

# Transfer and Bias in Norwegians' Judgments of Word Order in L2 English

Sunniva Briså Strætkevorn  
(Norwegian University of Science and Technology)

## *Abstract*

This study investigates transfer of V2 relative to certain adverbs in L1 Norwegian L2 English speakers of intermediate/advanced proficiency. Previous research has found transfer of V2 word order, especially relative to adverbs in elementary school children (i.e., beginners) (Westergaard 2002, 2003) as well as in highschoolers (Dahl et al. 2022; Listhaug et al. 2021). However, the question remains whether this tendency persists in intermediate/advanced adult speakers. The present study uses Acceptability Judgment Tasks (AJTs) to investigate the presence of V2 transfer in this speaker group. The use of AJTs to investigate transfer is complicated by the fact that there is sometimes talk of L2 users having a yes-bias (Romano and Guijarro-Fuentes 2023). In this study, I investigate the presence of such a bias in AJTs, and whether the emergence of a bias is potentially impacted by access to metalinguistic knowledge. I conducted two AJTs where access to metalinguistic knowledge differed. I used Signal Detection Theory (Bader and Häussler 2010) to get a measure of participants' sensitivity ( $d'$ ) and their decision criterion to investigate the presence of bias. The results revealed a clear no-bias in both experiments, suggesting we cannot always assume a yes-bias in L2 users. No detectable transfer of V2 relative to adverbs was found in this group of L2 users, whose performance was at ceiling. Furthermore, as the only way to detect V2 transfer in the data was for speakers to accept ungrammatical structures, any potential presence thereof may have been masked by this salient no-bias.

Keywords: transfer; bias; residual optionality; V2; Signal Detection Theory

## *1. Introduction*

The native language (L1) plays a central role in the acquisition of a second language (L2), and it is a well-known phenomenon that L2 users are prone to exhibit influence from their L1. This is usually referred to as *transfer* (Schwartz and Sprouse 1996; Sorace 2003), and it can manifest in areas such as word order (see, e.g., Westergaard 2002, 2003; Håkansson et al.

Strætkevorn, Sunniva Briså. 2023. 'Transfer and Bias in Norwegians' Judgments of Word Order in L2 English.' *Nordic Journal of English Studies* 22(2): 101–127.

2002; Bohnacker 2006; Rankin 2012). For instance, an L1 speaker of Norwegian learning English as an L2 may initially produce ill-formed sentences such as (2) and (4):

- (1) *Anna leser ofte vitenskapelige artikler.*  
[‘Anna often reads scientific papers.’ (lit. Anna reads often scientific papers.)]
- (2) \*Anna reads often scientific papers.
- (3) *Ingrid drikker alltid kaffe.*  
[‘Ingrid always drinks coffee.’ (lit. Ingrid drinks always coffee.)]
- (4) \*Ingrid drinks always coffee.

These examples show cases of non-facilitative transfer, where the use of a structure from the L1 results in ungrammaticality in the L2. Norwegian and English are both SVO languages, but while Norwegian exhibits so-called ‘Verb Second’ (‘V2’) in main clauses, English does not permit this word order with lexical verbs. Thus, whenever an adverb is present in English, the main verb appears after it, thereby occurring as the third element (henceforth referred to as ‘V3’). Previous research has attested transfer of V2 in native speakers of Norwegian in their early stages (grades 1–7) of learning English as an L2, especially relative to adverbs (Westergaard 2002, 2003). Moreover, this tendency appears to persist in learners at higher proficiency levels: Listhaug et al. (2021) and Dahl et al. (2022) both found traces of V2 transfer relative to adverbs in high school students (grades 11–12). However, we do not know whether this persists in advanced adult learners. Even if the non-target word order may not surface in production, it could potentially be detected in judgment tasks if these speakers still retain some residual optionality in their grammar (Sorace 2003). Moreover, L2 users are sometimes said to have a yes-bias in judgment tasks (see, e.g., Tokowicz and MacWhinney 2005; Stadt et al. 2016; Cheng and Mayberry 2020; Romano and Guijarro-Fuentes 2023). On the other hand, Domaas (2016), who investigated intermediate/advanced L2 users of English, found the L2 users to be systematically less willing to accept grammatical items than the control group, which is, essentially, a no-bias (although the exact term is never

used by Domaas herself). A bias can skew the data and potentially dilute differences between conditions, which can mask the presence or absence of potential transfer. If participants have a yes-bias, it can appear as though they have acquired the target structure because they tend to accept those items, when in reality they simply tend to accept items regardless of their grammaticality. On the other hand, a no-bias (i.e., under-acceptance of items, including grammatical ones), can make it appear as though participants are good at rejecting the non-target structures, when in fact, they just tend to reject items overall. As such, response bias should be accounted for in the data analysis. Furthermore, different types of judgment tasks can reveal different aspects of learners' mental grammar: a task that allows learners access to their metalinguistic knowledge is likely to provide insight into learners' competence, whereas a task that restricts metalinguistic access is likely to tap more into learners' performance (see, e.g., Spinner and Gass 2019). The purpose of this study is to investigate the presence of V2 transfer in intermediate/advanced L2 users of English, and to determine whether these speakers are biased in their judgments. Additionally, the study investigates how access to metalinguistic knowledge impacts participants' judgments and bias. Two acceptability judgment tasks were conducted, one which allowed access to metalinguistic knowledge, and another where access was restricted.

This article is organized as follows: section 2 contains the theoretical background for the present study, while section 3 outlines the methodology, research questions, and results of the two experiments conducted. The results are discussed in section 4, and concluding remarks are given in section 5.

