

Author's version, accepted for publication 9/27/2015

**Appeared in *Linguistic Approaches to Bilingualism*, 7(2), 199-219.
DOI: 10.1075/lab.15011.gru**

Title: Coreference and discourse coherence in L2:
The roles of grammatical aspect and referential form

Authors: Theres Grüter¹
Hannah Rohde²
Amy J. Schafer¹

Affiliations: ¹University of Hawai'i at Mānoa
²University of Edinburgh

Address for correspondence:

Department of Second Language Studies
University of Hawai'i at Mānoa
1890 East-West Road, Moore Hall, rm570
Honolulu, HI 96822
U.S.A.

E-mail: theres@hawaii.edu

Abstract

Discourse-level factors, such as event structure and the form of referential expressions, play an important role in native speakers' referential processing. This paper presents an experiment with Japanese- and Korean-speaking learners of English, investigating the extent to which discourse-level biases that have gradient effects in L1 speakers are also implicated in L2 speakers' coreference choices. Results from a story continuation task indicate that biases involving referential form were remarkably similar for L1 and L2 speakers. In contrast, event structure, indicated by perfective versus imperfective aspect, had a more limited effect on L2 speakers' referential choices. The L2 results are discussed in light of existing accounts of L1 reference processing, which assume that referential choices are shaped by speakers' continually updated expectations about what is likely to be mentioned next, and argued to reflect L2 speakers' reduced reliance on expectations.

Pronouns are among the most frequently used words in English. Yet neither a dictionary nor a grammar book will satisfy a learner who is trying to master their use in successful and fluent communication. This reflects the fact that the use of referring expressions is not guided by categorical definitions or rules alone, but critically involves gradient preferences and constraints, often conditioned on information beyond the grammar proper, such as discourse context and inferential reasoning. For example, when encountering the pronoun *He* in (1), the fact that it is marked for masculine gender constitutes a categorical cue ruling out Sue as a potential antecedent. Yet both John and Bill remain viable candidates, and indeed, felicitous continuations in which *He* corefers with either of them are easy to construct (1a,b).

- (1) John loaned Bill Sue's old textbook. He...
- a. ...thought Bill could use it.
 - b. ...was grateful for John's thoughtfulness.

Thus both antecedents are *permitted*; the question that remains is whether one of them might be *preferred*. If so, what are the cues that contribute to such biases, and at what point in the course of the comprehender's incremental construction of meaning do those cues play a role?

Psycholinguistic research has investigated the interaction of multiple probabilistic cues in referential processing in some depth, and within a variety of theoretical and experimental paradigms (e.g., Arnold, 2001, 2010; Grosz, Joshi, & Weinstein, 1995; Hobbs, 1979; Kaiser, 2011; Kehler, 2002; Stevenson, Crawley, & Kleinman, 1994), yet

almost exclusively with native speakers. Independently, research in the field of Second Language Acquisition (SLA) has identified pronouns as a particularly vulnerable domain for non-native speakers (e.g., Sorace & Filiaci, 2006; Sorace, 2011). The SLA literature has focused primarily on the impact of specific structural factors, such as subjecthood of potential antecedents, on the interpretation or choice of different referential forms (e.g., null versus overt pronouns; Belletti, Bennati, & Sorace, 2007). While there can be no doubt that structural factors are important for reference resolution, a critical finding that has emerged from the literature on referential processing in native speakers is that these are not the only factors at work, and that such factors may in fact reflect higher-level constraints related to topicness within a coherent discourse. Numerous studies have shown that a variety of higher-level contextual factors, such as event semantics, causal reasoning, and inference, critically influence native speakers' coreference biases (Arnold, 2001; Caramazza, Grober, Garvey, & Yates, 1977; Kehler, Kertz, Rohde, & Elman, 2008; Stevenson et al., 2004; *inter alia*). Consider (2a) and (2b).

- (2) John loaned Bill Sue's old textbook. He...
- a. ...handed it over carefully.
 - b. ...opened it to the first page and started reading.

As in (1a) and (1b) above, we have no difficulty understanding that *He* most likely refers to John in (2a) and Bill in (2b). Yet unlike in (1), there is no explicit mention of the other potential referent that helps us disambiguate the reference of *He* in (2). In order to make sense of (2a), we need to understand that the continuation constitutes an *elaboration* of

the first sentence: it specifies *how* the textbook was transferred from John to Bill; since John is the ‘giver’ (the Source referent), we infer that he must be the agent in the handing-over event, and hence the subject in (2a). Similarly, we can infer that the opening event described in (2b) happened right *after* the book changed possession; since Bill is now in possession of the book (i.e., Bill is the Goal referent of the handing event), it makes most sense for him to be the one opening it to the first page, and hence the subject in (2b). Contextual factors such as these, involving discourse-level reasoning and world knowledge, have received little attention in the SLA literature on pronoun interpretation so far.

