

## Animacy effects in the English genitive alternation

Dubois, Tanguy; Grafmiller, Jason; Szmrecsanyi, Benedikt; Paquot, Magali

*License:*

Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

*Document Version*

Peer reviewed version

*Citation for published version (Harvard):*

Dubois, T, Grafmiller, J, Szmrecsanyi, B & Paquot, M 2023, 'Animacy effects in the English genitive alternation: comparing native speakers and EFL learner judgments with corpus data', *Language and Cognition*.

[Link to publication on Research at Birmingham portal](#)

**Publisher Rights Statement:**

This is the Accepted Author Manuscript of an article to be published in a revised form in *Language and Cognition*, available at <http://journals.cambridge.org/action/displayJournal?jid=LCO>. This version is published under a Creative Commons CC-BY-NC-ND licence. No commercial re-distribution or re-use allowed. Derivative works cannot be distributed. © copyright holder.

**General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

**Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

1           **Animacy effects in the English genitive alternation: comparing native**  
2                           **speakers and EFL learner judgments with corpus data**

3  
4   Recent years have seen a heightened interest in the interface between language use and  
5   cognition in language learners. In this study, we investigate this interface further by  
6   conducting a rating task experiment on the intuitions of 25 native speakers and 101 low-  
7   intermediate to advanced learners of English as a Foreign Language regarding the  
8   acceptability of the genitive variants (*the beauty of nature/nature's beauty*) in different  
9   contexts. These ratings were then compared against existing corpus-based statistical models  
10  that predict which variant is most likely in spoken language use with two mixed-effects linear  
11  regression models. The first model focused on the animacy of the possessor in particular,  
12  which has been found to have a different effect for native speakers and EFL learners in  
13  language use, whereas the second model tested how the ratings relate to the predictions as a  
14  whole. Results show that there is a larger discrepancy between language use and intuitions of  
15  low-proficiency learners compared to native speakers, which is partially due to the fact that  
16  animate, collective and inanimate possessors affect the intuitions and the language use of  
17  learners differently.

18  
19  Keywords: rating task experiment, EFL learners, genitive alternation, usage-based SLA

## 1. Introduction

1 Recently, usage-based approaches to Second Language Acquisition (SLA) have investigated  
2 how learners of English as a Foreign Language (EFL) use the structural variants of alternation  
3 phenomena compared to native speakers (e.g., Jäschke & Plag, 2016; Kinne, 2020; Wulff &  
4 Gries, 2019). Usage-based approaches to SLA assume that language learning is driven by  
5 language experience (Ellis & Wulff, 2019, p. 41), and the same applies to the acquisition of the  
6 probabilistic constraints on the choice between alternating variants (Wulff & Gries, 2019, p.  
7 880). This raises the question as to how the probabilistic grammar of EFL learners develops as  
8 they become more proficient. With this question in mind Dubois et al. (2022) investigated how  
9 low-intermediate to advanced EFL learners from various L1 backgrounds (e.g., Spanish, Italian,  
10 Hindi, Chinese, Russian) choose between the *s*-genitive (1) and the *of*-genitive (2).

12

13 (1) the people's<sub>possessor</sub> mentality<sub>possessum</sub> (TLC: 2\_IN\_11)

14 (2) the mentality<sub>possessum</sub> of the people<sub>possessor</sub> (TLC: 2\_6\_ME\_69)

15

16 On the methodological plane, they collected 2302 genitive observations in the Trinity  
17 Lancaster Corpus (TLC, Gablasova, Brezina & McEnery 2019), a three-million-word corpus  
18 consisting of transcribed recordings from an official spoken language exam, namely the  
19 Graded Examinations in Spoken English (GESE). They annotated the observations for the  
20 probabilistic constraints that drive the choice of genitive variant for native speakers and then  
21 analyzed whether the learners differ from the native speakers with regard to the effect of any  
22 of the constraints using mixed-effects logistic regression (e.g., Gries, 2015). In line with  
23 previous research (Gries & Wulff, 2013), they found that learners and native speakers are  
24 remarkably similar: both are sensitive to structural persistence (e.g., Bock, 1982; Szmrecsanyi,

1 2006), the length of the possessor and the possessum, the final sibilancy of the possessor and  
2 the lexical density of the surrounding context (Dubois et al., 2022, p. 17).

3 Both native speakers and learners are also sensitive to possessor animacy, which  
4 captures the tendency for animate possessors to favor the *s*-genitive and inanimate  
5 possessors to prefer the *of*-genitive. One explanation for the effect of the animacy constraint  
6 is that speakers prefer to first produce linguistic elements that are easy to process (MacDonald,  
7 2013). Animate referents, which are easy to retrieve from memory (Branigan et al., 2008),  
8 would therefore be produced early, which means that an animate possessor favors the *s*-  
9 genitive where the possessor comes first. While animacy is normally the strongest constraint  
10 of the genitive alternation (Rosenbach, 2014), Dubois et al. (2022) found that B1 speakers are  
11 more likely than native speakers to use the *s*-genitive when the possessor is inanimate,  
12 whereas B2 speakers are more likely than native speakers to use the *of*-genitive when the  
13 possessor is animate (Dubois et al., 2022). This finding is reminiscent of other studies on the  
14 genitive in World Englishes and the development of the English genitive over time where the  
15 strength of the animacy constraint was found to fluctuate (Heller et al., 2017; Wolk et al.,  
16 2013).

17 Although Dubois et al. (2022) and similar studies aim to shed light on the implicit  
18 grammatical knowledge of learners regarding the constraints, there are fundamental  
19 differences between how statistical models and humans gather knowledge about  
20 distributional patterns. For example, regression models like the one used by Dubois et al.  
21 (2022) have difficulty dealing with multicollinearity, yet collinearity and redundancy are an  
22 integral part of language (Milin et al., 2016, p. 508). In general, corpus-based models are able  
23 to “reflect the characteristics of mental processes and structures yielding usage, even though  
24 we do not know the exact form of these mental representations” (Divjak & Arppe, 2013, p.

1 230). Put simply, corpus-based models show what speakers say, but not necessarily what they  
2 know (Ellis, 2017, p. 46; Kinne, 2020, p. 173). Applied to the genitive alternation, this means  
3 that it is unclear whether the probabilistic constraints that affect language use according to  
4 the corpus-based regression model actually derive from the implicit grammatical knowledge  
5 of speakers (Klavan & Divjak, 2016), or whether they are simply frequent patterns of language  
6 use.

7         Therefore, we compare the development of EFL learners' patterns of language use  
8 according to the regression model of Dubois et al. (2022) against the acceptability judgments  
9 of learners at different proficiency levels in a rating task experiment. Whereas experimental  
10 data is typically less valid ecologically than corpus data, rating task experiments do not  
11 primarily concern language production and instead tap into the intuitions of speakers  
12 regarding the acceptability of the variants (Arppe & Järvikivi, 2007, p. 132; Gilquin & Gries,  
13 2009; Klavan & Divjak, 2016, p. 361), which offers a different perspective on the development  
14 of probabilistic grammars for EFL learners across proficiency levels. This combination of corpus  
15 data and experimental ratings highlights the way in which cognition, operationalized as the  
16 speakers' knowledge of probabilistic constraints on variation, comes to the fore in EFL learners  
17 at different stages of language learning in "two different language usage situations, namely  
18 production and introspection" (Arppe & Järvikivi, 2007, p. 150). In our case, such an  
19 experiment would provide complementary evidence for the cognitive reality of the  
20 probabilistic constraints guiding the choice of genitive variant for native speakers and EFL  
21 learners and shed more light on the finding that learners are less sensitive to the animacy of  
22 the possessor while also providing insights on the development of the animacy constraint  
23 throughout language learning in general.

1            Similar experiments have been implemented successfully to complement corpus-  
2 based research for the particle placement alternation, the position of the prepositional phrase,  
3 the dative alternation and the future marker alternation in English (e.g., Bresnan & Ford, 2010;  
4 Engel et al., 2022; Engel & Szmrecsanyi, 2023; Kinne, 2020). It has also been used to  
5 complement corpus-based models with more than two possible variants in other languages,  
6 such as Russian (Divjak & Arppe, 2013) and Estonian (Klavan, 2020). While these studies  
7 generally find that the results from corpus-based models and the experiments largely  
8 converge (Klavan & Divjak, 2016, p. 17), the predictions from the corpus-based models do not  
9 correlate in a linear fashion with the acceptability of the variants according to individual  
10 speakers (Arppe & Järvikivi, 2007, pp. 149–150; Divjak et al., 2016, pp. 26–27; Kempen &  
11 Harbusch, 2005, p. 330). This is partially attributable to the fact that regression models are  
12 largely concerned with how frequently the variants occur in certain contexts. Indeed, the fact  
13 that a genitive variant is used very rarely in a certain context does not mean that speakers  
14 deem the variant unacceptable (Arppe & Järvikivi, 2007, pp. 149–150; Divjak et al., 2016, pp.  
15 26–27; Kempen & Harbusch, 2005, p. 344).

16            Against this backdrop, our study investigates the following research questions:

17

- 18        - How does possessor animacy affect the intuitions of EFL learners at different  
19            proficiency levels in a rating task experiment compared to correlations between  
20            animacy and genitive variant in a corpus of spoken learner language?
- 21        - How does the relationship between experimental ratings and corpus data vary by  
22            proficiency level of participants and speakers?

23

24        For our first research question, we expect that participants find the *s*-genitive more natural

1 when the possessor is animate, but this tendency should be weaker for B1 and B2 learners  
2 compared to native speakers in accord with the results from the regression model of Dubois  
3 et al. (2022), henceforth referred to as the reference corpus model. This means that  
4 preference for the *s*-genitive with animate rather than inanimate possessors should be less  
5 pronounced for B1 and B2 learners compared to native speakers. For our second research  
6 question, we expect that the intuitions from the learners in the experiment correspond to the  
7 patterns of language use. That is, the ratings from the participants in the experiment should  
8 correlate with the predictions from the reference corpus model, showing that the participants  
9 are sensitive to probabilistic constraints (Bresnan & Ford, 2010). For example, if the model  
10 predicts that the *s*-genitive is more likely in a given context, we expect participants to give  
11 higher ratings for the *s*-genitive in that same context as well.

