary candidate. For example, a rating of 1 for a PPP item is in favor of the plural verb with two plural nouns. A rating of 5 for a PPP item is in favor of PPS instead, as the only alternative to PPP.<sup>6</sup> Responses of 1 and 2, or 4 and 5 were grouped together as favorable toward one candidate, which lead to a categorical transformation of the speaker ratings into speaker preferences. A gradient transformation, in which 1 holds more value than 2, or 5 more than 4, would be ideal, but the nature of the AJT is such that we cannot be certain that the speakers all value a 1 or a 2 similarly. Therefore, we adopt this binary approach. In this vein, ratings of 3 were discarded, because it was not possible to interpret what the speakers would prefer. The end result of transforming the speakers' ratings was a preference ratio for singular versus plural verbs given the combination of noun phrases, calculated by adding all favorable ratings for each candidate and dividing that by the total number of ratings for the candidate pair.

We begin with the monolingual group and will then compare the model based on this group with the bilinguals.

Table 3 presents the preference ratios for the input–output pairs for the monolingual and bilingual participants, and the total number of middle-point ratings of 3 that were excluded. For double plural NP input, the monolingual speakers had a near-categorical preference for the plural verb (PPP, 0.93) over the singular verb (PPS, 0.07), whereas the bilingual speakers were more accepting of the singular verb (PPS, 0.18). Monolingual speakers preferred a singular verb for the mismatched SP input (SPS, 0.89), but not categorically (SPP, 0.11). In contrast, the bilingual speakers showed no clear preference for the singular or the plural verb (SPP, 0.44; SPS, 0.56). These ratios are then used as target figures to fit the constraint weights.

Table 4 illustrates how the model works for the given input of two NPs, with target outputs of either the NP pair with a plural or a singular verb. Each candidate is evaluated against the constraints and assigned the appropriate number of violations. For instance, the candidate PSP incurs a violation under the constraint AgrNP2. These violations are multiplied by the weight of the respective constraint, and the total penalty across the row is the Harmony of the candidate. The probability of each candidate is then calculated using the formula in (8), following Hayes

Table 4. Model of monolingual group, Experiment 1

| Input:<br>NP1&2<br>(PP) |     | *AGRNP1<br>-5 | *AGRNP2<br>-2 | NP2-V(SP>PS)<br>-1 | H  | Pr   | Target |
|-------------------------|-----|---------------|---------------|--------------------|----|------|--------|
| a.                      | PPP | 0             | 0             | 0                  | 0  | 1.00 | 0.93   |
| b.                      | PPS | -5            | -2            | -1                 | -8 | 0.00 | 0.07   |
| Input:<br>NP1&2<br>(PS) |     | *AGRNP1<br>-5 | *AGRNP2<br>-2 | NP2-V(SP>PS)<br>-1 | H  | Pr   | Target |
| c.                      | PSP | 0             | -2            | 0                  | -2 | 0.95 | 0.97   |
| d.                      | PSS | -5            | 0             | 0                  | -5 | 0.05 | 0.03   |
| Input:<br>NP1&2<br>(SP) |     | *AGRNP1<br>-5 | *AGRNP2<br>-2 | NP2-V(SP>PS)<br>-1 | H  | Pr   | Target |
| e.                      | SPP | -5            | 0             | 0                  | -5 | 0.12 | 0.11   |
| f.                      | SPS | 0             | -2            | -1                 | -3 | 0.88 | 0.89   |
| Input:<br>NP1&2<br>(SS) |     | *AGRNP1<br>-5 | *AGRNP2<br>-2 | NP2-V(SP>PS)<br>-1 | H  | Pr   | Target |
| g.                      | SSP | -5            | -2            | 0                  | -7 | 0.00 | 0.03   |
| h.                      | SSS | 0             | 0             | 0                  | 0  | 1.00 | 0.97   |

(2017), in which  $H(c)$  represents the Harmony of the candidate and  $H(x)$  the Harmonies in the set:

$$(8) Pr(c) = e^{H(c)} / \sum_x e^{H(x)}$$

The constraint weights were determined through a simple manual manipulation: all of the constraints were set to an arbitrary equal weight (e.g., -5). This generates a predicted probability for the candidates. The goal of the manipulation is to bring that set of predicted probabilities as close to the targets obtained from the judgment task as possible. The baseline, arbitrary constraint weights are adjusted in increments of  $\pm 0.5$  or  $\pm 1$  until the point of *diminishing returns* is met. This is the point at which adjustments to that constraint weight do not yield further improvements to the model. Through this approach, the constraint weights were determined simultaneously for all four inputs, to ensure that the model’s predicted probabilities approached the targets as closely as possible under the assumption that the constraint weights do not fluctuate depending on the input, that is, the grammar is the grammar. In Table 4, candidate a has a Harmony of 0 and candidate b a Harmony of -8, which results in a categorical predicted probability that candidate a should occur 100% of the time. The end result of this manipulation is categorical predictions for the inputs where both NPs have the same number (PP and SS), and