In this paper, we investigate one such factor, namely event structure, encoded by grammatical aspect. As we outline in more detail below, event structure has been shown to influence native speakers’ referential choices in subtle but consistent ways. Here we ask whether the same holds true for non-native speakers. We thus situate our study within a tradition of research on L1 reference processing that has emphasized the contribution of probabilistic expectations about how a discourse is likely to continue, in addition to any information structural factors (e.g., subject preference) that may be at work (e.g., Arnold, 2001; Kaiser, 2011; Kehler et al., 2008; Rohde & Kehler, 2014; Stevenson et al., 1994). We are not aware of any previous work on L2 reference processing within this tradition. Following much previous L1 work in this tradition, we adopt the story continuation paradigm (e.g., Arnold, 2001; Ferstl, Garnham, & Manouilidou, 2011; Stevenson et al., 1994), in which participants provide written continuations to discourse fragments similar to those in (1) and (2). This task requires both comprehension and production processes: participants first comprehend the presented discourse fragment, and then produce a

natural continuation. Their continuations provide an index of how they think the discourse is likely to continue and who will be mentioned next. Here we report findings from a story continuation experiment with Japanese- and Korean-speaking learners of English in which we manipulated aspect (perfective, imperfective) in the context sentence, as well as the referential form of the following subject (pronoun, free prompt), following an influential set of studies with native speakers (Rohde, Kehler, & Elman, 2006; Kehler et al., 2008; Ueno & Kehler, 2010). Our results indicate that biases involving the referential form of the subject of the continuation (prompt) were remarkably similar for native and non-native speakers, despite substantial differences in the inventories of pronominal forms in the learners' L1 compared to English. The manipulation of event structure, on the other hand, appeared to have a smaller effect on non-native compared to native speakers' referential choices. Drawing on existing accounts of what drives the influence of contextual factors like event structure on referential processing in a native language, we discuss what may account for the reduced effect of event structure in our L2 group, and outline potential implications of such an account for L2 processing more generally.

1. Event structure and referential form in L1 coreference processing

Tracking reference across sentences depends on a variety of factors—some related to properties of the available referents such as their grammatical or thematic roles (Arnold, 2001; Caramazza et al., 1977; McKoon, Greene, & Ratcliff, 1993; Stevenson et al., 1994), others related to properties of the unfolding discourse and the events being described (Hobbs, 1979; Kehler, 2002; Madden & Zwaan, 2003), and still others related

to the available referring expressions in a language's referential system and the felicity of those expressions for re-mentioning topical entities (Ariel, 1990; Gundel, Hedberg, & Zacharski, 1993; Grosz et al., 1995; Gatt, Krahmer, van Deemter, & van Gompel, 2014). The literature on L1 coreference processing is extensive, and beyond the scope of this paper to review comprehensively (see e.g., Arnold, 2010). We thus focus our discussion on previous coreference studies with L1 speakers that have investigated the specific cues of interest in our experiment: event structure and referential form of the anaphoric expression. Both cues influence L1 speakers' biases regarding who a sentence will be about, but they do so in different ways. Event structure is indicated by content in the preceding context and guides comprehenders' expectations about the relationship between that preceding sentence and the current sentence. Referential form is evident in the current sentence, at the point at which a referent is re-mentioned. In the study presented here, we manipulated event structure and referential form in order to address the question of how the use of these different types of cues may vary between L1 and L2 speakers.

Event structure can be encoded by various means, one of which is via a verb's grammatical aspect. Experiments that manipulate verb aspect have shown that the way that a context sentence portrays an event influences comprehenders' construction of situation models: Perfective aspect leads to the construal of an event as completed, with focus on its end state, whereas imperfective marked events are construed as ongoing (Moens & Steedman, 1988), with greater focus on individual participants in the event (e.g., Madden & Zwaan, 2003; Magliano & Schleich, 2000). These situation models in turn give rise to expectations about material in a subsequent continuation sentence