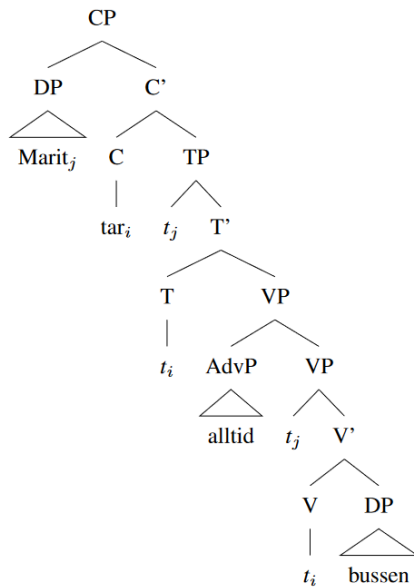
## 2. Theoretical background

### 2.1 Verb and adverb placement in Norwegian and English

In the following, I give a simplified description of some differences between word order in Norwegian and English main and relative clauses within a generative framework. In simple declarative main clauses, Norwegian and English display different word order whenever a frequency adverb such as *often* or *always* is present. In Norwegian, finite verbs move to C, while the subject moves to Spec-CP (this is the so-called V2-rule) (see, e.g., Áfarli and Eide 2003). For the present analysis, it is assumed that adverbs such as *ofte* ('often') and *alltid* ('always') are adjoined to the VP (see, e.g., Bobaljik and Thráinsson 1998), as they are preceded by the

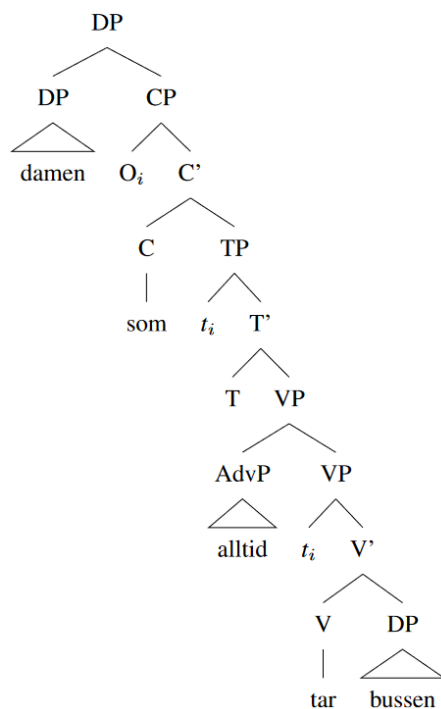
finite verb in main clauses, as shown in (5) and the corresponding tree diagram:

- (5) *Marit tar alltid bussen.* / \**Marit alltid tar bussen.*  
 Marit takes always bus-DEF. / \*Marit always takes bus-DEF.  
 [‘Marit always takes the bus.’]



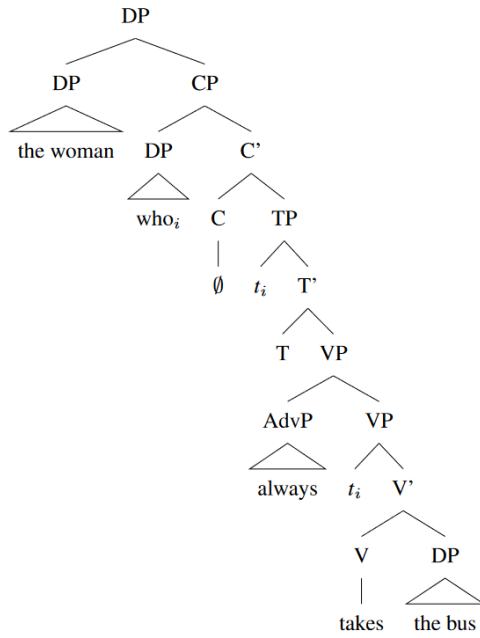
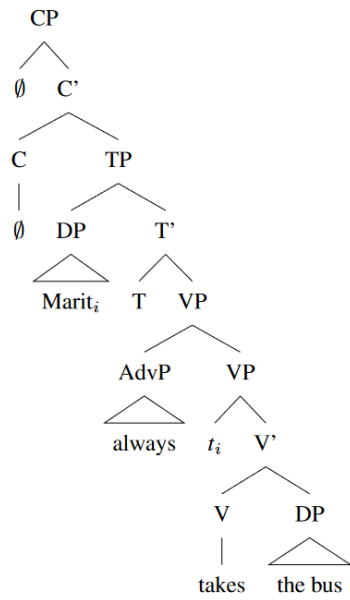
Regarding relative clauses, I follow the standard assumption for subordinate clauses in Norwegian, namely that the verb does not move out of the VP domain (Holmberg and Platzack 1995; Lohndal et al. 2020). Although some Norwegian dialects permit V2 in certain subordinate clauses (see, e.g., Bentzen 2005), this is not the case for relative clauses, as made apparent in (6) (the analysis of the relative pronoun is based on Áfarli and Eide 2003):

- (6) *Ola møtte damen som alltid tar bussen.* / \**Ola møtte damen som tar alltid*  
 Ola met woman-DEF who always takes bus-DEF. / \*Ola met  
 woman-DEF who takes always bus-DEF.  
 [‘Ola met the woman who always takes the bus.’]



In contrast to Norwegian, English displays the same word order in main and relative clauses (i.e., V3). This is because English main verbs occupy a low position in the clause, and lexical verbs do not move out of the VP (see, e.g., Pollock 1989; Cinque 1999; Radford 2004). As a result, adverbs adjoined to VP intervene between the subject and the predicate (relative pronoun analysis based on Radford 2004):

- (7) Marit always takes the bus. / \*Marit takes always the bus.
- (8) Ola met the woman who always takes the bus. / \*Ola met the woman [who takes always the bus].



Accordingly, English and Norwegian exhibit the same word order with lexical verbs in relative clauses (i.e., V3), whereas in main clauses their word order differs when an adverb is present.

### 2.2 *Transfer in L2 acquisition*

Several proposals have been made to account for the nature of L1 influence in L2 acquisition, and terms such as 'interference', 'crosslinguistic influence' (CLI), and, perhaps most frequently, 'transfer', are commonly used to describe the phenomenon (Sharwood Smith 1983; Schwartz and Sprouse 1996; Westergaard 2021). The distinct terms are sometimes employed in the context of distinguishing between learners' competence and performance (see Grosjean 2011). However, 'transfer' and 'CLI' are often used interchangeably. In this study, I will simply use the term *transfer* to refer to any kind of systematic influence from the L1 on the L2 during the acquisition process.

Several early models proposed some type of partial transfer, whereas Schwartz and Sprouse's (1996) Full Transfer Full Access maintained that the complete L1 grammar transfers. During the last few decades, a substantial body of research has provided compelling evidence in favor of full transfer, and there is general consensus that the partial transfer models of the 1990s cannot account for existing empirical evidence. Today, the debate is mainly concerned with whether transfer entails that the brain creates a *real* full copy of the L1 grammar (Schwartz and Sprouse 1996), or whether everything is simply *available* for transfer ('everything *does* transfer' vs. 'everything *may* transfer'), i.e., a Full Transfer Potential (Westergaard 2021).