12         Analysis of the ratings shows that the A2/B1 and B2 learners participating in the  
13 experiment are less sensitive to certain animacy distinctions involving animate, collective, and  
14 inanimate possessors. Overall, the intuitions from the participants in the experiment correlate  
15 with the patterns of language use of the speakers in the corpus, which provides additional  
16 evidence that the participants are sensitive to the significant probabilistic constraints included  
17 in the reference corpus model. However, this correlation is weaker for A2/B1 and B2  
18 participants, indicating that these learners do not always produce the genitive variant they  
19 prefer in the experiment.

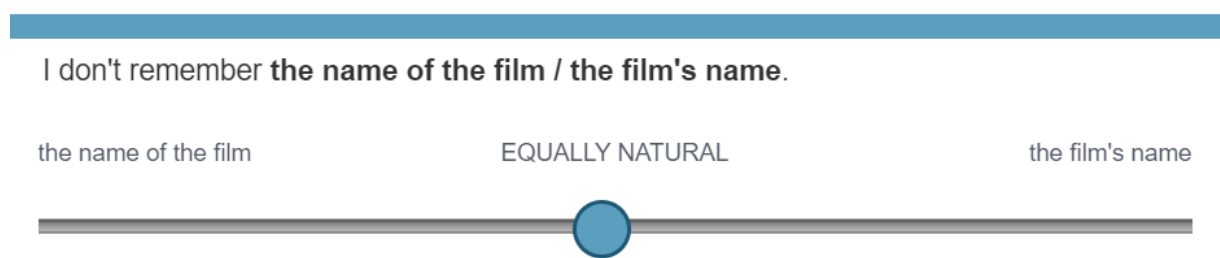
20         The structure of the paper is as follows. The methodology including the materials,  
21 participants, procedure and design of the experiment are presented in Section 2. The results  
22 from two mixed-effects linear regression models are presented in Section 3 and then  
23 discussed in Section 4.

24

## 1 2. Methodology

2 The rating task experiment was similar to the one used by Bresnan and Ford (2010) and more  
3 recently Engel et al., (2022). The participants, who were native speakers and learners of  
4 English, were shown an original corpus excerpt with both possible genitive variants. Using a  
5 slider bar, participants could indicate which variant they preferred. The position of the slider  
6 represented a rating from 0 to 100 for the *s*-genitive (*the film's name* in Figure 1). Higher  
7 scores indicated that the *s*-genitive was more natural, with a score of 100 indicating that this  
8 was the only acceptable variant in this context. Inversely, a score of 0 indicated that this  
9 variant was not acceptable at all compared to the *of*-genitive.

10



12 *Figure 1: An example of a target experimental item presented to participants*

13

14 The experiment focused in particular on the effect of possessor animacy on the ratings since  
15 B1 and B2 learners are less sensitive to this constraint according to the reference corpus model.  
16 Dubois et al. (2022) originally distinguished between five animacy levels based on the coding  
17 scheme of Zaenen and colleagues (2004; see also Wolk et al., 2013), viz. animate, collective,  
18 inanimate, locative and temporal possessors. Animate possessors “mostly concern humans or  
19 entities represented as humans and animals” (e.g. *god's plan*) (Dubois et al., 2022, p. 10).  
20 Collective possessors covered collective nouns like *government* or *company* and groups of  
21 animates who display an identity as a group (e.g., *the values of the parents*). Locatives referred



## Animacy in the genitive alternation

1 to locations that are fit for humans, as in *the south of France*. Possessors that were marked as  
2 temporal refer to a time period or a specific point in time (e.g., *next week*). All other possessors  
3 were considered inanimate.

4 Due to data sparsity issues however, the constraint was conflated into a binary  
5 distinction between animate and inanimate possessors in the reference corpus model. As a  
6 result, it is not clear whether animacy plays a less important role in the language use of B1  
7 and B2 learners compared to native speakers because of specific animacy levels. In this respect,  
8 Dubois et al. (2022) argue that learners struggle with locative and collective possessors in  
9 particular because many possessor noun phrases can fit either category depending on the  
10 context. In example (3) for instance, copied from example (11) in Dubois et al. (2022), *South*  
11 *Africa* could either refer to a location or a collective group of people.

12

13 (3) and er in May nineteen ninety four for the first time in South Africa's history all the  
14 races voted in democratic election (TLC:2\_6\_IT\_100)

15

16 In order to shed more light on the effect of animacy for our first research question, the rating  
17 task experiment was designed to test the effect of all five animacy levels on the ratings.  
18 Moreover, the experimental items stem from the 2302 observations Dubois et al. (2022)  
19 collected from the Trinity Lancaster Corpus (Gablasova et al., 2019) for the reference corpus  
20 model. This is relevant for the second research question because it allows us to compare the  
21 predictions from the reference corpus model, which reflect the patterns of language use of

Animacy in the genitive alternation

1 the speakers in the corpus, with the ratings of the participants in the experiment representing  
2 their intuitions regarding the naturalness of the variants.<sup>1</sup>

3

## 4 **2.1 Materials**

5 The rating task experiment consisted of target experimental items and items copied from the  
6 Oxford Quick Placement Test, a short proficiency test consisting of 60 questions of different  
7 types, most of which require the participant to fill in the blank in a sentence or text by selecting  
8 the right word among several options (Allan, 2001). The score participants received on the  
9 test from 0 to 60 was used to determine their CEFR proficiency level from A1 to C2 (Council of  
10 Europe, 2001).

11 The target experimental items were selected from the observations of the reference  
12 corpus model based on a variety of criteria. Since we were interested in the effect of possessor  
13 animacy on the ratings specifically, we controlled for the other effects that could influence the  
14 ratings. According to the reference corpus model, the choice of genitive variant is also  
15 influenced by the definiteness and the final sibilancy of the possessor, the length of the  
16 possessor and the possessum, the lexical density of the surrounding context as well as lexical  
17 effects. To account for these effects, for the experiment we only considered observations that  
18 had definite possessors, no final sibilant, and involved lexical items that exhibited relatively  
19 minimal bias for one variant or the other.<sup>2</sup> Because the length of the constituents was difficult

---

<sup>1</sup> We argue that the TLC data is representative of speakers' language use because the GESE exam is designed to resemble an authentic discussion between the learner and the examiner across four tasks that "represent a variety of settings in which spoken language is used with different degrees of formality, interactiveness, topic familiarity, and different interlocutor roles" (Gablasova et al., 2017, p. 619; see also Dubois et al., 2022).

<sup>2</sup> This meant that we only included observations for which the possessor and possessum head lemma had an intercept adjustment in the reference corpus model that did not exceed 1 or -1.

## Animacy in the genitive alternation

1 to control for while selecting the target items, we included the constraint as a fixed effect in  
2 the linear regression model we used to analyze the rating data (see Section 2.4). In the  
3 reference corpus model, the lexical density of the surrounding context was operationalized as  
4 the type-token ratio in the 100 words surrounding the genitive observation. Including this  
5 much context would make the experiment rather long and possibly disadvantage low-  
6 proficiency learners, so the surrounding context was reduced for the experimental items. As  
7 a consequence, the predictions from the corpus-based model would no longer be completely  
8 accurate for the corresponding corpus observation. It was nonetheless important that the  
9 predictions are accurate for our second research question on the correlation between the  
10 predictions from the corpus-based model and the ratings. Therefore, we fitted a new mixed-  
11 effects logistic regression model on the corpus data without type-token ratio, which provided  
12 a new prediction for the *s*-genitive that was based only on information available in the limited  
13 context of the experimental items. This model had an  $R^2_{\text{marginal}}$  of .37, an  $R^2_{\text{conditional}}$  of .75 and  
14 a C-value of .95, indicating an excellent fit (Baayen, 2008). Like the original model, the updated  
15 reference corpus model without type-token ratio predicted the likelihood of the *s*-genitive on  
16 a probability range from 0% to 100% for the 2302 observations collected from the Trinity  
17 Lancaster Corpus (Gablasova et al., 2019).

18         Additionally, we did not consider corpus observations where the genitive variant does  
19 not match the prediction of the reference corpus model. The model might have predicted the  
20 alternative genitive variant because it did not take into account an unknown source of  
21 variation, which we did not want to influence participants' ratings. The context of the  
22 experimental item could also not contain another genitive variant so as to prevent priming  
23 effects (see Szmrecsanyi, 2006; Engel et al., 2022). Finally, the target items were simplified  
24 wherever possible to make sure that all participants understood each target item regardless

1 of their proficiency level. Specifically, we transformed subclauses into main clauses and we  
2 removed or replaced words that did not figure in the New General Service List (New-GSL,  
3 Brezina & Gablasova, 2015), which contains a stable vocabulary of 2,122 highly frequent  
4 English words that learners are most likely to encounter. Since the predictors in the reference  
5 corpus model related to the features of the possessor and possessum noun phrases,  
6 simplifying the possessor and possessum would again affect the accuracy of the predictions  
7 of the model, so we did not simplify these words and instead discarded observations whose  
8 possessor and possessum noun phrases did not figure in the New-GSL. The experimental items  
9 were selected from the remaining observations.