(Rohde et al., 2006; Ferretti, Rohde, Kehler, & Crutchley, 2009). Contexts that describe completed events favor continuations that tell what happened next (as in continuation (2b) above), whereas contexts that describe ongoing events favor continuations that provide elaborations or explanations of the event (as in continuation (2a)), or even information about what happened to prevent the event from being completed. These different continuation types establish different *coherence relations*, i.e., the relationships between propositions that must be inferred in order for a discourse to make sense (Asher & Lascarides, 2003; Hobbs, 1979; Kehler, 2002; Mann & Thompson, 1988). Importantly for our purposes, different coherence relations yield different patterns of coreference (Kehler et al., 2008; see also Arnold, 2001). The study reported here will follow prior work on event structure that has focused on transfer-of-possession contexts and manipulated the aspect of the verb, as in (3). Rohde et al. (2006) elicited story continuations following sentences like (3), which contains two potential referents for the pronoun *He* that begins the continuation. The Source referent is in the subject position of the context sentence; the Goal is the indirect object.

(3) John_{Source} handed/was handing a book to Bob_{Goal}. He _____

The results showed that continuations that described a subsequent event occasioned by or resulting from the first event (OCCASION and RESULT coherence relations) were more likely to re-mention the referent associated with the end state of the transfer event, i.e., the Goal. Other types of continuations (ELABORATIONS, EXPLANATIONS, VIOLATED EXPECTATIONS) were more likely to re-mention the referent associated with the start state

of the event, i.e., the Source (see Kehler et al., 2008, for discussion of the nature of these biases). We will refer to ‘Goal continuations’ and ‘Source continuations’ based on which referent is re-mentioned as the subject of the continuation. In line with the observation that different coreference biases are associated with different coherence relations, and that event structure influences the distribution of coherence relations, prior work has found a strong effect of event structure, marked by verb aspect, on the pattern of coreference, with more Goal continuations following perfective than imperfective contexts (Kehler et al., 2008; Ueno & Kehler, 2010, for Japanese; Kim, Grüter, & Schafer, 2013, for Korean). Verb aspect thus guides coreference biases via a fairly complex dynamic process which depends on the comprehender’s ability to build a mental model of the transfer event being described, postulate a coherence relation that connects the context sentence to a likely continuation, and incrementally update their estimates of which referent that continuation is likely to be about.

Referential form, on the other hand, has been shown to influence coreference biases via another discourse-level mechanism: information structure (Rohde & Kehler, 2014). Specifically, the presence of a *pronominal* form—even the ambiguous *He* in (3)—induces a preference for continuations about the subject referent more so than an unconstrained prompt. This follows from the fact that reduced forms, such as pronouns in English, are the preferred referential form for re-mentioning the topic of the discourse (e.g., Ariel, 1990; Gundel et al., 1993; Grosz et al., 1995). While subjecthood and topichood are by no means coextensive, they often align in English, especially in a short passage like (3), where the discourse topic is typically understood to be the subject of the context sentence. (Since our experiment is comprised of such short passages, we cannot

distinguish between subject- and topichood here; see Rohde & Kehler, 2014, for evidence that it is topichood, independent of semantic next-mention biases, that drives native speakers' choices for referential form.) Comprehenders who encounter a pronoun, even a fully ambiguous one, can thus posit a higher probability that the intended referent is the preceding subject (qua topic), as compared to their estimates prior to encountering the pronoun. As such, referential form only influences coreference biases via prompt type at the onset of the continuation.

2. The potential roles of event structure and referential form in L2 coreference processing

In the study presented here, we examine the roles of event structure (perfective, imperfective) and referential form (pronoun, free prompt) in guiding L1 and L2 speakers' referential choices in passages like (3). To what extent either of these factors should be expected to play a different role in L2 compared to L1 reference resolution is not immediately clear. On the one hand, reference resolution is a paradigmatic example of a linguistic phenomenon involving the interface between syntax- and discourse-level information. Such phenomena have been argued to lead to persistent and possibly permanent L1-L2 differences (Sorace, 2011; Sorace & Filiaci, 2006). Moreover, the cues we explore are gradient, and there are no categorical or explicit rules dictating how they should combine (unlike, for example, a cue about gender, as in (1) above). If the integration of multiple and gradient cues involving information sources from various (linguistic and non-linguistic) domains is a general weakness in L2 processing, we would expect weaker effects for both manipulations in our study in the L2 compared to the L1

group. On the other hand, coreference is one of the most fundamental and universal elements of comprehension since it determines who or what a discourse is about. If resources are spread thin during L2 processing, coreference may stand as one of the best candidates for resource allocation given its importance to understanding the speaker's message. L2 speakers also have abundant practice with coreference processing from their L1. If transferable L1 knowledge or skill plays a modulatory role in L2 processing, differences between L1 and L2 speakers may be diminished in this domain.