**Table 5.** Model of bilingual group, Experiment 1

| Input:<br>NP1&2 |     | *AGRNP1 | *AGRNP2 | NP2-V(SP>PS) | H     | Pr   | Target |
|-----------------|-----|---------|---------|--------------|-------|------|--------|
| (PP)            |     | -5      | -3.5    | -1.25        |       |      |        |
| a.              | PPP | 0       | 0       | 0            | 0     | 1.00 | 0.82   |
| b.              | PPS | -5      | -3.5    | -1.25        | -9.25 | 0.00 | 0.18   |
| Input:<br>NP1&2 |     | *AGRNP1 | *AGRNP2 | NP2-V(SP>PS) | H     | Pr   | Target |
| (PS)            |     | -5      | -3.5    | -1.25        |       |      |        |
| c.              | PSP | 0       | -3.5    | 0            | -3.5  | 0.82 | 0.86   |
| d.              | PSS | -5      | 0       | 0            | -5    | 0.18 | 0.14   |
| Input:<br>NP1&2 |     | *AGRNP1 | *AGRNP2 | NP2-V(SP>PS) | H     | Pr   | Target |
| (SP)            |     | -5      | -3.5    | -1.25        |       |      |        |
| e.              | SPP | -5      | 0       | 0            | -5    | 0.44 | 0.44   |
| f.              | SPS | 0       | -3.5    | -1.25        | -4.75 | 0.56 | 0.56   |
| Input:<br>NP1&2 |     | *AGRNP1 | *AGRNP2 | NP2-V(SP>PS) | H     | Pr   | Target |
| (SS)            |     | -5      | -3.5    | -1.25        |       |      |        |
| g.              | SSP | -5      | -3.5    | 0            | -8.5  | 0.00 | 0.20   |
| h.              | SSS | 0       | 0       | 0            | 0     | 1.00 | 0.80   |

predicted probabilities for the mismatched input that come very close to matching the transformed judgments. The monolingual and bilingual models generated using this approach are presented in Tables 4 and 5.

### Discussion

Our judgment results confirm that the grammatical baseline conditions (PPP and SSS) are considered highly acceptable, while the ungrammatical conditions (PPS and SSP) are generally deemed unacceptable. The conditions with number conflicts between NP1 and NP2 show a clear preference for NP1 as the agreement controller, which indicates a general preference for the container over the quantity reading. This is consistent with the findings reported by Smith *et al.* (2018) for pseudo-partitives with container nouns in English.

We also found evidence of NP2 affecting participants' judgments. Sentences containing a matching NP2 were rated more highly than sentences that contained a mismatching NP2, but only if NP2 was plural. This asymmetry suggests that accepting NP2 as the agreement controller was not uniformly facilitated by NP2's proximity to the verb, a factor which has been argued to play a role in determining number agreement by Wegerer (2012) and others. The observed singular/plural asymmetry is consistent with theoretical claims to the effect that plural NPs are

marked or more salient in memory than singular NPs, thus increasing the likelihood that they are selected as agreement controllers if NP1 is singular. Note that in our experimental materials, a plural-marked NP1 was always introduced by a plural numeral, making NP1's plurality very obvious and thus reducing the likelihood of a singular NP2 triggering verbal agreement.

Some group differences were also observed, with the bilinguals showing higher overall ratings for sentences with plural auxiliaries, a smaller difference between NP1-match and NP1-mismatch sentences, and stronger NP2 influence in comparison to monolinguals. The bilingual group's overall preference for plural auxiliaries is inconsistent with transfer from Turkish, where pseudo-partitives never trigger plural agreement. The prediction derived from usage-based approaches that the bilinguals might focus more strongly on NP1 number than the monolinguals was also disconfirmed. Instead, the bilingual group's judgments were more strongly affected by NP2 number than were the monolingual group's judgments. Unlike what we had expected, the bilingual speakers' judgments were not measurably affected by the age at which they acquired German. We did find effects of their German proficiency, however, albeit not in the direction of the overall judgment pattern becoming more nativelike: bilinguals with higher German proficiency showed a larger discrepancy between both NP1 and NP2 match versus mismatch conditions but no evidence of reduced influence of a matching NP2.

In our GSC models, the weights of the constraints reflect the experimental results. AgrNP1 carries more than twice as much weight than AgrNP2 for monolingual speakers, meaning a stronger penalty is assigned for a mismatch in number between the verb and NP1, making NP1 the primary criterion for determining verbal agreement. Nonetheless, NP2's plurality—while not as strong a factor as the number of NP1—plays a role in that failing to minimally maintain the number of NP2 in the verb has a detrimental effect on a candidate. This can be seen most clearly in candidate f in Table 4 above. The mismatch between NP2 and the verb, compounded by the fact that NP2 is plural but the verb is singular (violating NP2-V(SP>PS)), brings the Harmony of the candidates closer together (candidate e: -5, candidate f: -3).

The relative ranking of the constraints (AgrNP1 > AgrNP2 > NP2-V(SP > PS)) held for both the monolingual and bilingual groups. However, we found that increasing the weight of AgrNP2 by 1.5 and the weight of NP2-V(SP > PS) by 0.25 resulted in a better fit for the bilingual judgment data. This reflects the importance placed on NP2 apparent in the bilingual judgments. By making these adjustments, the difference in Harmony values for the candidates in the NP1-mismatch conditions (Table 5, candidates c–f) is reduced. This allows the candidates to be near-equal in predicted probability, as seen in Table 5, candidates e and f.

Nonetheless, the fit of the model to the bilingual data is not ideal in the baseline conditions. Due to the nature of the constraints, the model generates categorical Harmony values for the grammatical conditions. Neither of the groups demonstrated categorical preferences in these conditions, although the monolinguals were close and the deviation from categorical could be attributed to momentary lapses in attention or any other number of factors. The bilinguals' preferences of approximately 0.8 for the grammatical candidate and approximately 0.2 for the ungrammatical candidate may stem from criteria not currently reflected in the set of

constraints that the grammatical candidate violates but the ungrammatical does not. The non-categorical judgment of our baseline conditions may reflect a lack of confidence in participants' judgments (compare e.g., Huang & Ferreira, 2020), especially among bilinguals. Our bilingual group contrasted the grammatical and ungrammatical baseline conditions less strongly (Table 1), and more often chose the middle point of the scale, than the monolinguals did (Table 3). There is also often a response bias toward acceptance in AJTs (Hammerly *et al.*, 2019), so that higher degree of confidence may be required in order to reject an ungrammatical stimulus item compared to accepting a grammatical one.