10 Previous research found that acceptability ratings do not correlate in a linear fashion  
11 with the predictions from regression models, especially when the model predictions indicate  
12 that the probability of a given variant is low (Arppe & Järvikivi, 2007, pp. 149–150; Divjak et  
13 al., 2016, pp. 26–27; Kempen & Harbusch, 2005, p. 330). Therefore, the correlation between  
14 the ratings and the predictions of the reference corpus model may not be as strong when the  
15 model predicts a high likelihood of the *s*-genitive compared to when the *s*-genitive is not likely  
16 to occur. To properly quantify the overall correlation between the ratings and the corpus-  
17 based predictions for our second research question, we made sure to include observations  
18 from the entire probability range. The distribution of the experimental items across animacy  
19 levels is shown in Table 1, which organizes the items by the probability of the *s*-genitive  
20 according to the reference corpus model. The number of items included in the experiment for  
21 each combination of probability and possessor animacy is given in brackets. In total, there are  
22 37 experimental items.

23

## Animacy in the genitive alternation

1 *Table 1: Distribution of target experimental items across probability bins and animacy levels. The number of observations*  
2 *used in the experiment for each combination of probability and animacy is shown in brackets.*

3

4 Probability of the s-genitive	Animacy of the possessor in the experimental items		
5 0%-20%	Temporal (3)	Animate (3)	Locative (3)
6 21%-40%	Temporal (2)	Inanimate (3)	Collective (3)
7 41%-60%	Animate (3)	Inanimate (2)	Locative (1)
8 61%-80%	Temporal (1)	Inanimate (2)	Collective (3)
9 81%-100%	Animate (3)	Collective (3)	Locative (2)

10

## 11 2.2 Participants

12 Power analysis using multiple simulations (Kumle et al., 2021) showed that to achieve a  
13 statistical power of .8, we should preferably collect data from at least 20 participants for each  
14 of the five corresponding CEFR proficiency levels featured in the TLC data, viz. B1, B2, C1, C2  
15 and native.<sup>3</sup> The participants for the study were recruited on Prolific (Prolific, 2023), an online  
16 platform where individuals can sign up to participate in paid online experiments. These  
17 individuals fill in information about themselves upon registration to Prolific, which then allows  
18 researchers to choose who is able to participate in their experiment via prescreening.

19 For the purposes of the study, it was important that the participants in our experiment  
20 resemble the speakers included in the TLC data of the reference corpus model as closely as  
21 possible. Therefore, we administered the questionnaire to 25 monolingual native speaker of  
22 British English holding British nationality, who grew up and currently live in the UK (13 female,

---

<sup>3</sup> A detailed report of the power simulation is available at  
[https://osf.io/3a6n5/?view\\_only=83930ab00f7643eba454e556b062f0be](https://osf.io/3a6n5/?view_only=83930ab00f7643eba454e556b062f0be).

## Animacy in the genitive alternation

1 12 male, age range: 19-67, M = 36, IQR = 24). The learner participants had to be raised  
2 monolingually with either Chinese, Hindi, Urdu, Spanish, Italian, Hungarian, Polish or Russian  
3 as their L1, which covers the most common L1 backgrounds featured in the TLC data. More  
4 generally, including learners from several typologically diverse L1 backgrounds increases the  
5 generalizability of the findings related to learners' proficiency level and helps ensure that the  
6 results are not influenced by transfer effects from specific L1 backgrounds.

7 Since the proficiency level of the learner participants is determined by their score on  
8 the items from the Oxford Quick Placement Test (see Section 2.1), the questionnaire was  
9 administered to learners until we reached a satisfactory number of participants for the B1, B2,  
10 C1 and C2 levels. Given that it was particularly challenging to obtain enough data from B1  
11 learners, we conflated the data from 18 B1 learners and 7 A2 learners. In total, the  
12 questionnaire was distributed to 101 EFL learners (54 female, 47 male, age range: 19-60, M =  
13 27, IQR = 15). Table 2 presents an overview of the distribution of the EFL learners by their L1  
14 and proficiency level.

15

16 *Table 2: Distribution of learner participants by mother tongue background and proficiency level*

	Chinese	Hindi/Urdu	Hungarian	Italian	Polish	Russian	Spanish	Total
A2/B1	0	0	3	3	6	3	10	25
B2	3	7	5	5	4	2	4	30
C1	10	3	1	6	3	2	3	28
C2	3	3	4	3	4	0	1	18
Total	16	13	13	17	17	7	18	101

17

1 The duration of the experiment varied considerably across participants, both among native  
2 speakers (range in minutes: 6:46-24:16, M = 13:55, IQR = 5:18) and learners (range in minutes:  
3 11:57-94:42, M = 24:44, IQR = 15:19). Nonetheless, no submissions were excluded from  
4 further analysis because all participants performed well on the more simple questions from  
5 the Oxford Quick Placement Test aimed at beginners and intermediate learners, so it is  
6 unlikely that they answered the questions randomly.

7

### 8 **2.3 Design and Procedure**

9 The survey was built in Qualtrics (Qualtrics, 2023) and administered online to participants  
10 recruited on Prolific. Before accessing the Qualtrics survey, the learner and native speaker  
11 participants on Prolific saw a slightly different description of the survey.<sup>4</sup> Since most items  
12 stemmed from a placement test (see Section 2.1), the description for the learner participants  
13 stated that we were interested in different ways of measuring how well speakers of different  
14 languages know English. Given that these items might seem particularly easy for native  
15 speakers, we explained that we also wanted to collect their data in order to test the accuracy  
16 of these measurements. This brief description also included the approximate duration of the  
17 survey and the payment participants would receive upon completion. From there, they could  
18 access the Qualtrics survey and give their consent. At the beginning of the survey, the  
19 prescreening questions were repeated. If the participants' answers were inconsistent with the  
20 information they reported on Prolific, they were automatically removed from the Qualtrics  
21 survey.

---

<sup>4</sup> The survey description and participant instructions are available at [https://osf.io/3a6n5/?view\\_only=83930ab00f7643eba454e556b062f0be](https://osf.io/3a6n5/?view_only=83930ab00f7643eba454e556b062f0be).

1           Because the experiment featured different types of questions, alternating target items  
2 and items from the Oxford Quick Placement Test could potentially have confused the  
3 participants, so questions of the same type were presented together instead. Hence, the  
4 participants answered the first 25 items from the Oxford Quick Placement Test before the  
5 target experimental items. The target items were preceded by the following instructions and  
6 an example where they could move the slider for practice:

7

8           “Below you will see sentences with two options to say the same thing, which are  
9 presented in a random order. Your task is to read each sentence carefully and choose  
10 which option sounds more natural to you. To do so, you can drag the slider towards  
11 the option that sounds more natural to you. The closer you place the slider to one end,  
12 the more natural that option seems to you, and the less natural the other option. If  
13 both seem equally natural, you can place the slider closer to the middle (you will have  
14 to use the slider to go the next question). If one option seems like the only possible  
15 option, you may move the slider all the way towards that option. Please feel free to  
16 use the full range of the slider.”

17

18 Each participant saw only one observation for each combination of probability bin and  
19 animacy level shown in Table 1. Therefore, each participant saw in total 15 experimental items  
20 out of the total of 37 experimental items. Since there were up to three observations for each  
21 combination of probability and animacy level (see Table 1), the 37 experimental items were  
22 distributed across three versions of the experiment, with each version containing a different  
23 item per combination. If there were fewer than three experimental items for a given  
24 combination, the same observations would be used in multiple versions. Within each version,



Animacy in the genitive alternation

1 we made sure that there were no more than two instances of the same possessor or  
2 possessum head noun so as to avoid lexical priming effects (Hartsuiker et al., 2008).  
3 Participants were automatically assigned to one of the three versions after receiving the  
4 instructions. The target items were presented in pseudorandom order and the position of the  
5 genitive variants on the scale was altered each time. This means that if the *s*-genitive figured  
6 on the right side of the slider as in Figure 1, the *s*-genitive would figure on the left in the  
7 following item.

8         After the 15 target experimental items, participants answered the remaining 35 items  
9 from the Oxford Quick Placement Test. At the end of the survey, participants could leave  
10 comments or feedback about the study. The participants were then automatically returned to  
11 Prolific and their survey was submitted for us to approve and remunerate the participant. To  
12 check whether the instructions were clear and whether the survey functioned without any  
13 technical issues, the survey was first piloted on five learner and six native speaker participants,  
14 who were made aware that they would participate in a pilot study before they started.

15

## 16 **2.4 Statistical analysis**

17 The ratings were analyzed as the response variable in two mixed-effects linear regression  
18 models using the *lme4* package (Bates et al., 2015) in R (version 4.2.2, R Core Team 2022).<sup>5</sup>  
19 Specifically, we employed treatment coding to investigate the effect of the predictors on the  
20 ratings. In accordance with the reference corpus model, native speakers are chosen as the  
21 reference proficiency level, which allows us to compare their intuitions with those of learners  
22 at different stages of language learning. To account for idiosyncrasies and irrelevant sources

---

<sup>5</sup> The dataset and the analysis are available at  
[https://osf.io/3a6n5/?view\\_only=83930ab00f7643eba454e556b062f0be](https://osf.io/3a6n5/?view_only=83930ab00f7643eba454e556b062f0be).

## Animacy in the genitive alternation

1 of variation in the rating data, both models featured intercept adjustments for the participant  
2 and the experimental item (Baayen et al., 2008; Gries, 2015). If the *s*-genitive was presented  
3 on the left side of the slider, the participants' ratings corresponded to a score from 0 to 100  
4 for the *of*-genitive. Because the reference corpus model predicted the likelihood of the *s*-  
5 genitive (Dubois et al., 2022), these ratings were transformed to also reflect a score for the *s*-  
6 genitive instead before analyzing the data.

7         The first model was concerned with our first research question and tested whether  
8 and how the five animacy levels influenced the ratings from the native speakers compared to  
9 the learners at different proficiency levels (Dubois et al., 2022). Hence, the first model  
10 included an interaction between the proficiency level of the participants and the animacy of  
11 the possessors in the experimental items. Collective possessors figure as the reference  
12 animacy level because this makes it easier to test the hypothesis in Dubois et al., (2022) that  
13 learners primarily struggle with collective and locative possessors. This model also featured  
14 the length of the possessor and the possessum as predictors because it was not possible to  
15 control for these constraints while designing the experiment (see Section 2.1). The length of  
16 the constituents was standardized by subtracting the median length and dividing the resulting  
17 centered value by two standard deviations (Gelman, 2008).