Let us also consider the potential difficulties associated with each manipulation separately. With regard to the event structure manipulation, it is important to bear in mind that the acquisition of aspect can be a difficult and protracted process in SLA (e.g., Bardovi-Harlig, 2000; Gabriele, 2009). It is thus possible that learners' incomplete acquisition of the semantics of grammatical aspect in English may limit their use of such a cue. In order to address — and rule out — this possibility here, we included an independent 'knowledge-of-aspect' task.

Another potential difficulty associated with the event structure cue is that the critical linguistic information – the perfective versus imperfective verb (*handed/was handing* in (3)) – is separated in the linguistic stream from the point at which a referential choice must be made (*He* __), i.e., in the subsequent sentence. Given that L2 processing is generally slower and less automatic (e.g., Kilborn, 1992; McLaughlin, 1987), it is possible that L2 speakers may have excellent understanding of the relationship between grammatical aspect and event structure, but do not carry out all of the steps of building an event structure as automatically and thoroughly as native speakers. By the time the event structure becomes relevant for the continuation, the specific cue to it – the grammatical

aspect information – would have been processed several words earlier, and may not be highly salient. Similarly, the L2 participants could construct a complete event structure, but it would have more time to decay, compared to that of a native speaker, if L2 participants spend more time comprehending the remainder of the sentence. Note that the relative timing with which the two cues of interest here are encountered in the discourse context differs: While the event structure cue is encountered in the preceding sentence, referential form is a cue available at the point of coreference interpretation, that is, when the participant encounters the prompt at the beginning of the continuation and must make a decision about who that continuation will be about. If memory access plays a critical role, we might thus expect a more native-like pattern for the referential form compared to the event structure manipulation in our L2 group. Alternatively, if all discourse-level cues present similar challenges for L2 speakers, we may see more limited L2 use of both the referential and event structural cues compared with L1 speakers.

A difference in the strength of effects for L1 and L2 speakers could also arise for somewhat different reasons. As hypothesized by Kehler et al. (2008), the influence of event structure on reference resolution results from a complex chain of probabilistic expectations about where a discourse is going. These expectations are assumed to be updated incrementally with each relevant cue as it is encountered. Underlying this account is a view of (native) language comprehension as a dynamic and forward-looking process, a view that aligns well with the abundant evidence of prediction in L1 processing (Altmann & Kamide, 1999; DeLong, Urbach, & Kutas, 2005; *inter alia*) – although there are open questions about the specific mechanisms underlying what are potentially a spectrum of processes involving linguistic pre-activation (DeLong, Troyer,

& Kutas, 2014). Meanwhile, the role of prediction, or linguistic pre-activation more generally in L2 processing is less clear (see Kaan, 2014, for discussion). If L2 comprehension relies less on the dynamic, forward-looking processes assumed by Kehler and colleagues to be responsible for the effects of event structure on reference resolution in L1, this should also manifest as a diminished effect of this manipulation in our L2 group. We return to these possibilities in the Discussion.

Finally, with regard to the referential form manipulation, we must consider crosslinguistic differences in the inventories of pronominal forms and form-meaning associations. In languages that have both null and overt pronouns, different forms may be associated with different referential preferences (e.g., previous subject vs non-subject; see Carminati, 2002, for Italian). Such form-dependent preferences have shown to be difficult for non-native speakers to acquire (see Sorace, 2011, for review). Moreover, preferences associated with null and overt forms in a learner's native language can influence their interpretation of referential forms in a second language (Roberts, Gullberg, & Indefrey, 2008). The L2 learners in our study are native speakers of Japanese or Korean, both null subject languages. In both Japanese and Korean, a null subject is frequently used to refer to a continued topic (Hwang, 1983; Kim, 1997; Obana, 2003). Both languages also have overt pronouns, yet they are used less frequently and in more restricted contexts (e.g., written and more formal genres), and it remains unclear whether the division-of-labor account proposed by Carminati (2002) for Italian extends to these East Asian languages. Thus given the different status of overt pronouns in Japanese and Korean compared to English, the Japanese- and Korean-speaking learners of English in our study may well treat the overt pronoun prompt in English differently than native English speakers.

Specifically, for Japanese- and Korean-speaking learners, unlike what has been observed for native English speakers, an overt pronominal form may not induce (as strong) a preference for a continuation about a previous subject, leading to a weaker effect of the referential form manipulation in the L2 group.