## Experiment 2

Our second experiment was a speeded forced-choice task (Staub, 2009), where participants had to choose a verb form (plural or singular) as their preferred sentence completion. The aim was to establish whether the same preference patterns would emerge in a production task as we saw in Experiment 1, and whether participants' agreement choices could be predicted from the above constraint weightings.

### Method

*Participants.* We recruited 47 German native speakers (mean age: 23.8 years; SD: 4.6) via the University's participant database. Our bilingual group consisted of 60 Turkish-German bilinguals (mean age: 32.2 years, SD: 9.3) who had all learned Turkish from birth. Fifteen of this group had also previously taken part in Experiment 1, but since the two testing sessions were almost 1 year apart, this was unlikely to affect the outcome of the current experiment.

The data from eight bilingual participants were excluded prior to analysis because of more than 50% errors in the baseline conditions (PP and SS). The remaining 52 participants had a mean AoA of German of 10.6 years (SD: 7.7, range 0–30), and their mean length of immersion was 19.9 years (SD: 12.7, range 2–54). Their mean proficiency in German of 43.2 out of 50 (SD: 5.4, range 31–50), which places this group's mean at the C1 level according to CEFR. Participants were recruited in the same way as for Experiment 1. Their average use of German in a normal week was 66%, and 19 of them considered German their dominant language. They filled in the same questionnaire about their language history and completed the same proficiency test as did the bilingual participants in Experiment 1.

*Materials.* We used the same materials as in Experiment 1, the only difference being that participants had to choose between a clause-final singular or plural auxiliary (*ist* "is" vs. *sind* "are") to complete the sentences. The number of experimental conditions was therefore reduced to four, with only NP1 and NP2 number manipulated, as illustrated in (9a–d).

- (9) a. NP1 PLURAL- NP2 PLURAL (PP)  
 Nina glaubt, dass drei Packungen Tabletten vorhanden ...  
 Nina believes that three packs pills at hand ...
- b. NP1 PLURAL - NP2 SINGULAR (PS)  
 Nina glaubt, dass drei Packungen Zucker vorhanden ...  
 Nina believes that three packs sugar at hand ...

- c. NP1 SINGULAR - NP2 PLURAL (SP)  
 Nina glaubt, dass eine Packung Tabletten vorhanden ....  
*Nina believes that one pack pills at hand ...*
- d. NP1 SINGULAR - NP2 SINGULAR (SS)  
 Nina glaubt, dass eine Packung Zucker vorhanden ...  
*Nina believes that one pack sugar at hand ...*  
 'Nina believes that one/three pack/s of pills/sugar ... at hand.'

The 24 critical sentence quadruplets were distributed over four presentation lists and mixed with 24 sentences from a different experiment that investigated subject-verb agreement with coordinated noun phrases and with 29 additional fillers with varying structures. Each list included 77 sentences in total. Before the experiment proper, participants completed eight practice sentences.

*Procedures.* The experiment was set up using the web-based platform Ibx Farm (Drummond, 2013), which allows for reaction time experiments to be carried out remotely. Each sentence was presented one word at a time in black letters on white background, at a rate of 500 ms per word plus a 50 ms inter-word pause. After the final word, two response options were presented (*ist* "is" vs. *sind* "are"). Participants were instructed to select the auxiliary that they thought provided a grammatical continuation to the sentence by pressing either the F or J key on the keyboard. Time-out was set at 2,000 ms, and participants received a warning message when reaching it. The position of the words was kept unchanged throughout the experiment. Experimental and filler sentences were randomized on a by-participant basis. Each experimental session began with eight practice trials. After each response, participants pressed a key to start the next trial. No feedback was provided except the time-out warning.

The monolingual group completed the experiment at home via a web link. The whole experiment lasted about 20 min, and participants received either course credit or a small fee for participation. The bilingual group was tested in a laboratory room. As with the first experiment, for the bilinguals this experiment was part of a larger test battery. The order of test languages and experiments was counterbalanced between participants. They received a fee for participation.

We analyzed the proportion of response choices (plural or singular) and reaction times (RTs). Data cleaning included the removal of time-outs (0.09% data points for the monolinguals and 2.6% for the bilinguals). To prevent results from being influenced by extreme outliers, we set a cutoff at below 50 ms and above 1,500 ms, which led to the further removal of 1.6% data points for the monolinguals and 5.2% for the bilinguals.

Responses were analyzed with mixed-effects logistic regression (Jaeger, 2008) with the bobyqa optimizer (Powell, 2009) applied, and RTs were analyzed with mixed-effects linear models (Baayen et al., 2008) and were log-transformed before analysis. Fixed factors were NP1 number, NP2 number, and Group. We used the same approach of finding the model with the maximal random structure that converged as described above. The model for the RTs included both factors by item and participant, but without interaction. The model for responses included the factors NP1 and NP2 by participant, and NP1 by item.