18         The second model dealt with the second research question concerning the overall  
19 relation between the patterns of language use of the speakers in the corpus and the intuitions  
20 of participants by investigating whether and how the ratings from the native speakers and EFL  
21 learners at different proficiency levels correlated with the corpus-based predictions from the  
22 reference corpus model for the experimental items under study. Specifically, this model  
23 featured the probability of the *s*-genitive according to the reference corpus model as a  
24 predictor, which was allowed to interact with the proficiency level of the participants to reveal

1 whether the strength of the correlation varies across participants from different proficiency  
2 levels. Importantly, the probability of the *s*-genitive was calculated based on the effect of all  
3 predictors present in the reference corpus model based on data from both native speakers  
4 and learners. This means that the prediction also incorporates the effect of the significant  
5 interaction between proficiency level and possessor animacy in the reference corpus model.  
6 Hence, if the experimental item was produced by a B1 learner in the original corpus data for  
7 example, the corpus-based prediction would capture the behavior of B1 learners specifically.  
8 Since B1 and B2 learners are less sensitive to the animacy of the possessor than native  
9 speakers according to the reference corpus model, the corpus-based prediction for an  
10 experimental item produced by a B1 or B2 learner might not be entirely accurate for a native  
11 speaker participant and vice versa. While this is admittedly not ideal, the experimental items  
12 were selected more or less evenly from all proficiency levels in the original corpus data, so all  
13 participants had to rate items produced by speakers with a different proficiency level.  
14 Therefore, we maintain that the correlation between the predictions of the reference corpus  
15 model and the ratings from the participants provides an accurate representation of the  
16 relation between the language use and intuitions of speakers at different proficiency levels.  
17 As for the first model on the effect of animacy, this is not an issue considering that we are  
18 mostly interested in whether the ratings for the *s*-genitive are higher or lower depending on  
19 the proficiency level of the participant.

20

### 21 **3. Results**

#### 22 **3.1 The animacy model**

Animacy in the genitive alternation

1 Table 3 presents the results from the first model on the effect of possessor animacy on the  
 2 ratings. This model has an  $R^2_{\text{marginal}}$  of .15 and  $R^2_{\text{conditional}}$  of .35.

3

4 *Table 3: Effects of the individual predictors in the animacy model with collective possessors as reference level. Positive*  
 5 *coefficients indicate an increase in ratings for the s-genitive. The reference level of the categorical predictors is given in*  
 6 *brackets. (Marginally) significant p-values are written in bold.*

Predictor	Coefficient	Std. error	t-value	p-value
(Intercept)	63.503	6.27	10.128	< <b>0.001</b>
Possessor length	-1.214	5.287	-0.230	0.82
Possessum length	-15.878	5.212	-3.046	<b>0.005</b>
Animacy (collective)				
Animate	16.916	8.391	2.016	<b>0.049</b>
Inanimate	-16.456	8.889	-1.851	<b>0.07</b>
Locative	-17.109	9.378	-1.824	<b>0.074</b>
Temporal	-20.857	8.986	-2.321	<b>0.024</b>
Proficiency (native)				
A2 & B1	1.959	5.363	0.365	0.715
B2	-11.019	5.146	-2.141	<b>0.033</b>
C1	-16.116	5.218	-3.089	<b>0.002</b>
C2	-11.402	5.929	-1.923	<b>0.055</b>

Animacy in the genitive alternation

Proficiency x animacy (collective)				
A2 & B1 + animate	-14.191	6.814	-2.083	<b>0.037</b>
A2 & B1 + inanimate	-6.64	6.813	-0.975	0.33
A2 & B1 + locative	-5.071	6.82	-0.744	0.457
A2 & B1 + temporal	-7.592	6.82	-1.113	0.266
B2 + animate	4.291	6.538	0.656	0.512
B2 + inanimate	13.391	6.535	2.049	<b>0.041</b>
B2 + locative	5.489	6.534	0.840	0.401
B2 + temporal	0.934	6.534	0.143	0.886
C1 + animate	9.226	6.629	1.392	0.164
C1 + inanimate	8.113	6.627	1.224	0.221
C1 + locative	7.001	6.63	1.056	0.291
C1 + temporal	8.148	6.630	1.229	0.219
C2 + animate	5.105	7.539	0.677	0.498
C2 + inanimate	3.288	7.526	0.437	0.662
C2 + locative	7.905	7.508	1.053	0.293
C2 + temporal	4.496	7.508	0.599	0.549

1

2 Whereas the length of the possessor does not influence the ratings, the model shows that the  
3 ratings for the *s*-genitive decrease when the length of the possessum increases, which echoes  
4 the finding that native speakers and learners alike tend to prefer the *of*-genitive when the  
5 possessum is long in spoken language production (Dubois et al., 2022), in conflict with the  
6 short-before-long principle (Behaghel, 1909). Since animacy is part of an interaction with

## Animacy in the genitive alternation

1 proficiency level, the main effects for animacy only pertain to the reference proficiency level,  
2 namely native speakers. The main effects for animacy show that native speakers give higher  
3 ratings for the *s*-genitive when the possessor is animate rather than collective, which in turn  
4 receive significantly higher ratings than temporal possessors. This pattern aligns with previous  
5 research stating that the *s*-genitive is preferred with referents that are higher on the animacy  
6 scale (Rosenbach, 2014, p. 232). The main effects for proficiency level compares the ratings  
7 for collective possessors across proficiency levels. The effects indicate that B2 and C1 learners  
8 give significantly lower ratings for the *s*-genitive than native speakers when the possessor is  
9 collective. The model features two significant interaction effects. While native speakers give  
10 higher ratings for the *s*-genitive when the possessor is animate rather than collective, this  
11 effect is much weaker for A2 and B1 learners, who give similar ratings for animate and  
12 collective possessors. B2 learners, by contrast, do not distinguish between inanimate and  
13 collective possessors as much as native speakers.<sup>6</sup>

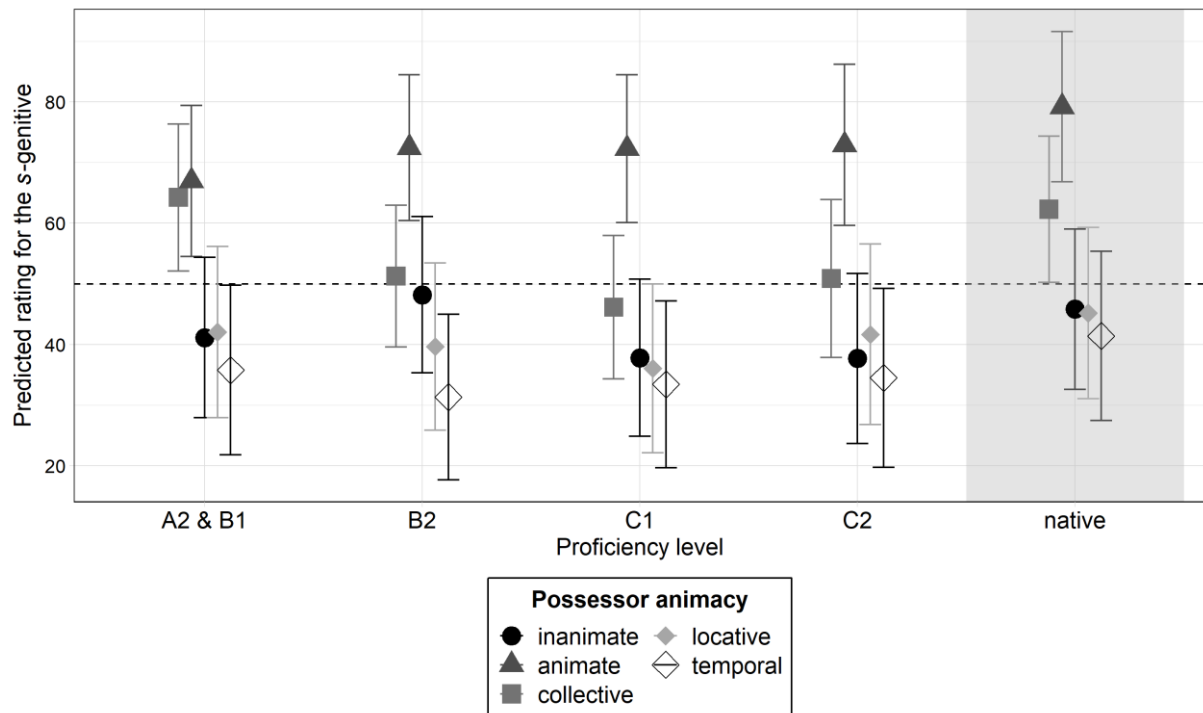
14 The effects pertaining to possessor animacy are visualized in the partial effects plot in  
15 Figure 2. Partial effects plots summarize the information from the model by changing the value  
16 of the relevant constraint, viz., possessor animacy, for each of the proficiency levels while  
17 holding the other constraints in the model, namely possessor and possessum length, at their  
18 default level (Fox, 2003). In this way, the plot shows how a change in possessor animacy affects  
19 the ratings for the *s*-genitive for native speakers and learners across proficiency levels.

20

---

<sup>6</sup> To ascertain that these interaction effects are not due to Spanish EFL learners, who account for about 40% of the A2 and B1 learners, we ran a model without these learners, which revealed the same patterns.