Native speakers of Norwegian generally reach a high proficiency level in English by the time they are in high school, and most adult speakers fall into the category of advanced proficiency. In Norway, English is taught in schools from the first grade, and Norwegians receive a lot of exposure to the language on an everyday basis through various channels. If we assume that all aspects of the L1 will or may transfer to the L2, we expect Norwegian L2 learners of English to (potentially) transfer V2 with lexical verbs relative to adverbs in main clauses—indeed, this is what Westergaard (2002, 2003) found, and Dahl et al. (2022) found that high school students (grades 11–12) were less consistent in rejecting V2 in main clauses relative to adverbs than they were at rejecting V2 in topicalized structures. Even though speakers of advanced proficiency will have

received a substantial amount of positive evidence of non-V2 for their L2 grammar to have restructured accordingly, they may still retain some optionality regarding V2 (see Beck 1998; Robertson and Sorace 1999), as this can even persist in speakers who have attained near-native proficiency (Sorace 2003). As such, it is possible that optionality regarding V2 can also be found in this group of advanced L2 speakers.

### *2.3 Previous research on verb placement in L2*

There is evidence in SLA research that acquiring verb placement poses a particular challenge for L2 learners if it differs in their L1. Not only does this apply to learners at early stages, but it also appears to persist at higher levels of proficiency. Westergaard's (2002, 2003) judgment study of L1 Norwegian elementary school children in grades 1–7 learning English as an L2 found 'massive transfer' of V2, especially relative to adverbs, while Robertson and Sorace (1999) found that L1 German speakers retain residual V2 optionality in their L2 English at advanced proficiency. Non-target-like verb placement at high proficiency level was also observed in Rankin's (2012) comparative learner corpus study, which investigated texts written by university students (L1 German or Dutch) and showed that they still exhibited transfer of V2 word order in their L2 English.

Judgment studies of learners at high school level also suggest that there are still traces of transfer pertaining to verb and adverb placement. Dahl et al. (2022) and Listhaug et al. (2021) both found that participants were less consistent in the ratings they assigned to these structures, which might suggest that these speakers still retain some residual optionality.

### *2.4 Acceptability Judgment data and Signal Detection Theory*

Judgment data play an important role in SLA research, as they allow researchers to investigate what the L2 learner's developing grammar looks like by comparing it to what we assume is allowed in the target language (i.e., the standard set by native speakers). In an Acceptability Judgment Task (AJT), participants report whether and (usually) to what extent a specific string of words sounds 'good' or 'bad' to them (Schütze and Sprouse 2014). AJTs are thus an indirect measurement method, as the data collected are a reported perception of acceptability (Chomsky 1965; Schütze 1996; Sprouse and Almeida 2012). While the method is not exempt from criticism (see, e.g., Selinker 1972), AJT data are generally



considered both valid and valuable in linguistic research (see, e.g., Schütze and Sprouse 2014). What must be kept in mind when using judgments for SLA research, however, is that the acquisition of an L2 often involves obtaining some form of metalinguistic knowledge. We therefore cannot know whether the judgments elicited from participants reflect their spontaneous intuitions, or whether they also rely on consulting their metalinguistic knowledge when producing the judgment. Essentially, this means that different results may be obtained depending on whether the tasks employed are online or offline. Abbas et al. (2021) found online measures to be reflective of the automated activation of morpho-syntactic structures, whereas offline judgment tasks allow participants to use metalinguistic strategies as well. One way to control and restrict access to metalinguistic knowledge during a judgment task is to implement a time constraint (Sorace 1996). Thus, a timed judgment task will likely tell us more about participants' implicit knowledge, whereas untimed AJTs are more likely to measure explicit knowledge (Bialystok 1979; Ellis 2005; Loewen 2009; Morgan-Short et al. 2014; Godfroid et al. 2015; Spinner and Gass 2019).

From a theoretical point of view, we use judgment data to determine whether participants know which sentences are allowed by the target language grammar and which are not, to inform us how reliably participants can discriminate between good and bad sentences. Most judgment studies that collect acceptability ratings compare mean acceptability scores between conditions as a proxy for discriminability. However, comparing mean judgments only provides an imperfect measure of discriminability, because statistically significant differences between condition means can arise as long as there is some degree of discriminability either large or small. Having a clear estimate of discriminability can be crucial for inference about the status of particular constructions in developing grammars, particularly when optionality is at issue. Comparing mean acceptability scores across conditions also has another drawback when it comes to inference: such analysis does not take into account or provide information about (potentially confounding) factors that influence participants' performance in a task. One such factor is *response bias*. Unlike random variability and noise, a response bias is systematic (see Huang and Ferreira 2020), and as such, it can obscure (the researcher's measurement of) participants' ability to discriminate between good and bad items (e.g., does a low acceptability score mean that a

participant thinks an item is bad, or are they just biased to give low scores overall?). There are some accounts of L2 learners having a yes-bias in judgment tasks, i.e., they tend to show acceptance for ungrammatical items (e.g., Tokowicz and MacWhinney 2005; Stadt et al. 2016; Cheng and Mayberry 2020; Romano and Guijarro-Fuentes 2023), though no-biases have also occasionally been found (see Domaas 2016).

A way to more accurately measure discriminability and account for response bias is to apply Signal Detection Theory (SDT) to judgment data (Bader and Häussler 2010; Dillon and Wagers 2019; Huang and Ferreira 2020). First, SDT provides a measurement of sensitivity,  $d'$  (Green and Swets 1966)<sup>1</sup>, which indicates participants' ability to distinguish good items from bad ones. The higher the  $d'$  value, the more reliably participants can discriminate between good and bad sentences of a particular type. Lower  $d'$  values indicate that participants have a harder time discriminating bad from good sentences (i.e., they either accept bad sentences or reject good sentences more often). A  $d'$  of zero indicates that participants cannot tell the difference between good and bad forms of a sentence type. Second, SDT lets us calculate participants' bias,  $c^2$ , which indicates their threshold for giving one type of response. If participants have a high threshold for giving a yes-response (i.e., if they are conservative in their judgments), they have a no-bias, whereas a low threshold (i.e., if participants are liberal in their judgments) constitutes a yes-bias.