**Table 6.** Mean reaction times in milliseconds and proportions of plural responses (with standard deviations in parentheses), Experiment 2

| Condition | Reaction times |            | % Plural responses |             |
|-----------|----------------|------------|--------------------|-------------|
|           | Monolinguals   | Bilinguals | Monolinguals       | Bilinguals  |
| PP        | 512 (249)      | 650 (321)  | 97.6 (15.4)        | 93.4 (24.9) |
| PS        | 501 (253)      | 646 (340)  | 95.8 (20.0)        | 82.6 (38.0) |
| SP        | 474 (255)      | 706 (334)  | 6.0 (23.8)         | 29.2 (45.6) |
| SS        | 451 (235)      | 628 (326)  | 2.3 (15.3)         | 4.5 (20.7)  |

*Predictions.* If the same constraint weightings which underlie speakers' judgments also determine their verb number choices in a task that resembles production, we expect to see the same preference patterns here as we did in Experiment 1. However, using a fixed word-by-word presentation rate and timing participants' responses put them under greater processing pressure in comparison to Experiment 1. If the influence of NP2 number that we observed in Experiment 1 was a (proximity-related) attraction effect, then we might see this effect increase in the current task because the mental representation of NP1, including its number feature, may be comparatively more difficult to retrieve or re-access at the point at which the verb must be chosen.

### Results

The proportions of participants' plural responses and mean RTs can be seen in Table 6.

The bilinguals took longer to respond on average than the monolinguals across all conditions, which was reflected in a main effect of Group (Estimate =  $-0.28$ , SE =  $0.1$ ,  $t = -2.7$ ). We also found a significant interaction of NP1 and Group (Estimate =  $-0.14$ , SE =  $0.057$ ,  $t = -2.5$ ), due to the monolinguals responding faster to singular than to plural NP1 conditions and the bilinguals showing the reversed pattern. There were no other significant effects or interactions in the RT data. Note that RT data may not be very informative, however, since the decision about verb number could in principle be made earlier during the presentation of the subject phrase, and not only when the question prompt appeared. We will thus not analyze or discuss the RT results any further.

The proportions of plural responses seen in Table 6 again reveal a strong preference for agreement with NP1 across both groups. In conditions with a singular NP1, a higher rate of plural responses was elicited when NP2 was plural compared to when both NPs were singular. The difference between the NP2 singular and plural conditions was smaller for sentences containing a plural NP1. A rather striking group difference appears in the SP condition, where the bilinguals were nearly five times as likely as the monolinguals to choose a plural auxiliary.

Results of the omnibus model on the responses, with sum-coded contrasts for all factors, are presented in Table 7.

**Table 7.** Main effects and interactions for plural responses, Experiment 2

|                   | Estimate | SE   | z      |
|-------------------|----------|------|--------|
| Group             | -0.03    | 0.26 | -0.11  |
| NP1               | -6.18    | 0.36 | -17.05 |
| NP2               | -1.01    | 0.26 | -3.8   |
| NP1 x NP2         | -1.1     | 0.5  | -2.05  |
| NP1 x Group       | -2.96    | 0.5  | -5.7   |
| NP2 x Group       | 1.05     | 0.4  | 2.5    |
| NP1 x NP2 x Group | 0.93     | 0.84 | 1.13   |

glmer(responses ~ NP1\*NP2\*Group + (1+NP1+NP2|Participant) + (1+NP1|Item), data, family = "binomial", control = glmerControl(optimizer = "bobyqa"))

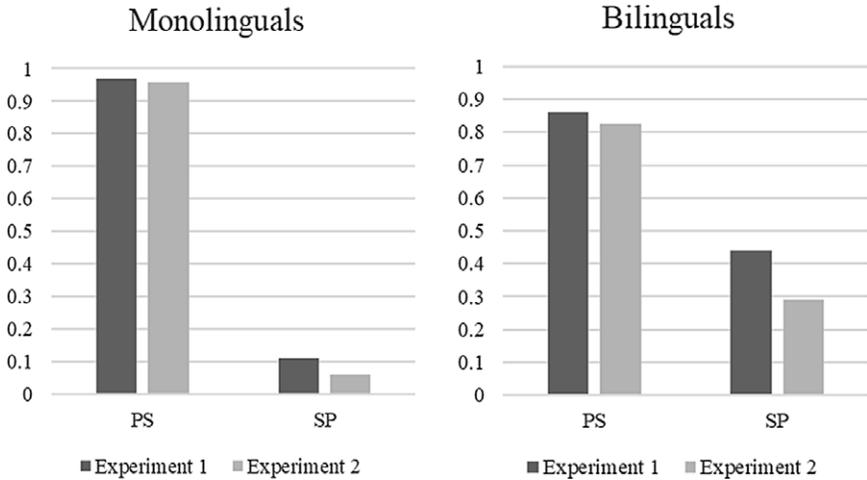
Main effects of both NP1 and NP2 were observed, and the two factors also interacted, indicating that the likelihood of NP1 triggering agreement was modulated by NP2 number. Both NP1 and NP2 interacted with the factor Group. These two-way interactions are due to the NP1 agreement preference being stronger in the monolingual group and the difference between singular and plural NP2 conditions larger in the bilingual group. By releveling for NP1 number, we found again a significant NP2 by Group interaction for sentences with a singular NP1 ( $\beta = 1.46$ ,  $SE = 0.58$ ,  $z = 2.53$ ); this interaction was not significant for plural NP1 ( $\beta = 0.64$ ,  $SE = 0.6$ ,  $z = 1.12$ ). The interaction reflects the fact that the increase in plural responses triggered by a plural-marked NP2 (SP > SS) was larger in our bilingual group, although it is statistically significant for both groups (ML:  $\beta = -1.78$ ,  $SE = 0.98$ ,  $z = -1.99$ ; BL:  $\beta = -2.5$ ,  $SE = 0.44$ ,  $z = -5.6$ ).