## Animacy in the genitive alternation



1

2 *Figure 2: Partial effects plot of the interaction between possessor animacy and proficiency level. Predicted ratings*  
 3 *on the y-axis are for the s-genitive. The vertical distance between the shapes reflects the effect size of the predictor.*  
 4 *Errors bars represents confidence intervals (95%). The plotted probabilities are calculated with the standardized*  
 5 *length of the possessor and possessum at their default level, viz. 0. The effects pertaining to the native speakers*  
 6 *are shaded in grey.*

7

8 As suggested by the interaction effects presented in Table 3, Figure 2 shows that the ratings  
 9 for animate possessors and collective possessors are very similar at the A2/B1 level whereas  
 10 native speakers clearly give higher ratings for the s-genitive when the possessor is animate  
 11 rather than collective. Specifically, it seems that A2 and B1 learners give slightly higher ratings  
 12 to collective possessors while at the same time giving lower ratings to animate possessors.  
 13 Turning to the interaction effect between B2 learners and native speakers regarding collective  
 14 and inanimate possessors, we can see that B2 learners give significantly lower ratings for the  
 15 s-genitive than native speakers when the possessor is collective (see Table 3) while at the same  
 16 time giving higher ratings when the possessor is inanimate. As a result, there is barely a  
 17 difference between the ratings for collective possessors and inanimate possessors at the B2

1 level as opposed to native speakers who give higher ratings for the *s*-genitive when the  
 2 possessor is collective.

3

### 4 **3.2 The corpus prediction model**

5 Table 4 presents the results from the second model. The model has an  $R^2_{\text{marginal}}$  of .14 and a  
 6  $R^2_{\text{conditional}}$  of .34.

7

8 *Table 4: Effects of the individual predictors in the corpus prediction model. Positive coefficients indicate an increase in*  
 9 *ratings for the s-genitive. The reference level of the categorical predictors is given in brackets. (Marginally) significant p-*  
 10 *values are written in bold.*

Predictor	Coefficient	Std. error	t-value	p-value
(Intercept)	31.044	5.313	5.843	<b>&lt;.001</b>
Corpus-based prediction	53.313	8.987	5.932	<b>&lt;.001</b>
Proficiency (native)				
A2 & B1	10.063	4.732	2.127	<b>0.034</b>
B2	0.361	4.539	0.080	0.937
C1	-6.501	4.594	-1.415	0.158
C2	-5.209	5.221	-0.998	0.319
Proficiency x corpus-based prediction				
A2 & B1 + corpus-based prediction	-30.717	7.246	-4.239	<b>&lt;.001</b>



## Animacy in the genitive alternation

B2 + corpus-based prediction	-13.570	6.933	-1.957	<b>0.051</b>
C1 + corpus-based prediction	-6.442	7.007	-0.919	0.358
C2 + corpus-based prediction	-4.255	7.944	-0.536	0.592

1

2

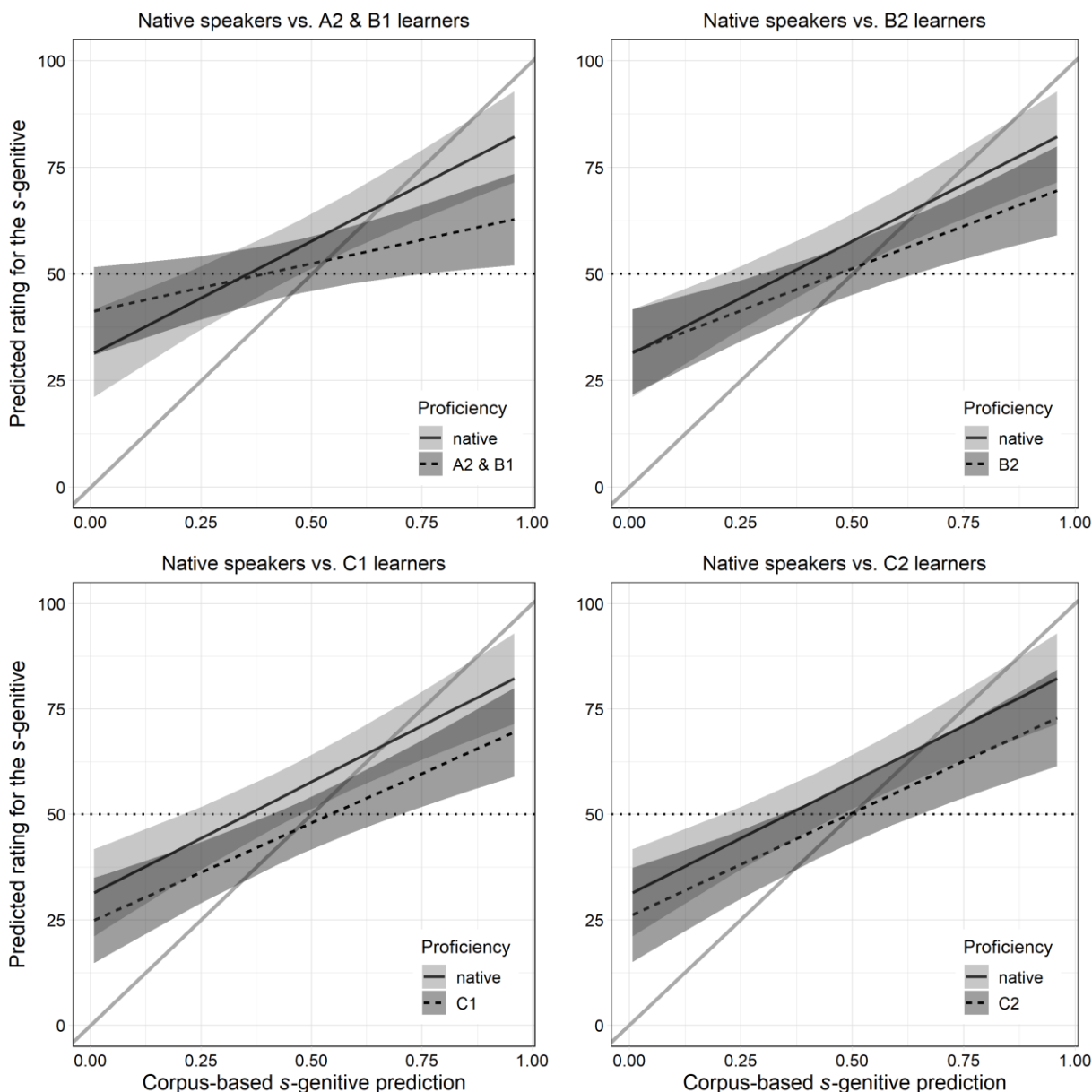
There is a significant effect for the prediction of the reference corpus model which shows that when the model predicts a higher likelihood for the *s*-genitive, the rating for the *s*-genitive also increases. Since the corpus-based prediction is part of a significant interaction term with proficiency level, the main effect shown in Table 4 only pertains to the reference proficiency level, namely native speakers. The interaction suggests that A2 and B1 speakers are less sensitive to the effect of the corpus-based predictions compared to native speakers, indicating that their ratings correspond less closely to the predictions of the reference corpus model. As their proficiency level increases, the ratings of the learners become more similar to those of the native speakers, and by extension, to the language use of the speakers in the corpus. Hence, while the interaction is still marginally significant for B2 learners, the coefficients gradually approach zero in the C1 and C2 levels.

13

The partial effects plot in Figure 3 visualizes the correlation between the ratings and the corpus-based predictions for native speakers and each learner proficiency level separately.

15

## Animacy in the genitive alternation



1  
2 *Figure 3: Partial effects plot of the interaction between proficiency level and the corpus-based prediction for the*  
3 *s-genitive for native speakers and each learner proficiency level separately. Predicted ratings on the y-axis are for*  
4 *the s-genitive. The grey diagonal line represents a perfect fit between corpus predictions and ratings.*

5  
6 The figure shows that the ratings from the native speakers are most sensitive to the  
7 predictions of the reference corpus model: their slope is steeper compared to the slopes for  
8 the learners, indicating that an increase in the predicted likelihood of the s-genitive leads to a  
9 larger increase in ratings for the s-genitive. Specifically, the intercept in Table 4 indicates that

1 native speakers give the *s*-genitive a low rating of 31 when the reference corpus model  
2 predicts the *of*-genitive, whereas they rate the *s*-genitive approximately 50 points higher when  
3 the *s*-genitive is very likely in language use. Overall, the slopes for the learners become steeper  
4 as their proficiency level increases, so they gradually approximate native speakers. The upper  
5 left panel of Figure 3 shows that A2 and B1 learners, who differ significantly from native  
6 speakers, exhibit the flattest curve. Specifically, the curve for the A2 and B1 learners seems to  
7 indicate that A2 and B1 learners give lower ratings for the *s*-genitive than native speakers  
8 when the *s*-genitive is highly likely according to the reference corpus model while also giving  
9 higher ratings than native speakers when the *s*-genitive is unlikely, i.e., when the reference  
10 corpus model predicts a preference for the *of*-genitive (see Table 4). B2 learners also rate the  
11 *s*-genitive slightly lower than native speakers when the reference corpus model predicts an *s*-  
12 genitive, but they give the same ratings as native speakers when the model predicts an *of*-  
13 genitive. Interestingly, it seems that C1 and C2 learners give slightly lower ratings for the *s*-  
14 genitive in general compared to native speakers, which might point to a slight overall  
15 preference for the *of*-genitive.<sup>7</sup>

16 The overall repeated measures correlation between the predictions from the  
17 reference corpus model and the ratings from all participants is moderate at  $r = 0.38$  ( $p < 0.001$ )  
18 (Bakdash & Marusich, 2017). If we consider that ratings above 50 represent a preference for  
19 the *s*-genitive, we can also examine how often the participants' intuitions and the predictions  
20 from the reference corpus model agree on the same genitive variant in a categorical way.

---

<sup>7</sup> One could argue that the slightly lower slopes of the C1 and C2 learners align better with the corpus-based predictions as they more closely follow the grey diagonal representing a perfect fit. If learners also exhibit a slight preference for the *of*-genitive compared to native speakers in their language use, then this finding is not surprising considering that the corpus-based predictions stem from the language use of both native speakers and learners.