While an account focusing on crosslinguistic differences may predict a weaker effect of the referential form manipulation in the L2 group, such an account predicts no difficulties for the event structure manipulation: Both Japanese and Korean mark verbal aspect grammatically. Moreover, and critically for our purposes, two recent studies confirm the effect of aspect on coreference biases in transfer-of-possession contexts discussed above in Japanese and Korean: Ueno and Kehler (2010) reported that Japanese (native) speakers produced more Source continuations following imperfective than perfective context sentences; Kim et al. (2013) reported a similar effect of aspect for Korean (native) speakers. Thus given that Japanese and Korean speakers show aspect-driven coreference effects in their L1, if transfer is at work in this domain, it could only serve to *minimize* L1-L2 differences for the event-structure manipulation in our experiment. We now turn to the experiment.

3. Method

3.1 Participants

A total of 87 participants, recruited from the University of Hawai‘i community, took part in this study for either course credit or US\$10. Thirty-nine participants were L1 speakers of English (L1 group) and 48 were L2 speakers whose L1 was either Japanese ($N=23$) or Korean ($N=25$). Participants in the L2 group were primarily international or

exchange students at the time of testing. They were first exposed to English between the ages of 8 and 13 years in school in Japan/Korea; their length of exposure to English in the U.S. varied considerably, yet most had spent less than one year in the U.S. at the time of testing (median = 6 months).

Three measures of English language proficiency were collected from L2 participants: (i) performance on a written cloze test (Brown, 1980), (ii) self-ratings of English language ability, and (iii) performance on the Versant English Test, a commercially available assessment of oral fluency (Pearson, 2011). L1 participants completed only (i) and (ii) given previous evidence showing that native speakers typically perform at ceiling on (iii) (Pearson, 2011). Descriptive statistics of participants' performance on these proficiency measures are summarized in Table 1. Scores were significantly higher in the L1 than in the L2 group on the cloze test ($t(85)=11.8, p<.001$) as well as on self-ratings (overall and subskills, all $U>36, p<.001$).¹ The Japanese and Korean L2 learners did not differ in their cloze test scores ($t(46)<1$) or self-ratings (all $t(46)<1.3$); on the Versant Test, the Korean subgroup obtained higher scores ($U=192, p=.048$). Based on the alignment of Versant scores with the general level descriptors of the Council of Europe framework (Pearson, 2011), the majority of L2 participants in this study fall into the categories of Independent Users (B1, B2; $N=27$) or advanced Basic Users (A2; $N=19$) of English. In order to include proficiency as a predictor in our statistical models, a single proficiency score for each L2 participant was calculated by converting scores from the Versant and Cloze tests to z -scores and averaging the two.

¹ We report t statistics for comparisons in normally distributed data, and non-parametric Mann-Whitney U for non-normal distributions.

This combined z-score measure significantly correlated with overall self-ratings ($r(48)=.48, p=.001$), and did not differ significantly between the Japanese ($M= -.19, SD=.66$) and the Korean subgroups ($M= .17, SD=.99; t(46)=1.49, p=.14$).

Table 1

Performance on proficiency measures (means and ranges)

	<i>Age</i> (in years)	<i>Cloze test</i> ¹ (proportion acceptable responses)	<i>Versant</i> <i>English Test</i> ² (overall score, range 20-80)	<i>Self-rated</i> <i>overall</i> <i>English</i> <i>proficiency</i> (out of 10)
L1 English (n=39)	24 (18-66)	0.84 (.60-.98)	--	9.3 (7-10)
L2 English (n=48)	24 (18-51)	0.55 (.24-.80)	51 (34-80)	6.0 (2-9)
L1 Japanese (n=23)	25 (18-51)	0.54 (.36-.68)	49 (40-61)	6.2 (4-9)
L1 Korean (n=25)	23 (20-32)	0.56 (.24-.80)	53 (34-80)	5.8 (2-8)

¹Brown (1980), ²Pearson (2011)

3.2 Materials and procedure

Participants completed two experimental tasks: a written story continuation task, adapted from Rohde et al. (2006), and the knowledge-of-aspect task, which was designed to assess participants' understanding of verb aspect in English. The story continuation task is the measure of interest, addressing our key research question: Do L2 speakers show sensitivity to manipulations of event structure and referential form in the choices they make in their continuations? For outcomes from this task to be interpreted meaningfully, however, independent evidence is needed to establish that participants have the necessary linguistic knowledge that underlies the event structure manipulation, namely that they understand the basic semantics of grammatical aspect in English: Perfective denotes completed events, whereas imperfective describes ongoing or incomplete events. The knowledge-of-aspect task was included to establish whether participants have this knowledge.