Additional analyses of the bilingual group's responses revealed a significant interaction of NP2 number and proficiency (Estimate = -0.1,  $SE = 0.03$ ,  $t = -3.1$ ) but no effects or interactions with German AoA. The interaction with proficiency reflects the fact that participants with higher German proficiency were less likely to choose a plural verb in the presence of a plural NP2 compared to less proficient participants. This suggests that the impact of NP2 on the choice of verb number weakened with increasing German proficiency. Other factors such as length of exposure to German, language dominance, or usage of German did not yield any significant effects or interactions.

#### *Comparison of outcomes and GSC models for Experiments 1 & 2*

The speaker's choices were used as ratios and compared with the transformed judgments. Participants' agreement preferences were generally very similar across the two experiments (see Table 8 for a full overview), although both participant groups chose fewer plural verbs in the number conflict conditions than their judgments might have led us to expect. Figure 1 illustrates the two experiments' outcomes for conditions with a number conflict (PS and SP).

The monolingual group's results demonstrate that the transformed speaker judgments from Experiment 1 could indeed be used to represent speaker preferences.



**Figure 1.** Ratios of plural responses based on converted speaker judgments (Experiment 1) and agreement choices (Experiment 2) for number conflict cases for each participant group.

**Table 8.** Comparison of speaker performance from Experiment 1 (judgment) and Experiment 2 (forced choice), and model predictions for both groups (ML: monolinguals, BL: bilinguals)

| Input | Output | Judgment |      | Forced choice |      | Predicted probabilities |      |
|-------|--------|----------|------|---------------|------|-------------------------|------|
|       |        | ML       | BL   | ML            | BL   | ML                      | BL   |
| PP    | PPP    | 0.93     | 0.82 | 0.98          | 0.93 | 1.00                    | 1.00 |
|       | PPS    | 0.07     | 0.18 | 0.02          | 0.07 | 0.00                    | 0.00 |
| PS    | PSP    | 0.97     | 0.86 | 0.96          | 0.83 | 0.95                    | 0.82 |
|       | PSS    | 0.03     | 0.14 | 0.04          | 0.17 | 0.05                    | 0.18 |
| SP    | SPP    | 0.11     | 0.44 | 0.06          | 0.29 | 0.12                    | 0.44 |
|       | SPS    | 0.89     | 0.56 | 0.94          | 0.71 | 0.88                    | 0.56 |
| SS    | SSP    | 0.03     | 0.20 | 0.02          | 0.05 | 0.00                    | 0.00 |
|       | SSS    | 0.97     | 0.80 | 0.98          | 0.95 | 1.00                    | 1.00 |

Based on the comparison of the transformed judgments and the results of the forced-choice task, the model we developed using the results of Experiment 1 remained almost valid when reevaluated using the results of Experiment 2. The model’s predictions were able to capture speakers’ behavior, although the predictions from the model are more categorical than speaker behavior. This is likely due to a number of factors that could not be represented or incorporated into the model, such as item-specific interpretations or extraneous reasons for selecting/rating an item unrelated to the relationship between the number of the NPs and the verb.

Table 8 shows a comparison of participants' preferences in Experiments 1 and 2, and the model's predictions. The set of candidates with an input of singular NP1 and plural NP2 (SP candidates) is the source of the most discrepancy between the speaker's production and the model's predictions. While the transformed judgments align well with the model's predictions (not surprising, given that they were used to generate the model), the weight of the NP2-V(SP > PS) constraint needed to be reduced by 0.75 to  $-0.25$  to bring the predictions for the SP candidates set in line with speaker behavior on the forced-choice task. With this new weight, the predictions for these candidates change to 0.06 for SPP and 0.94 for SPS, which is perfectly aligned with the results.

PP and SS candidates remained unaffected by any changes to the model due to the categorical nature of the violations incurred in the model by these candidates.

Regarding the bilingual results, we can see similar alignments and discrepancies between the transformed judgments and the forced-choice results as for the monolingual group, with SP candidates showing the highest deviation from the models' prediction (as well as the transformed judgments). By applying the same weight change as for the monolinguals, that is, reducing the weight of NP2-V(SP > PS) by 0.75 from  $-1.25$  to  $-0.5$ , the model's predictions align more closely with the results of Experiment 2: 0.73 for SPS and 0.27 for SPP. The other candidates with number conflict did not change much between Experiments 1 and 2, same as for the monolinguals.

Given that the model predicts categorical behavior for the PP and SS candidate sets, but the speakers do not behave categorically, we can assume that there is some allowable variability in the predictions. A question that remains open at this point is how much variability, or deviation from the predicted probabilities, is allowed by the model itself or is attributable to constraints we were unable to include, such as semantic factors or even L2 status and comprehension skills of the participants.

## General discussion

Our study is the first to systematically investigate German speakers' preferences for singular versus plural agreement with pseudo-partitive subject phrases, and the first to use GSC as an example of a stochastic modeling in an attempt to quantify the relative strength of the constraints that determine participants' judgments. To explore the extent to which these constraints and their relative weightings can be acquired under suboptimal acquisition conditions, we compared native German speakers with native Turkish speakers who acquired German as a non-native language or alongside their family language. We also examined whether number agreement preferences remained consistent across different tasks (judgment vs. production). Given the different nature of the tasks, we asked whether a single model could account for both the judgment and the forced-choice production results.