1 Hence, for each of the participant proficiency levels, we selected a small, balanced sample  
2 consisting of 100 observations for which the reference corpus model predicts that the *s*-  
3 genitive is more likely, and 100 for which it predicts that the *of*-genitive is more likely. The  
4 intuitions of the native speakers correspond to the predicted variant in 80% of the cases. By  
5 comparison, the intuitions from participants at the A2 and B1 level match the predicted  
6 variant in only 68% of the cases. As learners become more proficient, their intuitions become  
7 more similar to actual language use, with an agreement of 73% at the B2 level, 77% at the C1  
8 level and 81% at the C2 level.

9

## 10 **4. Discussion**

### 11 **4.1 (Methodological) summary**

12 The present study investigated the interface between implicit grammatical knowledge and  
13 language. To do so, we compared the intuitions of native speakers and EFL learners regarding  
14 the naturalness of the genitive variants in a rating task experiment against the use of the  
15 genitive variants in language production, for which we took as a reference point the corpus-  
16 based regression model of Dubois et al. (2022). The ratings of 25 native speakers and 101 EFL  
17 learners from various L1 backgrounds with a proficiency level ranging from A2 to C2 were  
18 analyzed with two separate mixed-effects linear regression models. The first model focused  
19 on our first research question and compared how possessor animacy, operationalized as a  
20 five-level distinction between animate, collective, inanimate, locative and temporal  
21 possessors, influences the intuitions of native speakers and EFL learners in the rating task  
22 experiment compared to the language use of the speakers in the corpus. The second model  
23 was aimed at our second research question, which investigates how the intuitions from the  
24 participants in the experiment relate to the patterns of language use in general. Therefore,

1 the second model tested how the predictions from the reference corpus model, which reflect  
2 spoken language use, correlate with the ratings of the participants in the experiment as this  
3 would indicate that native speakers and EFL learners are generally sensitive to probabilistic  
4 constraints driving the choice of variant (Bresnan & Ford, 2010).

5

6 **4.2 RQ1: How does possessor animacy affect the intuitions of EFL learners at**  
7 **different proficiency levels in a rating task experiment compared to**  
8 **correlations between animacy and genitive variant in a corpus of spoken**  
9 **learner language?**

10 In accordance with previous research (Rosenbach, 2014, p. 232), the results from the first  
11 model show that native speakers find the *s*-genitive most natural when the possessor is  
12 animate, followed by collective, inanimate, locative and temporal possessors. By contrast, A2  
13 and B1 learners find the *s*-genitive more or less equally natural with animate and collective  
14 possessors. Figure 2 suggests that this is due to A2 and B1 learners giving slightly higher ratings  
15 for collective possessors while at the same time giving lower ratings for animate possessors.  
16 B2 learners, on the other hand, find the *s*-genitive equally natural when the possessor is  
17 collective or inanimate, whereas native speakers clearly prefer the *s*-genitive when the  
18 possessor is collective. Figure 2 suggests that this is again due to a combination of B2 learners  
19 giving slightly lower ratings for collective possessors (see Table 3) and higher ratings for  
20 inanimate possessors.

21 At first glance, the lack of distinctions between certain animacy levels in the ratings of  
22 A2/B1 and B2 participants in the experiment reflect the tendency for B1 and B2 learners to  
23 rely less on the animacy constraint than native speakers in spoken language production

1 (Dubois et al., 2022). There are, however, important differences. According to Dubois et al.  
2 (2022), B1 and B2 learners are less sensitive to animacy because they have difficulty  
3 determining the animacy of possessors like *region, province, world, India* etc. in specific  
4 contexts where these possessors can be interpreted as either locatives, which take the *of-*  
5 genitive, or collectives, which prefer the *s-*genitive. As a result, there is a trade-off between  
6 these two categories, causing learners to use the *s-*genitive more often with locative  
7 possessors and the *of-*genitive more often with collective possessors. Although the  
8 experiment certainly suggests that B2 learners find the *s-*genitive less natural than native  
9 speakers when the possessor is collective up until C1 level (see Table 3), this does not apply  
10 to the A2 and B1 learners. The participants also do not appear to experience difficulties with  
11 locative possessors and instead A2/B1 and B2 learners struggle with animate and inanimate  
12 possessors respectively. This is surprising considering that the easiest strategy for learners  
13 would be to rely on the very strong animacy constraint in a categorical way by always using  
14 the *s-*genitive when the possessor is animate and the *of-*genitive when the possessor is  
15 inanimate, as suggested by prescriptivists (Murphy, 2012; cited in Heller et al., 2017, p. 10).

16 Moreover, while a trade-off between locative and collective possessors seems  
17 plausible, it is unclear how learners would fail to differentiate between animate and collective  
18 possessors or collective and inanimate possessors, unless there are important distinctions  
19 beyond our five-level annotation of animacy that we did not take into account. In general,  
20 animacy is difficult to operationalize in all its facets, since the linguistic animacy distinctions  
21 are language-specific rather than biological and not diachronically stable (Wolk et al., 2013;  
22 Zaenen et al., 2004). Rather than a trade-off, the more likely explanation is that learners  
23 struggle with each of the categories independently. According to Dubois et al. (2022), the  
24 animacy constraint requires learners to first determine the animacy of the possessor noun

## Animacy in the genitive alternation

1 phrase in the context. Once this has been determined, they can produce the preferred genitive  
2 variant for this animacy category. For a possessor like *dog* for example, learners would first  
3 have to determine that it is an animate possessor. Since animate possessors prefer the *s*-  
4 genitive, they would be more likely to produce an *s*-genitive with *dog* as a possessor. With this  
5 in mind, learners' struggle with independent animacy categories could either result from the  
6 fact that the learners have not yet fine-tuned how these larger animacy categories affect the  
7 choice of variant, e.g., collectives prefer the *s*-genitive, but not as much as animate possessors,  
8 or they simply do not have enough experience with certain individual possessor noun phrases  
9 to determine their animacy in a consistent way to begin with. The latter explanation seems  
10 more plausible since what makes the animacy constraint more complicated than other  
11 constraints is the multitude of nouns that have to be categorized correctly based on the  
12 context without overt marking (see Dubois et al., 2022).

13 In fact, low-proficiency learners might not even consider the animacy of the possessor  
14 when choosing the genitive variant because learners do not start out with knowledge of these  
15 categories and their associations with the variants. Instead, they are exposed to genitive  
16 exemplars with thousands of different nouns, each of which has its own preferred genitive  
17 variant. Once they learn how many of these nouns pattern with the genitive variants, they  
18 may recognize that nouns behaving the same way share the same animacy, which then  
19 gradually influences language production (see Abbot-Smith & Tomasello, 2006, p. 284; Azazil,  
20 2020, p. 420; Divjak et al., 2021, p. 78).

21 In such an account, the animacy constraint would capture item-specific effects in the  
22 early stages of acquisition, so the most plausible explanation for the differences between  
23 native speakers and low-proficiency learners regarding the animate, inanimate and collective  
24 possessors is that the nouns included in the experiment for these categories are somehow

1 problematic. According to usage-based approaches, language learning is driven by input (Ellis  
2 & Wulff, 2019), so the frequency of possessors nouns can be expected to influence their  
3 learnability considerably. To compare the frequency of the possessor nouns across animacy  
4 levels, we checked their ranking in the New-GSL (Brezina & Gablasova, 2015). The expectation  
5 is that the animate, collective and inanimate possessors feature less frequent nouns on  
6 average than locative and temporal possessors, so they should be found lower on the list. The  
7 average ranking of animate (593), collective (688) and inanimate (970) possessors in the list is  
8 indeed on lower on average than temporal possessors (289). Most locative possessors are  
9 country names, which are not included in the New-GSL, but it is highly likely that learners are  
10 also very familiar with words like America and India, so overall it seems that temporal and  
11 locative possessors were represented by easier nouns in the experiment. Another difficulty  
12 with collective possessors in the experiment is that they mostly feature singular collective  
13 nouns with plural reference like *association*, *community* and *government*, which are  
14 potentially confusing for learners (e.g., Dziemianko, 2008). Nonetheless, one would need to  
15 adopt a more sophisticated approach to collect more conclusive evidence considering that the  
16 overall frequency of a noun does not correlate perfectly with how frequently it is encountered  
17 as a possessor in genitive exemplars.

18

#### 19 **4.3 RQ2: How does the relationship between experimental ratings and corpus** 20 **data vary by proficiency level of participants and speakers?**

21 The second model showed that there is a positive correlation between the predictions from  
22 the reference corpus model and the ratings in the experiment: when the reference corpus  
23 model predicts a high likelihood for the *s*-genitive, the participants in the experiment also give



1 higher ratings for the *s*-genitive. These results suggest that the significant constraints on which  
2 the reference corpus model is based (see Dubois et al., 2022), i.e., the constraints that drive  
3 spoken language use, are indeed part of speakers' implicit linguistic knowledge as posited by  
4 the Probabilistic Grammar framework (Bresnan & Ford, 2010; Engel et al., 2022). However,  
5 the correlation between the predictions from the reference corpus model and the ratings is  
6 weaker for the A2/B1 learners and to a lesser extent the B2 learners compared to native  
7 speakers.

8         This difference could be explained in two ways. On the one hand, the linear regression  
9 model might have inferred that the ratings from the A2/B1 and B2 learners correspond less  
10 well to the predictions from the reference corpus model because these learners tended to  
11 stick to the middle range of the rating scale, as implied by their flatter curve in Figure 3. One  
12 reason for this might be for example that A2/B1 and B2 learners are less confident about their  
13 choices than native speakers. In this case, their intuitions about the naturalness of the variants  
14 would be in principle be similar to those of native speakers, but they did not want to give  
15 ratings on the extreme ends of the slider given that a high rating for one variant implies that  
16 the other variant is much less natural in our experiment design. In other words, although  
17 learners might have a clear preference for one variant that aligns with the intuition of native  
18 speakers, they might not have wanted to exclude the other variant completely. To test this  
19 possibility, we compared to what extent the ratings of A2/B1, B2 and native speakers deviated  
20 from the middle point of the rating scale, namely 50. For example, if a learner gave a rating of  
21 75 for the *s*-genitive, the deviation from 50 is 25. If A2/B1 and B2 learners stick more to the  
22 middle range of the rating scale, their deviations from 50 should on average be smaller than  
23 those of the native speakers. Wilcoxon tests revealed that neither A2/B1 learners ( $p = 0.99$ )  
24 nor B2 learners ( $p = 0.34$ ) gave ratings that are significantly closer to the middle point of the

1 scale than native speakers. Hence, the interaction effect is probably not due to these learners  
2 sticking to the middle range of the rating scale.