Participants completed a language background questionnaire, including the self-rating of their English language ability reported above, prior to the experimental session in the lab. During a 60-to-90-minute experimental session, the story continuation task was administered first, followed by the written cloze test, and then the knowledge-of-aspect task. Following this session, L2 participants also completed the Versant English Test (20 minutes).

3.2.1 Story continuation task

The experiment employed a 2×2 design, varying grammatical aspect of the verb in the context sentence (perfective/imperfective) and referential form in the continuation

prompt (pronoun/free), as in (4). In the pronoun-prompt condition, a pronoun, always referentially ambiguous between the two individuals mentioned in the context sentence, was provided as the first word of the continuation. In the free-prompt condition, there were no restrictions on the continuation.

- (4) a. Emily brought a drink to Melissa. (She) _____ [perfective]
b. Emily was bringing a drink to Melissa. (She) _____ [imperfective]

Twenty critical items with transfer-of-possession verbs (5 per condition) and 20 fillers appeared in semi-random order in each of four lists, counterbalanced for experimental condition. The critical items consisted of 10 transfer-of-possession verbs (*bring, feed, give, mail, pass, push, roll, serve, take, throw*), each presented twice, with different Source/Goal names and theme objects and with different verb aspect. The list of experimental stimuli is provided in Appendix 1. As in (4), the Source of the transfer-of-possession event was always the syntactic subject of the context sentence, and the Goal was the indirect object. The fillers contained non-transfer-of-possession events involving one or two individuals. Half of the fillers were presented with a prompt for the continuation, consisting either of a pronoun or an adverb; the other half imposed no restrictions. The task was presented through a web-based interface. On each trial, a context sentence appeared on the screen with or without the subsequent sentence-initial pronoun prompt, followed by a text box. Participants were instructed to imagine a natural story continuation for each context sentence, and type the first continuation that came to mind, avoiding humor. The context sentence (and prompt) remained on the screen while

the participant typed the continuation. A continuation was required for each story before the participant could move on to the next trial, and no backtracking was permitted.

3.2.2 Knowledge-of-aspect task

The goal of this task was to assess whether L2 participants consistently associate perfective and imperfective aspect with complete and ongoing events respectively. In English, the interpretation of the imperfective *be -ing* as ongoing occurs with verbs of all event classes. The same is the case for the the imperfective marker *-ko iss-* in Korean (Lee, 2006). In Japanese, however, the imperfective marker *-te i-* denotes ongoing events when combined with most verbs, yet yields a resultative reading with others, specifically achievement verbs (Gabriele, 2009; Shirai, 2000). Prior work on aspect in L2 acquisition has shown that Japanese learners of English, even at advanced levels of proficiency, do not consistently rule out resultative interpretations of imperfective-marked achievement verbs in English (Gabriele, 2005, 2009). Given that transfer-of-possession verbs are most often classified as achievement verbs, a null effect of aspect for the L2 group in the story continuation task could thus be attributed to learners' deriving the same resultative interpretations from the context sentences in both the perfective and imperfective condition, that is, not distinguishing between the two. The knowledge-of-aspect task, inspired by a story compatibility task originally designed by Gabriele (2005), was included so that this possibility could be ruled out.

Participants read stories describing events that were either complete or ongoing. Following the story, they were asked to judge the truth of a (written) test sentence uttered at a particular point in time by an observer (the cartoon character Pikachu), as

exemplified in (5). The story remained visible on the screen until a judgment was made by clicking on one of three radio buttons labeled *true*, *false* or *not sure*.

(5) *Story beginning:*

Brenda is at the hospital visiting Anne. [picture of soup]

This is the bowl of soup that Brenda will feed to Anne.

At 11:00, Brenda is ready with the soup and a spoon.

Story end, completed condition:

At 12:00, the bowl is empty and Anne wipes her mouth.

Story end, ongoing condition:

At 11:05, Brenda puts the first spoonful of soup into Anne's mouth.

Test sentence:

At [12:00/11:05], Pikachu says: "Brenda is feeding the bowl of soup to Anne."