Our main finding is that both monolingual and bilingual German speakers generally preferred NP1 as the agreement controller for pseudo-partitive subjects with container nouns. Nevertheless, NP2 also affected participants' preference patterns but in an asymmetrical way: Plural NP2s were more likely to trigger plural

agreement than singular NP2s were to trigger singular agreement. The results from our two experiments aligned rather well, demonstrating that the constraint weights computed from our judgment data could predict speakers' performance in a task that resembled production. Although both participant groups showed similar performance patterns, we also found some group differences, most notably a stronger influence of NP2 number on the bilingual speakers' judgments and production choices. The AoA of German did not measurably affect our bilingual participants' agreement preferences, but proficiency did. Recall that variable number agreement with pseudo-partitive subjects can be formally accounted for by their structural ambiguity. Analyzing NP1 as the head of the overall subject phrase—and hence, as the agreement controller—is assumed to correspond to a referential or “container” reading (Selkirk, 1977). Although our study was not designed to probe participants' semantic interpretation of pseudo-partitive subjects directly, their general preference for NP1 to determine verbal agreement suggests a preference for a referential over a quantity reading of container nouns.

Agreement with NP2 was also permitted to some degree, however. Our GSC models captured participants' response patterns by assigning a larger weight to the requirement that the verb agree with NP1 than to the constraint favoring agreement with NP2, and with an additional (but weaker) constraint accounting for the observed singular/plural asymmetry. A parallel asymmetry has previously been reported in the literature on agreement attraction. Although permissible variation of the kind we have investigated is different from ungrammaticality, it is conceivable that the same processing mechanisms or constraints that give rise to agreement attraction errors also influence speakers' agreement computation with pseudo-partitives. The fact that plural NP2s had a larger effect on agreement than singular NP2s is consistent with the assumption that plural NPs are more salient than singular ones and thus more likely to trigger agreement (e.g., Eberhard, 1997; Staub, 2009). In our study, the hypothesized greater salience of plural noun phrases might have made a plural NP2 more likely to be analyzed as the subject head compared to a singular NP2. What we cannot tell on the basis of our results, however, is whether agreement with a plural NP2 reflects an initial analysis of the subject phrase in which NP2 is the head, with agreement computed correctly, or whether it reflects the erroneous retrieval of a noun phrase that was not encoded as the subject head. Our finding that added time pressure in Experiment 2 did not lead to a larger NP2 effect could be taken to indicate that the observed NP2 influence on participants' agreement preferences does not reflect a memory retrieval error.

The lack of any major task effects in our study—in contrast, for example, to the task effects observed by Franck *et al.* (2015) for AJT versus self-paced reading—would be expected if our two tasks do not actually tap into different processes. In both of our tasks, participants had to evaluate how well a given verb form fitted the context. However, note that in Experiment 2, participants knew that they would have to select a singular or plural auxiliary following the presentation of the subject phrase, and it was possible for them in principle to make this decision upon having seen NP1 only. This could also perhaps explain why NP2 influence did not increase in Experiment 2 despite added time pressure and the fact that participants could only see one word at a time.

A different line of explanation could come from studies such as Villata and Franck (2020) which show that in comprehension tasks often (though not always) an attractor matching the features of the verb actually facilitates processing (in terms of reading times and accuracy), in contrast to agreement attraction studies where the majority of responses still matches the agreement controller. This difference between comprehension and production studies could also explain the higher ratings for SPP items compared to production of plural verbs in SP items. Still, this facilitation effect has not been found for grammatical sentences, and as stated above, in our number conflict conditions both verb forms are grammatical (see e.g., Shen et al., 2013; Tanner et al., 2014; Tucker et al., 2015, among others).

Another difference between the two tasks is that in Experiment 1 participants were able to provide graded judgments, whereas in Experiment 2 they were forced to choose between two candidate verb forms, which always resulted in one of the forms being selected as the winner, even though both might have been deemed acceptable to some degree. This might explain the somewhat higher amount of variation observed across almost all conditions in Experiment 1 compared to Experiment 2.

Our two participant groups' agreement preferences were very similar. GSC modeling yielded the same relative weighting of the three constraints for both groups, showing that gradient constraints on subject–verb number agreement can be acquired even under suboptimal acquisition conditions. However, in cases of a number conflict between the two NPs, the bilinguals accepted and chose a plural-marked NP2 as an agreement controller more readily than the monolingual group did. This was reflected in our models in that the constraint favoring agreement with NP2 carried more weight in the bilingual than in the monolingual model, and thus reducing the impact of the constraint favoring agreement with NP1.

As we noted earlier, the observed group differences are unlikely to reflect influence from Turkish, a language in which pseudo-partitive subjects always take a singular (or unmarked) verb. Neither did our bilingual participants over-apply NP1 agreement relative to the monolinguals, as might have been expected from the perspective of usage or frequency-based approaches to L2 acquisition and use. Rather, the bilingual groups' stronger focus on NP2 number, in particular in the SP condition, might indicate a stronger influence of noun proximity on agreement preferences, amplified by plural-marked NP2s being more salient than singular ones. In other words, it appears that agreement computation is affected by surface-level cues (noun proximity and plural morphology) more strongly in bilinguals than in monolinguals. Note that from the perspective of resource limitation accounts of L1/L2 differences, we might have expected added time pressure to further enhance group differences (compare e.g., Sato & Felser, 2010), but this was not the case. The results of Experiment 2 did in fact show smaller between-group differences than did the results from Experiment 1.

It is conceivable that our bilingual participants were more likely than the monolinguals to compute an underspecified or shallow representation of the subject phrase when this was first encountered, or that they had more difficulty re-accessing the subject's representation for agreement computation (compare Reifegerste et al., 2020). Uncertainty about which of the two NPs was or should be taken as the syntactic head would lead participants' verb form choices to be influenced by other

factors, including noun proximity. We might expect more proficient L2 speakers to be more likely to create fully specified grammatical representations than less proficient ones. The analysis of the bilinguals' verb choices in Experiment 2—albeit not the results from Experiment 1—do indeed suggest that the influence of NP2 decreases with increasing German proficiency.