SDT analysis is most commonly applied to binary judgment tasks, such as a Yes/No (YN) task. A YN task has two types of stimuli ('good' and 'bad'), and two types of response ('yes' or 'no', i.e., participants can accept or reject the stimuli).

---

<sup>1</sup> Sensitivity  $d'$  is measured in standard deviations and is calculated by subtracting the z-scored proportion (i.e., the normal distribution) of false positives (F) from the proportion of true positives (T) (i.e., this constitutes all the trials to which participants responded 'Yes'):  $d' = z(T) - z(F)$ .

<sup>2</sup> I use criterion location  $c$  to calculate bias.  $C$  describes the distance between the threshold for giving one type of response and the midpoint of the two distributions:  $c = -\frac{1}{2} (z(T) + z(F))$ .

This yields four possible outcomes<sup>3</sup>, which are shown Table 1.

Table 1. Categories of judgment in SDT

		Stimuli (item)	
		Good	Bad
Response	Acceptable	True positive	False positive
	Unacceptable	False negative	True negative

It should be noted that even though participants make a binary decision, we do not have to assume that the underlying construct is binary. Generally, acceptability is understood as a gradient rather than a clear-cut divide between the categories 'acceptable' and 'unacceptable'. But we can still *measure* acceptability with a binary decision: participants will then have to set a threshold for themselves below which they reject items and above which they accept them (Huang and Ferreira 2020).

### 3. The present study

The aim of the present study is to investigate whether adult intermediate and advanced L1 Norwegian L2 English speakers still exhibit transfer of V2 (i.e., residual optionality) relative to certain adverbs in main and relative clauses. I hypothesize that participants will show higher acceptance for V2 in main clauses than in relative clauses because of potential transfer from their L1 Norwegian. I sought to answer the following research questions:

1. Do adult intermediate and advanced L2 English users exhibit residual V2 optionality relative to certain adverbs (such as *often* and *always*)?
2. Do these speakers have a bias in judgments in L2 English?
3. How does access to metalinguistic knowledge affect participants' judgments and bias?

<sup>3</sup> Huang and Ferreira (2020) use the terms *hit*, *miss*, *false alarm* and *correct rejection* for these four scenarios. Here I have instead used the terms *true positive* (= 'hit'), *false negative* (= 'miss'), *true negative* (= 'correct rejection') and *false positive* (= 'false alarm'), as they are arguably more intuitive.

### *3.1 Method*

The project was registered with the Norwegian Centre for Research Data (NSD), and participants gave their informed consent through an online form. I conducted two experiments: Experiment 1 was an untimed AJT, whereas Experiment 2 was an AJT that used Rapid Serial Visual Presentation (RSVP), thereby imposing a time constraint on the reading of the sentences. The two experiments contained the same items but used different participants. I set the following exclusion criteria: for both experiments, participants whose accuracy score was lower than 70% on filler items were excluded from the analysis. Additionally, for Experiment 1, participants who had taken less than five minutes to complete the survey were excluded.

### *3.2 Experiment 1*

Experiment 1 was an untimed AJT (YN task). This experiment was conducted to investigate judgments when metalinguistic awareness is accessible to participants.

#### *3.2.1 Participants*

The participants in Experiment 1 ( $n = 96$ , aged 18–85, mean age 42.5) were mainly recruited through social media and represented a wide range of the Norwegian population in terms of age and educational background (34 participants reported having studied English at university level). The only criteria for participating were that their (only) native language was Norwegian, that they had some knowledge of English (which virtually all Norwegians have), and that they were at least 18 years of age. A control group of native speakers of English ( $n = 25$ , aged 27–76, mean age 43) participated in the experiment. These were recruited through contacts in the US who shared the experiment in their social networks.

#### *3.2.2 Materials and procedure*

I used a  $2 \times 2$  factorial design that crossed clause type and verb position. Clause type had two levels: Main Clause (MC) and Relative Clause (RC), and verb position had two levels: V2 and V3. An example of a target item is shown in Table 2.

Table 2. Example item from the experiment

<b>MC/V3</b>	The baker was careful but the apprentice <b>often burned</b> the croissants in the oven.
<b>MC/V2</b>	*The baker was careful but the apprentice <b>burned often</b> the croissants in the oven.
<b>RC/V3</b>	The baker disliked the apprentice who <b>often burned</b> the croissants in the oven.
<b>RC/V2</b>	*The baker disliked the apprentice who <b>burned often</b> the croissants in the oven.

There were 24 target items in the experiment, which all had four versions (as shown in Table 2). The items were distributed across four separate lists, so that each list contained only one version of each item. As such, each list contained six different lexicalizations of each condition. In addition to the target items, there were 36 fillers, half of which were ungrammatical. Each list was pseudo-randomized so that items of the same type did not immediately follow each other (see Spinner and Gass 2019). Participants were randomly assigned to one of the four different lists. In order to ensure that the items were matched for length and complexity, the MC versions were comprised of two coordinated main clauses (the structural manipulation, i.e., V2 or V3, occurring inside the latter), whereas the RC versions comprised one main clause with an embedded/relative clause containing the structural manipulation. The items only varied by content words and in some cases function words that did not affect the structural manipulation (see Sprouse and Almeida 2017).

The experiment was set up as an electronic survey on the platform Nettskjema<sup>4</sup>, to which participants were given access through a weblink. Participants were instructed that the purpose of the survey was *not* to test their proficiency level in English, but to see what their intuitions were (see Spinner and Gass 2019). Each item was accompanied by the question 'Er dette en god setning?' ('Is this a good sentence?'), to which participants had to respond either *Ja* ('Yes') or *Nei* ('No'). Six items were displayed on each page of the survey, and participants had the opportunity to move back and forth between pages and potentially change their answers. The survey included a short practice phase which contained items representing different levels of acceptability so that participants could get a feeling of how they would use the response buttons before they began the experiment itself. The practice items were not included in the analysis.

<sup>4</sup> <https://nettskjema.no/>.