For native English speakers, the test sentences with imperfective-marked verbs are false in the completed condition and true in the ongoing condition.² Critically, if learners – as a

² Note that interpretation of the story requires the reader to make inferences about what is likely to have happened between the Story beginning and the Story end. Readers who are inattentive or draw unexpected inferences may thus not converge on the expected truth value judgments in all cases, and uniform performance may not necessarily be observed, even in the L1 group. What is of interest here is whether participants' judgments differ significantly between the two conditions.

possible effect of L1 transfer – derive a resultative reading of this sentence, this should lead to an increased rate of ‘true’ judgments in the completed condition, as well as a potentially decreased rate of ‘true’ judgments in the ongoing condition. The task consisted of 10 experimental items with an imperfective-marked transfer-of-possession verb following a story in which the transfer was portrayed as completed (k=5; Condition 1) or ongoing (k=5; Condition 2), as illustrated in (5). The transfer-of-possession verbs were the same as those used in the story continuation task. An additional 12 items were included as controls to ensure that L1 and L2 speakers respond similarly in this task when no relevant interpretive differences exist crosslinguistically. Control items consisted of perfective-marked achievement verbs following a completed (k=4; true) or ongoing (k=4; false) event, as well as imperfective-marked accomplishment verbs following a completed event (k=4; false). Participants always completed this task after the story continuation task, in one of four semi-randomly ordered lists. Lists were counterbalanced for the presentation of verbs in Conditions 1 vs 2.

3.3 Results

We begin by reporting the results from the knowledge-of-aspect task, as this task constitutes a critical prerequisite for the interpretation of the results from the story continuation task.

3.3.1 Knowledge-of-aspect task

‘Not sure’ responses were excluded from the analysis as they were rare in both groups (L1: 3.1%, L2: 5.4%). Participants’ mean percentage of ‘true’ judgments in the

two critical conditions is illustrated in Figure 1. Participants in all groups generally judged sentences with imperfective-marked transfer-of-possession verbs true in the case of ongoing events (Condition 2), and false in the case of completed events (Condition 1), indicating that participants in all groups generally derived a progressive interpretation for imperfective-marked transfer-of-possession verbs in English. In order to probe for potential effects of L1 transfer regarding resultative interpretations, L2 participants' performance in the two critical conditions was analysed using mixed-effects logistic regression models, implemented in R using the lme4 package. Separate models were created for Japanese and Korean speakers, with the truth value judgment (true/false) modeled as a binary outcome and Condition (1 vs 2) as the fixed-effect predictor with deviation coding (-.5, .5; Barr, 2013). The best fitting models were those with maximal random effects structure (i.e., random intercepts and slopes for participants and items). The use of mixed-effects models permits the inclusion of terms to account for both participant- and item-variability within a single model (see Jaeger, 2008, on the advantages of such models over separate by-participant and by-item ANOVAs for data like this). These models indicate that both Japanese- and Korean-speaking learners were significantly more likely to judge as true a sentence with an imperfective-marked transfer-of-possession verb describing an ongoing event than a completed event (Japanese: $b=4.09$, $p<.001$; Korean: $b=4.64$, $p<.001$). This indicates that, as a group, the L2 learners in this study understand the interpretive consequences of manipulating verbal aspect with English transfer-of-possession verbs.

Nevertheless, closer inspection of participants' performance in Condition 1 suggests that the frequency of 'true' judgments was somewhat elevated in the L2 groups

compared to the L1 group. We thus fit a mixed-effects logistic regression model of the judgments in Condition 1 alone, with L1 (English, Japanese, Korean) as the fixed-effect predictor and ‘English’ set as the reference level. The best fitting model contained random intercepts for participants and items. This model indicates that L2 learners in both subgroups (Japanese and Korean) were significantly more likely than native English speakers to judge sentences in Condition 1 true (Japanese vs English: $b=1.57, p=.001$; Korean vs English: $b=1.08, p=.03$). In order to better understand whether these differences are attributable to a few L2 learners (in each subgroup) deriving a resultative interpretation of imperfective-marked achievement verbs, individual response profiles in Condition 1 were inspected. A total of 17 participants (10 Japanese, 4 Korean, 3 English) were found to have provided the expected ‘false’ judgment on only 3 or fewer of the 5 items in this condition. We think the most likely explanation of these profiles is that they reflect factors other than aspect, such as occasional lapses in attention, especially in the native English speakers. Recall, however, that the semantics of the Japanese progressive marker is more different from English imperfective than the semantics of the Korean progressive marker. The somewhat higher proportion of Japanese than Korean participants who showed this profile could thus be reflective of incomplete understanding of the semantics of aspect in English in at least some of these participants. Therefore, on the conservative assumption that this profile potentially reflects incomplete understanding or atypical use of the semantics of aspect, we conducted all analyses of the data from the story continuation task on both the full dataset ($N=87$) and on a more conservative reduced dataset excluding these 17 individuals ($N=70$). In the conservative

dataset, there were no statistically significant differences between L1 and L2 participants in Condition 1 (L1 vs L2: $b=0.55, p=.28$).