3 This supports the second possible explanation, which is that the ratings from A2/B1  
4 and B2 learners correspond less well to the predictions of the reference corpus model  
5 compared to native speakers because these learners do not always share the same intuitions  
6 as native speakers. More specifically, the intuitions from the native speakers in the experiment  
7 appear to correspond more closely to how native speakers and learners use the genitive  
8 variants in spoken language production. The question arises as to whether this difference in  
9 intuitions appears in specific contexts. In this respect, the partial effects plot in Figure 3 shows  
10 that the intuitions of A2 and B1 learners diverge from those of native speakers both in contexts  
11 where speakers would use the *s*-genitive and in contexts where they would use the *of*-genitive  
12 in language production. In *s*-genitive contexts, A2 and B1 learners find the *of*-genitive more  
13 natural than native speakers in the experiment. In *of*-genitive contexts, A2 and B1 learners  
14 find the *s*-genitive more natural than native speakers. B2 learners share the same intuitions  
15 as native speakers in *of*-genitive contexts, but they differ from native speakers in *s*-genitive  
16 contexts, where they find the *of*-genitive slightly more natural.

17

#### 18 **4.4 Synopsis**

19 While there are subtle differences in the way in which the animacy constraint affects the  
20 intuitions and spoken language production of A2/B1 and B2 learners, both sources of data  
21 point to the fact that low-proficiency learners are generally less sensitive than native speakers  
22 to certain animacy levels (see Dubois et al., 2022). In general, the results from our study are  
23 reminiscent of the findings of Wolk et al. (2013) that the effect of animacy is weakening over  
24 time in different alternation phenomena and more importantly Heller et al. (2017), who found

## Animacy in the genitive alternation

1 that the effect of animacy on the choice of genitive variant is also weaker for speakers from  
2 different World Englishes compared to native varieties. In line with Dubois et al. (2022), we  
3 argue that there are differences between learners and native speakers with regard to  
4 possessor animacy because the constraint requires learners to pick up on the statistical  
5 association between the genitive variants and many different possessor nouns. As a result,  
6 low-proficiency learners lack the necessary language exposure to develop abstract animacy  
7 categories they can rely on during language production like native speakers.

8         However, it is important to keep in mind that the models present the aggregated  
9 results across the individual participants. It is therefore unclear whether the results from the  
10 model derive from the A2/B1 and B2 learners as a group or by some individuals in particular  
11 who developed different strategies from the other learners at the same proficiency level. This  
12 is all the more relevant in the context of learner language since it features even more  
13 individual variability compared to native speakers (Jäschke & Plag, 2016, pp. 511–512; Kinne,  
14 2020, p. 128). Inspired by Kinne (2020), we therefore calculated Cronbach's alpha on the 15  
15 ratings of each participant to get an idea of how much individual variability there is at each  
16 proficiency level. As expected, the native speakers and the C2 learners are the most consistent  
17 in their ratings with a Cronbach's  $\alpha$  of .95, followed by the C1 learners (Cronbach's  $\alpha$  = .94)  
18 and the B2 learners (Cronbach's  $\alpha$  = .9), all of whom exhibit excellent internal consistency  
19 above .9 (Gliem & Gliem, 2003). By comparison, the A2 and B1 learners are much less  
20 consistent with a Cronbach's  $\alpha$  of .72. This implies that the A2 and B1 learners are indeed not  
21 very homogenous with respect to how natural they find the genitive variants with different  
22 possessors in various contexts.

23         A closer look at the ratings from individual A2/B1 participants showed that 10 out of  
24 25 participants (40%) already follow the expected pattern whereby s-genitives are generally

1 considered more natural with animate possessors than with collective possessors. In the B2  
2 group, a similar pattern emerges: half the participants (15 out of 30) find the *s*-genitive more  
3 natural with collective possessors than inanimate possessors like native speakers. Since  
4 learners build an interlanguage grammar that accommodates the language input they have  
5 received in a rational way (Ellis, 2006), the different strategies of learners ultimately depend  
6 on the exemplars they have encountered in their previous language experience (Ellis & Wulff,  
7 2019). Because low-proficiency learners have less experience with the language, their  
8 interlanguage grammar is built on fewer exemplars that can be categorized in different ways,  
9 which naturally leads to more variability across different learners.

10 More generally, we found that the intuitions from the participants correlate with the  
11 patterns of spoken language use, which provides further evidence that native speakers and  
12 EFL learners alike are sensitive to the implicit probabilistic constraints that drive language use,  
13 as posited by the Probabilistic Grammar framework (Bresnan & Ford, 2010; Engel et al., 2022).  
14 However, this correlation is slightly weaker for the A2/B1 and B2 learners, which implies that  
15 these learners are more likely than native speakers to produce a genitive variant that they do  
16 not find most natural in a given context. As learners become more proficient, their intuitions  
17 and language use become more similar.

18 To some extent, this result can be explained by the fact that low-proficiency  
19 participants are less consistent in their ratings, which could attenuate the relevant statistical  
20 association in the model, and by extension, the regression slope in Figure 3.<sup>8</sup> This is also  
21 partially due to the effect of the animacy constraint. In accord with previous literature, the  
22 reference corpus model found that the animacy of the possessor is one of the strongest  
23 determinants of genitive choice in spoken language production (Rosenbach, 2014). Hence, the

---

<sup>8</sup> We thank an anonymous reviewer for this observation.

1 prediction of the reference corpus model was largely determined by the effects of possessor  
2 animacy, with animate and collective possessors strongly favoring the *s*-genitive and  
3 inanimate, locative and temporal possessors disfavoring the *s*-genitive (Dubois et al., 2022).  
4 The fact that the effect of the animacy constraint influences the language use and the  
5 intuitions of the A2/B1 and B2 learners slightly differently contributes to why the intuitions of  
6 these learners correspond less well to the predictions of the reference corpus model.

7         However, the effect of the animacy constraint does not overlap with the findings from  
8 the corpus-prediction model well enough to account for this finding in its entirety. Another  
9 explanation is that spoken language involves different kinds of time-sensitive, processing-  
10 related factors that influence language production and might cause a speaker to produce a  
11 variant that they find less natural but that is easier to process in a given context. These  
12 restrictions did not apply during the experiment, so the participants could always choose the  
13 variant they found more natural. This could lead to differences between the intuitions of the  
14 participants and their spoken language use, especially for low-proficiency learners for whom  
15 the processing-related considerations during spoken language production are arguably more  
16 important given that their language production is in all likelihood less efficient and automatic  
17 than that of more advanced learners and native speakers (Segalowitz, 2003). In reality  
18 however, most processing-related constraints are controlled for during the experiment and  
19 previous research showed that even low-intermediate learners are surprisingly similar to  
20 native speakers when it comes to the actual use of the genitive variants in spoken language  
21 production (Dubois et al., 2022; see also Gries & Wulff, 2013).

22         Therefore, we cannot exclude the alternative possibility that low-proficiency learners  
23 differ from native speakers because of factors that are specific to the rating task experiment.  
24 In this regard, the intuitions feature much unexplained variability that does not correspond to

1 the patterns of spoken language use, especially at the lower-proficiency levels. On the one  
2 hand, this implies that some of the constraints driving language use are not relevant during  
3 introspective tasks like the rating task experiment. Unfortunately, it is not possible to test the  
4 relevance of each individual constraint since the corpus-based prediction results from the  
5 combined effect of all significant constraints, not to mention most constraints are controlled  
6 for apart from possessor animacy. On the other hand, this divergence between corpus-based  
7 predictions and ratings is to be expected given that introspection yields slightly different  
8 results than corpus data (Arppe & Järvikivi, 2007, p. 152).

9 More generally, rating task experiments offer the participant more freedom than  
10 forced choice tasks for example, so learners might have a different interpretation of what  
11 sounds natural to them as they complete the experiment. On a related note, the ratings of  
12 especially low-proficiency learners could be influenced by prescriptivism and their previous  
13 teaching experience, meaning that they could give higher ratings for the variant they were  
14 taught is 'right' although this does not align with their implicit grammatical knowledge. Overall,  
15 it is not entirely clear to what extent the ratings capture implicit grammatical knowledge as  
16 opposed to strategic behavior, or whether it is natural for participants to gauge the  
17 acceptability of the variants on such a fine-grained scale as in the present design to begin with  
18 (see Stefanowitsch, 2006, p. 73). Future research should therefore investigate whether there  
19 are differences between what these ratings capture at low-proficiency levels compared to  
20 higher proficiency levels and native speakers. Nonetheless, it is reassuring that the strength  
21 of the correlation between corpus-based predictions and ratings for native speakers and  
22 learners across several proficiency levels and L1 backgrounds is comparable to recent studies  
23 that adopt the same approach on native speakers of English exclusively (see Engel, 2022, p.  
24 176; Engel & Szmrecsanyi, 2023, p. 370).

1

## 2 **Data Availability Statement**

3 The data, the code with the analyses, the survey description and participant instructions are  
4 available at [https://osf.io/3a6n5/?view\\_only=83930ab00f7643eba454e556b062f0be](https://osf.io/3a6n5/?view_only=83930ab00f7643eba454e556b062f0be).