Before beginning the experiment, participants responded to a questionnaire on their language background, which included doing a self-report of their proficiency level in L2 English on a scale based on levels of the Common European Framework of Reference (CEFR). Participants were also asked to report if they had a diagnosis that could affect their language learning. In the L2 group, 4 participants reported having a reading disability, whereas 3 participants reported having an additional native language. In the control group, no participants reported having a reading disability or an additional native language. No participants in either group were under 18 years of age. Unfortunately, due to a technical error in Nettskjema, it was not possible to match the participants' survey responses with their individual background information, and thus it was not possible to exclude individual participants from the experiment based on their answers in the background form. However, outliers were removed during data analysis according to the exclusion criteria mentioned in section 3.1.

### 3.2.3 Results of Experiment 1

The results from the L2 speaker group ( $n = 88$ ) in Experiment 1 are summarized in Table 3. For the V3 items, a 'yes'-response constitutes a true positive while a 'no'-response constitutes a false negative, whereas for the V2 items, a 'no'-response constitutes a true negative while a 'yes'-response constitutes a false positive.

A paired samples  $t$ -test was conducted on individual participants' scores to see whether there was a significant difference in sensitivity  $d'$  between the MC and RC conditions.<sup>5</sup> The  $t$ -test revealed that the difference was not significant ( $t = -0.11, p = 0.91$ ). The values of  $c$  are both positive, which indicates that the group had an overall no-bias (see Huang and Ferreira 2020). Looking at bias in individual participants, it was found that for the MC condition, 42 participants had zero bias. The no-bias found in the aggregate analysis essentially came from 41 participants (the remaining 5 participants each had a small yes-bias). For the RC condition, 38 participants displayed zero bias, while 47 participants had a no-bias (a yes-bias was found in the remaining 3 participants).

---

<sup>5</sup> As the  $d'$  function cannot handle perfect accuracy, I added 0.5 to all data cells for participants for whom this was the case (cf. Hautus 1995).

The results for the native speaker control group ( $n = 23$ ) are summarized in Table 4.

Table 3. The L2 group's responses in Experiment 1

		Clause type			
		MC		RC	
Verb position		V3 (good)	V2 (bad)	V3 (good)	V2 (bad)
Response	Acceptable	445	15	441	9
	Unacceptable	83	513	87	519
	Mean Acc. (sd)	0.907 (0.11)		0.909 (0.10)	
	Range Acc. by subj.	0.58-1.00		0.58-1.00	
Measure	Sensitivity $d'$	2.911		3.094	
	Bias $c$	0.449		0.572	

Table 4. The control group's responses in Experiment 1

		Clause type			
		MC		RC	
Verb position		V3 (good)	V2 (bad)	V3 (good)	V2 (bad)
Response	Acceptable	112	12	114	9
	Unacceptable	26	126	24	129
	Mean Acc. (sd)	0.862 (0.16)		0.880 (0.14)	
	Range Acc. by subj.	0.42-1.00		0.58-1.00	
Measure	Sensitivity $d'$	2.244		2.451	
	Bias $c$	0.238		0.287	

A paired samples  $t$ -test revealed that the difference in sensitivity for MC and RC was not significant ( $t = -0.54$ ,  $p = 0.59$ ). The values of  $c$  are positive for both MC and RC, indicating a no-bias in the group. As for bias in individuals, I found that for the MC condition 6 participants displayed zero bias. The no-bias found in the aggregate analysis came from 13

participants (the remaining four participants had a yes-bias). For the RC condition, 12 participants displayed zero bias, while 9 participants displayed a no-bias (the remaining two displayed a yes-bias).

In summary, there is virtually no detectable transfer of V2 in the L2 group. The values of  $d'$  (which are measured in standard deviations) are extremely high in this group, indicating high sensitivity. Overall, their performance was at ceiling. In fact, they showed a higher sensitivity and accuracy in their judgments than the L1 control group. It is possible that test-taking anxiety (Purpura 2004) may have played a role: even though the instructions explicitly stated that the purpose of the task was *not* to test their English skills, the L2 group may still have wanted to perform well and thus have been more thorough than the control group. It is also possible that some participants in this group (especially those who reported having studied English at university) may have participated in similar experiments before, potentially giving them an advantage, as providing judgments is 'a learned skill' (Spinner and Gass 2019). The results also showed that there was a clear no-bias in both groups, which was largest in the L2 group. One possible explanation is that a no-bias could arise because of the untimed nature of the task: when participants can spend as much time as they like to judge the sentences and consult their metalinguistic knowledge (Sorace 1996), they might also become more critical in their assessment. I therefore wanted to investigate whether the no-bias is present under timed conditions.

### 3.3 Experiment 2

I conducted Experiment 2 as a follow-up to Experiment 1 to investigate whether judgments would differ when access to metalinguistic knowledge is constrained. The aim was to discover whether the prominent no-bias found in Experiment 1 is present in a more cognitively demanding task where certain factors that allow deliberation would be less likely to affect participants' judgments.

#### 3.3.1 Participants

The participants in Experiment 2 were recruited through social media as well as on campus at a major university in Norway. They were all L1 Norwegian L2 English speakers ( $n = 27$ , aged = 20–69, mean age = 32.5), and, crucially, they were not the same participants as in Experiment 1. 15



of the participants reported having studied English at university. A control group of native speakers of English ( $n = 31$ ) also participated in the experiment, none of whom had participated in Experiment 1. These participants were resident in Norway at the time of testing, and more than half of them reported that they had at least intermediate proficiency in Norwegian. No significant differences in judgments were found between those who reported that they knew Norwegian and those who did not. Participants who reported that they had more than one native language were excluded from the study.

### *3.3.2 Materials and procedure*

Experiment 2 was set up on the platform *Ibex Farm*<sup>6</sup> (Drummond 2018) which was accessible through a weblink. The experiment contained the same items and used the same factorial design as Experiment 1. The items were displayed to participants using RSVP: sentences were presented word by word on the screen, each one displaying for 400 ms. After each sentence had finished displaying, participants had to rate it as either 'Good' or 'Bad'. As such, Experiment 2 differed from Experiment 1 in three aspects which all restrained access to metalinguistic knowledge: 1) only one sentence was presented at a time (as opposed to chunks of six), thereby disallowing immediate comparison with other sentences, 2) participants could not reread any of the sentences (and potentially reconsider their assessment of them), and 3) the task involved a time constraint, ensuring that judgments were more immediate. Before beginning the experiment, participants answered a short questionnaire about their language background, similar to that in Experiment 1. The experiment began with a short practice phase so participants could get used to the mode of presentation. The practice items were not included in the analysis.