One important question that we have not addressed here is the potential influence of semantic factors on speakers' agreement choices such as the degree to which NP1 lends itself to a quantity reading (compare Smith *et al.*, 2018), or the notional plurality of the overall noun phrase. That is, *eine Packung Tabletten* ("one pack of pills") may be more likely to be conceived of as referring to multiple entities compared to *eine Packung Zucker* ("one pack of sugar"). Sensitivity to semantic factors may differ between monolinguals and bilinguals. Previous studies of agreement production observed a lack of susceptibility to conceptual number in less proficient L2 speakers (Hoshino *et al.*, 2010; Jackson *et al.*, 2018), for instance. Future investigation that includes semantic factors would add to further understanding the variability observed here in regard of L1 and L2 agreement preferences in pseudo-partitive sentences. Another factor of interest is the role of NP modification or "heaviness," as this may affect an NP's relative salience and the likelihood of it being selected as an agreement controller. These factors should also ultimately be included in a more comprehensive multiple-constraint model of agreement with pseudo-partitives.

## Conclusion

Using two different experimental tasks in combination with GSC modeling, we examined variability in number agreement with German pseudo-partitives. Our results show that for pseudo-partitive subjects with container nouns, agreement with NP1 is generally preferred. Monolingual and Turkish/German bilingual speakers showed similar agreement preferences, which demonstrates that the constraints that determine agreement, and their relative weighting, can be acquired even under suboptimal input conditions. The obtained group differences are argued to be due to our bilingual group's agreement choices being more prone to influences of surface-level information in comparison to the monolinguals. Future research will show whether this observation generalizes to other language combinations and to other phenomena displaying optionality or permissible variation. On the methodological side, we hope to have demonstrated the usefulness of using weighted constraint-based models of grammar for capturing subtle differences between monolingual and bilingual speakers' performance patterns.

**Acknowledgments.** We thank our student assistants Daria Fisuneneko, Marilena Tsopanidi, Simge Sargin, and Cilem Cicek for their help with the data collection. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer 317633480—SFB 1287, Project B04.

## Notes

1. Note that German *ein/e* is ambiguous between the indefinite determiner "a" and the numeral "one."
2. We will subsequently refer to this group as "monolinguals" so as to contrast them with our Turkish/German bilingual participants.

3. A complete list of our experimental stimuli are accessible at the Open Science Framework website at <https://osf.io/ns7gm/>.
4. We used a scale that resembles German school grades, where “1” means “excellent” and “5” translates as “insufficient.”
5. Factors are coded as follows: Group: bilinguals = -0.5, monolinguals = 0.5; NP1: plural = -0.5, sing = 0.5; NP2: plural = -0.5, sing = 0.5.
6. Of course, there are other alternatives. Without asking the participants for the reason they rated the item the way they did, we can only assume that the critical variable of verb number is the deciding factor.

## References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.
- Berg, T. (1998). The resolution of number conflicts in English and German agreement patterns. *Linguistics*, *36*, 41–70.
- Bock, K., & Miller, C. A. (1991). Broken Agreement. *Cognitive Psychology*, *23*, 45–93.
- Boersma, P., & Pater, J. (2016). Convergence properties of a gradual learning algorithm for Harmonic Grammar. In J. McCarthy & J. Pater (Eds.), *Harmonic Grammar and Harmonic Serialism*. Sheffield, (pp. 389–434). Sheffield: Equinox.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.
- Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge University Press.
- Drummond, A. (2013). Ibox farm. Online server: <http://spellout.net/ibexfarm>.
- Eberhard, K. M. (1997). The Marked Effect of Number on Subject–Verb Agreement. *Journal of Memory and Language*, *36*, 147–164.
- Franck, J., Colonna, S., & Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Frontiers in Psychology*, *6*, 349. <https://doi.org/10.3389/fpsyg.2015.00349>
- Goldrick, M., Putnam, M., & Schwarz, L. (2016). Coactivation in bilingual grammars: A computational account of code mixing. *Bilingualism: Language and Cognition*, *19*, 857–876.
- Grestenberger, L. (2015). Number marking in German measure phrases and the structure of pseudo-partitives. *The Journal of Comparative Germanic Linguistics*, *18*(2), 93–138.
- Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology*, *110*, 70–104.
- Hartsuiker, R. J., Schriefers, H. J., Bock, K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory & Cognition*, *31*, 1316–1326.
- Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject–verb agreement. *Journal of Memory and Language*, *48*(4), 760–778.
- Haskell, T. R., & MacDonald, M. C. (2005). Constituent structure and linear order in language production: Evidence from subject–verb agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 891–904.
- Haskell, T. R., Thornton, R., & MacDonald, M. C. (2010). Experience and grammatical agreement: Statistical learning shapes number agreement production. *Cognition*, *114*(2), 151–164.
- Häussler, J. (2009). The Emergence of Agreement Attraction Errors During Sentence Comprehension. Doctoral dissertation, University of Konstanz, Germany.
- Hayes, B. (2017). Varieties of noisy harmonic grammar. Proceedings of the Annual Meetings on Phonology (Vol. 4), 1–17. <https://doi.org/10.3765/amp.v4i0.3997>
- Hennig, M., Schneider, J. G., Osterwinter, R., & Steinhauer, A. (2016). Duden – das Wörterbuch der sprachlichen Zweifelsfälle. In M. Hennig (Ed.), *Der Duden in zwölf Bänden – das Standardwerk der deutschen Sprache (Band 9)*. Berlin: Dudenverlag.