5

## 6 **References**

- 7 Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based  
8 account of syntactic acquisition. *The Linguistic Review*, 23(3), 275–290.
- 9 Allan, D. (2001). *Oxford Placement Test 1*. Oxford University Press.
- 10 Arppe, A., & Järvikivi, J. (2007). Every method counts: Combining corpus-based and experimental  
11 evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory*, 3(2), 131–159.
- 12 Azazil, L. (2020). Frequency effects in the L2 acquisition of the catenative verb construction –  
13 evidence from experimental and corpus data. *Cognitive Linguistics*, 31(3), 417–451.
- 14 Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*.  
15 Cambridge University Press.
- 16 Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random  
17 effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- 18 Bakdash, J. Z., & Marusich, L. R. (2017). Repeated Measures Correlation. *Frontiers in Psychology*, 8,  
19 456.
- 20 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4.  
21 *Journal of Statistical Software*, 67(1), 1–48.
- 22 Behaghel, O. (1909). Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern.  
23 *Indogermanische Forschungen*, 25, 110–142.
- 24 Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to  
25 sentence formulation. *Psychological Review*, 89(1), 1–47.

## Animacy in the genitive alternation

- 1 Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical  
2 function assignment and word order during production. *Lingua*, *118*(2), 172–189.
- 3 Bresnan, J., & Ford, M. (2010). Predicting syntax: Processing dative constructions in American and  
4 Australian varieties of English. *Language: Journal of the Linguistic Society of America*, *86*(1),  
5 168–213.
- 6 Brezina, V., & Gablasova, D. (2015). Is There a Core General Vocabulary? Introducing the New  
7 General Service List. *Applied Linguistics*, *36*(1), 1–22.
- 8 Council of Europe. (2001). *Common European Framework of Reference for languages: Learning,*  
9 *teaching, assessment*. Cambridge University Press.
- 10 Divjak, D., & Arppe, A. (2013). Extracting prototypes from exemplars: What can corpus data tell us  
11 about concept representation? *Cognitive Linguistics*, *24*(2), 221–274.
- 12 Divjak, D., Dąbrowska, E., & Arppe, A. (2016). Machine Meets Man: Evaluating the Psychological  
13 Reality of Corpus-based Probabilistic Models. *Cognitive Linguistics*, *27*(1), 1–33.
- 14 Divjak, D., Milin, P., Ez-zizi, A., Józefowski, J., & Adam, C. (2021). What is learned from exposure: An  
15 error-driven approach to productivity in language. *Language, Cognition and Neuroscience*,  
16 *36*(1), 60–83.
- 17 Dubois, T., Paquot, M., & Szmrecsanyi, B. (2022). Alternation phenomena and language proficiency:  
18 The genitive alternation in the spoken language of EFL learners. *Corpus Linguistics and*  
19 *Linguistic Theory*.
- 20 Dziemianko, A. (2008). Subject-Verb Concord with Collective Nouns or the Count-Mass Distinction:  
21 Which is More Difficult for Polish Learners of English? *Poznań Studies in Contemporary*  
22 *Linguistics*, *44*(4), 449–468.
- 23 Ellis, N. C. (2006). Cognitive perspectives on SLA: The associative-cognitive CREED. *AILA Review*, *19*,  
24 100–121.
- 25 Ellis, N. C. (2017). Cognition, Corpora, and Computing: Triangulating Research in Usage-Based  
26 Language Learning. *Language Learning*, *67*(S1), 40–65.



## Animacy in the genitive alternation

- 1 Ellis, N. C., & Wulff, S. (2019). Cognitive approaches to Second Language Acquisition. In J. W.  
2 Schwieter & A. Benati (Eds.), *The Cambridge Handbook of Language Learning* (pp. 41–61).  
3 Cambridge University Press.
- 4 Engel, A. (2022). *The register-specificity of probabilistic grammars in English and Dutch—Combining*  
5 *corpus analysis and experimentation*. KU Leuven.
- 6 Engel, A., Grafmiller, J., Rosseel, L., & Szmrecsanyi, B. (2022). Assessing the complexity of lectal  
7 competence: The register-specificity of the dative alternation after give. *Cognitive Linguistics*,  
8 33(4), 727–766.
- 9 Engel, A., & Szmrecsanyi, B. (2023). Variable grammars are variable across registers: Future temporal  
10 reference in English. *Language Variation and Change*, 34(3), 355–378.
- 11 Fox, J. (2003). Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software*, 8, 1–  
12 27.
- 13 Gablasova, D., Brezina, V., & McEnery, T. (2019). The Trinity Lancaster Corpus: Development,  
14 description and application. *International Journal of Learner Corpus Research*, 5(2), 126–158.
- 15 Gablasova, D., Brezina, V., Mcenery, T., & Boyd, E. (2017). Epistemic Stance in Spoken L2 English: The  
16 Effect of Task and Speaker Style. *Applied Linguistics*, 38(5), 613–637.
- 17 Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in*  
18 *Medicine*, 27(15), 2865–2873.
- 19 Gilquin, G., & Gries, S. Th. (2009). Corpora and experimental methods: A state-of-the-art review.  
20 *Corpus Linguistics and Linguistic Theory*, 5(1), 1–26.
- 21 Gliem, J. A., & Gliem, R. R. (2003). Calculating, Interpreting, And Reporting Cronbach’s Alpha  
22 Reliability Coefficient For Likert-Type Scales. *Proceedings of the 2003 Midwest Research to*  
23 *Practice Conference in Adult, Continuing, and Community Education*, 82–88.
- 24 Gries, S. Th. (2015). The most under-used statistical method in corpus linguistics: Multi-level (and  
25 mixed-effects) models. *Corpora*, 10(1), 95–125.

## Animacy in the genitive alternation

- 1 Gries, S. Th., & Wulff, S. (2013). The genitive alternation in Chinese and German ESL learners:  
2 Towards a multifactorial notion of context in learner corpus research. *International Journal of*  
3 *Corpus Linguistics*, 18(3), 327–356.
- 4 Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic  
5 priming persists while the lexical boost decays: Evidence from written and spoken dialogue.  
6 *Journal of Memory and Language*, 58(2), 214–238.
- 7 Heller, B., Szmrecsanyi, B., & Grafmiller, J. (2017). Stability and fluidity in syntactic variation world-  
8 wide: The genitive alternation across varieties of English. *Journal of English Linguistics*, 45(1),  
9 3–27.
- 10 Jäschke, K., & Plag, I. (2016). The dative alternation in German-English interlanguage. *Studies in*  
11 *Second Language Acquisition*, 38(3), 485–521.
- 12 Kempen, G., & Harbusch, K. (2005). The Relationship between Grammaticality Ratings and Corpus  
13 Frequencies: A Case Study into Word Order Variability in the Midfield of German Clauses. In  
14 S. Kepser & M. Reis (Eds.), *Linguistic Evidence: Empirical, Theoretical and Computational*  
15 *Perspectives* (pp. 329–350). De Gruyter Mouton.
- 16 Kinne, A. (2020). *Particle Placement in English L1 and L2 Academic Writing*. Presses universitaires de  
17 Louvain.
- 18 Klavan, J. (2020). Pitting corpus-based classification models against each other: A case study for  
19 predicting constructional choice in written Estonian. *Corpus Linguistics and Linguistic Theory*,  
20 16(2), 363–391.
- 21 Klavan, J., & Divjak, D. (2016). The cognitive plausibility of statistical classification models: Comparing  
22 textual and behavioral evidence. *Folia Linguistica*, 50(2), 355–384.
- 23 Kumle, L., Vö, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed  
24 models: An open introduction and tutorial in R. *Behavior Research Methods*, 2528–2543.
- 25 MacDonald, M. C. (2013). How language production shapes language form and comprehension.  
26 *Frontiers in Psychology*, 4(Article 226), 1–16.

## Animacy in the genitive alternation

- 1 Milin, P., Divjak, D., Dimitrijević, S., & Baayen, R. H. (2016). Towards cognitively plausible data science  
2 in language research. *Cognitive Linguistics*, 27(4), 507–526.
- 3 Murphy, R. (2012). *English Grammar in Use. A self-study reference and practice book for intermediate*  
4 *learners of English*. Cambridge University Press.
- 5 Prolific. (2023). *Prolific* [Computer software]. Prolific. <https://www.prolific.co>
- 6 Qualtrics. (2023). *Qualtrics* [Computer software]. Qualtrics. <https://www.qualtrics.com>
- 7 R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for  
8 Statistical Computing. <https://www.R-project.org/>
- 9 Rosenbach, A. (2014). English genitive variation—The state of the art. *English Language and*  
10 *Linguistics*, 18(2), 215–262.
- 11 Segalowitz, N. (2003). Automaticity and Second Languages. In *The Handbook of Second Language*  
12 *Acquisition* (pp. 382–408). John Wiley & Sons, Ltd.  
13 <https://doi.org/10.1002/9780470756492.ch13>
- 14 Stefanowitsch, A. (2006). Negative evidence and the raw frequency fallacy. *Corpus Linguistics and*  
15 *Linguistic Theory*, 2(1), 61–77.
- 16 Szmrecsanyi, B. (2006). *Morphosyntactic persistence in spoken English. A corpus study at the*  
17 *intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Mouton  
18 de Gruyter.
- 19 Wolk, C., Bresnan, J., Rosenbach, A., & Szmrecsanyi, B. (2013). Dative and genitive variability in Late  
20 Modern English: Exploring cross-constructural variation and change. *Diachronica*, 30(3),  
21 382–419.
- 22 Wulff, S., & Gries, S. T. (2019). Particle Placement in Learner Language. *Language Learning*, 69(4),  
23 873–910.
- 24 Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., O'Connor, M. C.,  
25 & Wasow, T. (2004). Animacy encoding in English: Why and how. In B. Webber & D. Byron

Animacy in the genitive alternation

- 1 (Eds.), *Proceedings of the 2004 ACL Workshop on Discourse Annotation, Barcelona, July 2004*
- 2 (pp. 118–125). Association for Computational Linguistics.
- 3