### *3.3.3 Results of Experiment 2*

The results from the L2 English group ( $n = 20$ ) in Experiment 2 are summarized in Table 5.

---

<sup>6</sup> <https://github.com/addrummond/ibexfarm>. The experiment ran on a version of *Ibex Farm* hosted at the Norwegian University of Science and Technology (NTNU).

Table 5. The L2 group's responses in Experiment 2

		Clause type			
		MC		RC	
Verb position		V3 (good)	V2 (bad)	V3 (good)	V2 (bad)
Response	Acceptable	103	13	103	8
	Unacceptable	17	107	17	112
	Mean Acc. (sd)	0.875 (0.14)		0.896 (0.11)	
	Range Acc. by subj.	0.42-1.00		0.58-1.00	
Measure	Sensitivity $d'$	2.308		2.574	
	Bias $c$	0.081		0.214	

A paired samples  $t$ -test revealed that the difference in sensitivity between the two conditions was not significant ( $t = -1.20, p = 0.24$ ). Positive values of  $c$  indicate a no-bias in analysis of the group in aggregate, which was slightly higher in the RC condition than the MC condition. When looking at individual bias in the group, I found that for the MC condition, 7 participants displayed zero bias and 9 individuals had a no-bias (the remaining four displayed a yes-bias). For the RC condition, 6 participants displayed zero bias, while the aggregate no-bias came from 11 participants (three individuals had a yes-bias).

The results for the native speaker control group ( $n = 26$ ) are summarized in Table 6. A paired samples  $t$ -test revealed that the difference in sensitivity between the two conditions was not significant ( $t = -1.18, p = 0.25$ ). The values of  $c$  are negative and indicate a slight yes-bias for both conditions. As for bias in individuals, I found that for the MC condition 9 participants displayed zero bias, whereas the yes-bias came from 9 participants, the remaining 8 displaying a no-bias. In the RC condition, 13 participants displayed zero bias. The yes-bias came from 11 participants, while the remaining 6 displayed a no-bias.

To see whether there was a significant difference in bias in the timed and untimed AJT, I ran a Wilcoxon rank sum test. The difference in bias was not significant either for the MC condition ( $W = 959.5, p = 0.51$ ) or for the RC condition ( $W = 961.5, p = 0.50$ ).

Table 6. The control group's responses in Experiment 2

		Clause type			
		MC		RC	
Verb position		V3 (good)	V2 (bad)	V3 (good)	V2 (bad)
<b>Response</b>	<b>Acceptable</b>	137	21	145	19
	<b>Unacceptable</b>	19	135	11	137
	<b>Mean Acc. (sd)</b>	0.872 (0.14)		0.903 (0.11)	
	<b>Range Acc. by subj.</b>	0.50-1.00		0.58-1.00	
<b>Measure</b>	<b>Sensitivity <math>d'</math></b>	2.271		2.638	
	<b>Bias <math>c</math></b>	-0.031		-0.153	

In summary, no detectable transfer of V2 was found in the L2 group, as the difference in sensitivity for MC and RC was not significant. Just as in Experiment 1, the L2 group's performance was at ceiling, and their accuracy was comparable to the control group. Once again, a no-bias was found in the L2 group. However, this bias was not found to be significantly smaller than the no-bias found in the untimed AJT. Unlike in the untimed AJT, the control group displayed a very small yes-bias for both conditions. It is possible that these speakers were more liberal in their acceptance of items due to their multilingual context; nevertheless, those who had knowledge of Norwegian did not show any significant difference in performance from those who did not.

#### 4. Discussion

This study aimed to answer three research questions: 1) whether Norwegian intermediate to advanced users of L2 English still show traces of V2 transfer relative to adverbs (i.e., residual optionality), 2) whether these speakers have a bias in their judgments, and 3) how access to metalinguistic knowledge affects speakers' judgments and bias. I found no tangible evidence of transfer of V2 relative to adverbs in these constructions in either of the two experiments. This stands in contrast to the findings of Westergaard (2002, 2003), as well as potentially to Dahl et al. (2020) and Listhaug et al. (2021), which suggests that when learners have reached advanced proficiency, any potential remnant of V2 transfer

or optionality is untraceable—at least with the tools I have used. It is nevertheless noteworthy that no trace of V2 transfer was detected, especially as instructions were given in Norwegian, which could potentially have primed participants with that language (cf. Jensen and Westergaard 2022). Even though speakers were slightly more accurate in their judgments of the RC items than the MC items (the condition in which L1-transfer could potentially manifest), the paired *t*-tests showed that there was no significant difference between participants' sensitivity for MC and RC items in either the timed or the untimed experiment. Therefore, the results do not lend support to V2-optionality being present in these speakers' grammars. However, a crucial finding in this study is the salient no-bias observed in the L2 groups in both experiments, which contrasts with the assumed yes-bias in L2 users (Tokowicz and MacWhinney 2005; Stadt et al. 2016; Cheng and Mayberry 2020; Romano and Guijarro-Fuentes 2023). I can only speculate why: perhaps, as these L2 users had such high proficiency, they may also have felt highly confident in their own judgments, striving to ensure that they did not accept any items that retained any ungrammaticality whatsoever. Moreover, this finding leaves open another possibility: the no-bias may have diluted the presence of V2 transfer in the data, because, essentially, the only way to determine the presence of V2 transfer/residual optionality in these experiments is for participants to accept the MC/V2 items. If participants are biased towards rejecting any item they feel uncertain about, the data simply will not be able to reveal any V2 transfer, even if it is present in participants' grammar. Given this no-bias, then, I cannot exclude the possibility that these speakers do retain residual V2-optionality in their grammar. Furthermore, the no-bias found in this study appears to align with the findings of Domaas (2016), whose participants were similar to those in the current study, and whose results arguably indicated a no-bias.

Another somewhat surprising finding is that the respective L2 groups in the two experiments displayed higher sensitivity (as well as higher accuracy) than their respective L1 control groups. It is possible that the L2 users felt under more pressure than the native speakers to perform well on what they may have perceived as a test of their L2 skills. As such, they may have put in more effort in reading the sentences thoroughly. It is also possible that some of the L2 speakers were more used to reading English with a view to spotting grammatical errors than the L1 speakers, as relatively many of them reported having studied English at university,

which is likely to have provided them with at least some metalinguistic insight. Moreover, it is plausible that some had participated in similar experiments before, and as such, they may have relied on their judgment-providing skills (Spinner and Gass 2019), consequently performing better than the control group.