- Hölscher, C., & Hemforth, B.** (2000). Subject-Verb Agreement in German: Evidence from Production and Comprehension Experiments. In B. Hemforth and L. Konieczny (Eds.), *German Sentence Processing* (pp. 279–310). Dordrecht: Kluwer Academic Publishers.
- Hoshino, N., Dussias, P. E., & Kroll, J. F.** (2010). Processing subject-verb agreement in a second language depends on proficiency. *Bilingualism: Language and Cognition*, **13**, 87–98.
- Huang, Y., & Ferreira, F.** (2020). The application of signal detection theory to acceptability judgements. *Frontiers in Psychology*, **11**. doi: [10.3389/fpsyg.2020.00073](https://doi.org/10.3389/fpsyg.2020.00073)
- Humphreys, K. R., & Bock, K.** (2005). Notional number agreement in English. *Psychonomic Bulletin & Review*, **12**(4), 689–695.
- Jackson, C. N., Mormer, E., & Brehm, L.** (2018). The production of subject-verb agreement among Swedish and Chinese second language speakers of English. *Studies in Second Language Acquisition*, **40**(4), 907–921.
- Jaeger, C.** (1992). *Probleme der syntaktischen Kongruenz*. Tübingen: Niemeyer.
- Jaeger, T. F.** (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, **59**(4), 434–446.
- Lago, S., & Felser, C.** (2018). Agreement attraction in native and non-native speakers of German. *Applied Psycholinguistics*, **39**, 619–647.
- MacWhinney, B.** (2005). A unified model of language acquisition. In J. Kroll & A. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 49–67). Oxford: Oxford University Press.
- McDonald, J.** (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgement performance by late second language learners. *Journal of Memory and Language*, **55**, 381–401.
- Poplack, S., & Cacoullos, R. T.** (2016). Data before models. *Bilingualism: Language and Cognition*, **19**(5), 893–894.
- Powell, M. J.** (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06*, University of Cambridge, Cambridge, 26–46.
- Reifegerste, J., Jarvis, R., & Felser, C.** (2020). Effects of aging on native and nonnative sentence processing: Evidence from subject-verb agreement in German. *Journal of Memory and Language*, **111**, 104083.
- Sağ, Y.** (2018). The semantics of Turkish numeral constructions. In U. Sauerland & S. Solt (Eds.), *Proceedings of Sinn und Bedeutung 22*, vol. 2, ZASPiL 61, pp. 307–324. ZAS, Berlin.
- Sato, M., & Felser, C.** (2010). Sensitivity to morphosyntactic violations in English as a second language. *Second Language*, **9**, 101–118.
- Scontras, G.** (2014). *The Semantics of Measurement*. Doctoral dissertation, Harvard University.
- Selkirk, E.** (1977). Some remarks on noun phrase structure. In A. Akmajian, P. Culicover, & T. Wasow (Eds.), *Studies in formal syntax* (pp. 285–316). New York: Academic Press.
- Shen, E. Y., Staub, A., & Sanders, L. D.** (2013). Event-related brain potential evidence that local nouns affect subject-verb agreement processing. *Language and Cognitive Processes*, **28**(4), 498–524.
- Slabakova, R.** (2019). The Bottleneck Hypothesis updated. In T. Ionin & M. Rispoli (Eds.), *Three streams of generative language acquisition research: Selected papers from the 7th meeting of generative approaches to language acquisition–North America, University of Illinois at Urbana-Champaign* (pp. 319–345). Amsterdam: John Benjamins.
- Smith, G., Franck, J., & Tabor, W.** (2018). A self-organizing approach to subject-verb number agreement. *Cognitive Science*, **42**, 1043–1074.
- Smolensky, P., Goldrick, M., & Mathis, D.** (2014). Optimization and quantization in gradient symbol systems: a framework for integrating the continuous and the discrete in cognition. *Cognitive Science*, **38**(6), 1102–1138.
- Staub, A.** (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, **60**, 308–327.
- Stickney, H.** (2009). *The Emergence of DP in the Partitive Structure*. Doctoral dissertation, Amherst University, Massachusetts.
- Tanner, D., Nicol, J., & Brehm, L.** (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of memory and language*, **76**, 195–215.
- telc gGmbH** (Ed) (2015). *Kompetenzcheck Deutsch Beruf, A1 A2 B1 B2 C1*. Frankfurt am Main, 1. Auflage.
- Thornton, R., & MacDonald, M. C.** (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, **48**(4), 740–759.

- Tucker, M. A., Idrissi, A., & Almeida, D.** (2015). Representing number in the real-time processing of agreement: Self-paced reading evidence from Arabic. *Frontiers in Psychology*, *6*, 347.
- Vigliocco, G., Butterworth, B., & Semenza, C.** (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, *34*(2), 186–215.
- Villata, S., & Franck, J.** (2020). Similarity-based interference in agreement comprehension and production: Evidence from object agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(1), 170.
- Wegerer, M.** (2012). Die Numeruskongruenz von Subjekt und finitem Verb im Deutschen. Doctoral dissertation, Universität Wien, Austria.
- Wilson, C.** (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science*, *30*(5), 945–982.

---

**Cite this article:** Jessen, A., Schwarz, L., and Felser, C. (2021). Gradience in subject–verb number agreement: Can bilinguals tune in? *Applied Psycholinguistics* *42*, 1523–1551. <https://doi.org/10.1017/S0142716421000357>