As regards the impact of access to metalinguistic knowledge, the effect of experiment type (timed/untimed AJT) on bias was not significant. Therefore, this study does not lend any strong support to the assumption that a no-bias arises as a result of unrestrained access to metalinguistic knowledge. Nevertheless, it should be kept in mind that both groups in this study performed at ceiling. Therefore, there is still the possibility that an observable difference could emerge in participant groups with lower proficiency, such as the groups in Dahl et al. (2022) and Listhaug et al. (2021), who were less consistent in rejecting V2 relative to adverbs. Another point to consider is that the strong no-bias found in Experiment 1 originated from less than half of the participants in the experiment, whereas the majority of the remainder were unbiased (i.e., most of them had perfect accuracy). This seems to imply that, at least for a subgroup of the participants, untimed conditions make them more critical, i.e., they can freely scrutinize the sentences and look for elements they disapprove of once metalinguistic knowledge is accessible. Moreover, it is even possible that there would have been an observable effect of experiment type had the participant group in Experiment 2 been larger; however, at the time of data collection, opportunities for participant recruitment were severely restricted due to the ongoing pandemic. Thus, based on the results of the current study, I certainly cannot dismiss the possibility of a potential effect of experiment type. I therefore suggest that for the purposes of future research, the study ought to be replicated in groups with lower proficiency and of more equal size.

An alternate explanation for the no-bias must also be considered. As the participants did not receive any explicit instruction on how to treat punctuation in the task (see Spinner and Gass 2019), it is possible that they rejected some items based on their punctuation. Comma use differs slightly in Norwegian and English: Norwegian sometimes employs a comma where English does not. This could potentially explain why the no-bias was slightly smaller in the timed AJT, as words were presented individually and for a short time only, giving the participants less opportunity to assess punctuation and potentially reject the item based on,

e.g., a comma they considered to be missing. However, the potential effect of this cannot be determined. Moreover, as the difference between bias in the two experiments was shown not to be significant, I will refrain from further speculation here.

### *5. Conclusion*

The findings of this study essentially leave us with two possibilities: as no conclusive evidence of residual optionality/transfer of V2 relative to adverbs was found in these intermediate/advanced L2 users of English, it is possible that speakers who have attained this level of proficiency do not exhibit any residual optionality with respect to verb and adverb placement, or at least not to any degree that could be detected by an acceptability judgment task. While I fully acknowledge that judgment data alone cannot give a full picture of speakers' competence, this is nevertheless an interesting finding. If, on the other hand, transfer/residual optionality indeed *was* present, it would have been masked by the salient no-bias found in both experiments, as the only way to uncover it was for participants to accept the ungrammatical items. Moreover, the no-bias seemed to arise independently of experiment type. And finally, there is a correlation with the findings in Domaas (2016), whose participant group was similar to those in this study. This could imply that a no-bias rather than a yes-bias is a more common trend in L2 users whose proficiency level is very high. Future research should thus aim to investigate whether the no-bias arises particularly in highly proficient groups, rather than the supposedly typical yes-bias.

### *Acknowledgements*

The author thanks Associate Professor Dave Kush for his input on the design of the experiments. Thanks to the anonymous reviewers and Anna Giskes, Ingrid Bondevik, and Kristin Klubbo Brodahl for their valuable comments.

### *References*

Abbas, Nawras, Tamar Degani, and Anat Prior. 2021. Equal opportunity interference: Both L1 and L2 influence L3 morpho-syntactic

- processing. *Frontiers in Psychology* 12: 673535. doi.org/10.3389/fpsyg.2021.673535.
- Åfarli, Tor A., and Kristin M. Eide. 2003. *Norsk generativ syntaks*. Oslo: Novus.
- Bader, Markus, and Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46(2): 273–330. doi.org/10.1017/S0022226709990260.
- Beck, Maria-Luise. 1998. L2 acquisition and obligatory head movement: English-speaking learners of German and the Local Impairment Hypothesis. *Studies in second language acquisition* 20: 311–348.
- Bentzen, Kristine. 2005. What's the better move? On verb placement in Standard and Northern Norwegian. *Nordic Journal of Linguistics* 28(2): 153–188.
- Bialystok, Ellen. 1979. Explicit and implicit judgements of L2 grammaticality. *Language Learning* 29: 81–103. doi.org/10.1111/j.1467-1770.1979.tb01053.x.
- Bobaljik, Jonathan D., and Höskuldur Thráinsson. 1998. Two heads aren't always better than one. *Syntax* 1(1): 37–71.
- Bohnacker, Ute. 2006. When Swedes begin to learn German: From V2 to V2. *Second Language Research* 22(4): 443–486.
- Cheng, Qi, and Rachel I. Mayberry. 2020. When event knowledge overrides word order in sentence comprehension: Learning a first language after childhood. *Developmental Science* 24: 1–12. doi.org/10.1111/desc.13073.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. New York: Oxford University Press.
- Dahl, Anne, Kjersti Faldet Listhaug, and Guro Busterud. 2022. The role of L1 Norwegian and L2 English in the acquisition of verb placement in L3 German. In *Generative SLA in the age of minimalism*, edited by Tania Leal, Elena Shimanskaya, and Casilde A. Isabelli, 191–212. Amsterdam: John Benjamins.
- Dillon, Brian, and Matt Wagers. 2019. Approaching gradience in acceptability with the tools of signal detection theory. In *Cambridge handbook of experimental syntax*, edited by Grant Goodall, 60–96. Cambridge: Cambridge University Press.

- Domaas, Kari. 2016. I wonder who that took the dog out this morning: an investigation of L1 Norwegian transfer in L2 English indirect questions. Master's thesis. Trondheim: NTNU.
- Drummond, Alex. 2018. Ibex farm [Computer software]. <https://github.com/addrummond/ibexfarm>.
- Ellis, Rod. 2005. Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition* 27: 141–172. doi.org/10.1017/S0272263105050096.
- Godfroid, Aline, Shawn Loewen, Sehoon Jung, Ji-Hyun Park, Susan Gass, and Rod Ellis. 2015. Timed and untimed grammaticality judgments measure distinct types of knowledge. *Studies in Second Language Acquisition* 37: 269–297. doi.org/10.1017/S0272263114000850.
- Green, David M., and John. A. Swets. 1966. *Signal detection theory and psychophysics. Vol. 1*. New York: Wiley.
- Grosjean, François. 2011. An attempt to isolate, then differentiate, transfer and interference. *International Journal of Bilingualism* 16(1): 11–21.
- Håkansson, Gisela, Manfred Pienemann, and Susan Sayehli. 2002. Transfer and typological proximity in the context of second language processing. *Second Language Research* 18(3): 250–273.
- Hautus, Michael J. 1995. Corrections for extreme proportions and their biasing effects on estimated values of  $d'$ . *Behavior Research Methods, Instruments, & Computers* 27: 46–51. doi.org/10.3758/BF03203619.
- Holmberg, Anders, and Christer Platzack. 1995. *The role of inflection in Scandinavian syntax*. Oxford: Oxford University Press.
- Huang, Yujing, and Fernanda Ferreira. 2020. The application of signal detection theory to acceptability judgments. *Frontiers in Psychology* 11(73): 1–11. doi.org/10.3389/fpsyg.2020.00073.
- Jensen, Isabel Nadine, and Marit Westergaard. 2022. Syntax matters: Exploring the effect of linguistic similarity in third language acquisition. *Language Learning* 73(2): 374–402. doi.org/10.1111/lang.12525.
- Listhaug, Kjersti Faldet, Guro Busterud, and Anne Dahl. 2021. French as a third language in Norway: The influence of the prior languages in the acquisition of word order. In 'Syntax, semantics and acquisition: In honor of Hans Petter Helland', edited by Christine Meklenborg, Terje Lohndal, and Kathrine A. Østby. *Oslo Studies in Language* 12(1): 125–144.



- Loewen, Shawn. 2009. Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In *Implicit and explicit knowledge in second language learning, testing and teaching*, edited by Rod Ellis, Shawn Loewen, Catherine Elder, Rosemary Erlam, Jenefer Philp, and Hayo Reinders, 94–112. Bristol: Multilingual Matters.
- Lohndal, Terje, Marit Westergaard, and Øystein A. Vangsnes. 2020. Verb Second in Norwegian: Variation and acquisition. In *Rethinking Verb Second*, edited by Rebecca Woods and Sam Wolfe, 770–789. Oxford: Oxford University Press.
- Morgan-Short, Kara, Mandy Faretta-Stutenberg, Katherine A. Brill-Schuetz, Helen Carpenter, and Patrick C. Wong. 2014. Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition* 17: 56–72. doi.org/10.1017/S1366728912000715.
- Pollock, Jean Yves. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry* 20(3): 365–424.
- Purpura, James. 2004. *Assessing grammar*. Cambridge: Cambridge University Press.
- Radford, Andrew. 2004. *English syntax: An introduction*. Cambridge: Cambridge University Press.
- Rankin, Tom. 2012. The transfer of V2: inversion and negation in German and Dutch learners of English. *International Journal of Bilingualism* 16(1): 139–158.
- Robertson, Daniel, and Antonella Sorace. 1999. Losing the V2 constraint. In *The development of second language grammars. A generative approach*, edited by Elaine C. Klein and Gita Martohardjono, 317–361. Amsterdam: John Benjamins.
- Romano, Francesco, and Pedro Guijarro-Fuentes. 2023. Task effects and the yes-bias in heritage language bilingualism. *International Journal of Bilingual Education and Bilingualism*. doi.org/10.1080/13670050.2023.2206949.
- Schütze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- Schütze, Carson T., and Jon Sprouse. 2014. Judgment data. In *Research methods in linguistics*, edited by Robert J. Podesva and Devyani

- Sharma, 27–50. Cambridge: Cambridge University Press. doi.org/10.1017/CBO9781139013734.004.
- Schwartz, Bonnie D., and Rex A. Sprouse. 1996. L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12(1): 40–72.
- Selinker, Larry. 1972. Interlanguage. *International review of applied linguistics in language teaching* 10: 209–232. doi.org/10.1515/iral.1972.10.1-4.209.
- Sharwood Smith, Michael. 1983. Crosslinguistic aspects of second language acquisition. *Applied Linguistics* 4(3): 192–199.
- Sorace, Antonella. 1996. The use of acceptability judgments in second language acquisition. In *Handbook of second language acquisition*, edited by Tej K. Bhatia and William C. Ritchie, 375–409. New York: Academic Press.
- Sorace, Antonella. 2003. Near-nativeness. In *The handbook of second language acquisition*, edited by Catherine J. Doughty and Michael H. Long, 130–152. Oxford: Blackwell.
- Spinner, Patti, and Susan M. Gass. 2019. *Using judgments in second language acquisition research*. New York: Routledge.
- Sprouse, Jon, and Diogo Almeida. 2012. Assessing the reliability of textbook data in syntax: Adger's *Core syntax*. *Journal of Linguistics* 48(3): 609–652. doi.org/10.1017/S0022226712000011.
- Sprouse, Jon, and Diogo Almeida. 2017. Design sensitivity and statistical power in acceptability judgment experiments. *Glossa* 2(1): 1–32. doi.org/10.5334/gjgl.236.
- Stadt, Rosalinde, Aafke Hulk, and Petra Sleeman. 2016. The influence of L2 English and immersion education on L3 French in the Netherlands. *Linguistics in the Netherlands* 33(1): 152–165.
- Tokowicz, Natasha, and Brian MacWhinney. 2005. Implicit and explicit measures of sensitivity to violations in second language grammar. An event-related potential investigation. *Studies in Second Language Acquisition* 27(2): 173–204.
- Westergaard, Marit R. 2002. Minimal input in L2 acquisition: How Norwegian children learn English word order. Teaching and learning foreign languages. *Acta Didactica* 3: 199–218. Oslo: Unipub.
- Westergaard, Marit. 2003. Unlearning V2: Transfer, markedness, and the importance of input cues in the acquisition of word order in English by Norwegian children. *EUROSLA Yearbook* 3: 77–101.

Westergaard, Marit. 2021. Microvariation in multilingual situations: The importance of property-by-property acquisition. *Second Language Research* 37(3): 379–407.