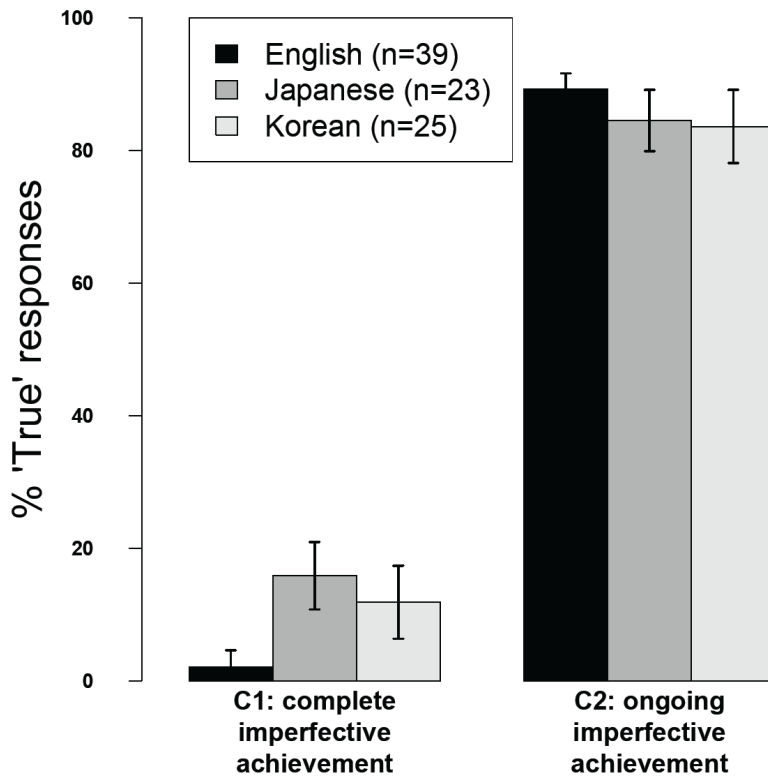


Figure 1. Mean percentage of ‘true’ judgments by condition and group. Error bars represent 95% confidence intervals of the means, adjusted for the repeated-measures design (Bakeman & MacArthur, 1996).

3.3.2 *Story continuation task*

We now turn to our principal research question regarding coreference processing:

How does the use of event structure and referential form vary between L1 and L2 speakers’ story continuations? To answer this question we considered three types of

choices that participants made in the story continuation task: first, who the continuation was about, as assessed by whether the referent of the syntactic subject of the continuation was the Source or the Goal of the context sentence event (e.g., *Emily* versus *Melissa*, in (6) below); second, what type of referential expression was used to express that syntactic subject in the free prompt conditions (e.g., *she* versus a name); and third, what coherence relation could be inferred to hold between the context sentence that described the transfer event and the participant's story continuation (e.g., Result versus Occasion). Although these choices are closely linked to each other, we begin by considering them individually. We then return to the relationships among them, and how those might differ for L1 versus L2 speakers, in the Discussion section.

Our coding of the continuations followed protocols designed by Rohde and colleagues in earlier work and thus allows comparison with related work testing speakers of English, Japanese, or Korean in their native language. Two trained judges worked separately to annotate continuations for each of the measures of interest. Disagreements among the coders were resolved by a third coder (the second author). Details of the coding for each dependent measure are provided below in conjunction with the results for that measure. The examples in (6) are taken from the participants' continuations and show samples of the annotators' coding.

- (6) Context: Emily brought/was bringing a drink to Melissa. (She) _____
- a. She gave her Coke. (SOURCE; PRONOUN; ELABORATION)
 - b. She thought Melissa was thirsty. (SOURCE; PRONOUN; EXPLANATION)
 - c. Emily dropped it on the ground. (SOURCE; NAME; VIOLATED EXPECTATION)

- d. Melissa drank it. (GOAL; NAME; OCCASION)
- e. Melissa said "Thank you." (GOAL; NAME; RESULT)
- f. Melissa brought a snack to share. (GOAL; NAME; PARALLEL)
- c. She did not want it. (AMBIGUOUS)
- d. It was Coke. (OTHER)

As with the knowledge-of-aspect task, all data was analyzed using mixed-effects logistic regression models, using the 1.1.7 version of the lme4 package in R 3.0.2. We modeled reference choice as a binary outcome (Source versus Goal), likewise for referential form (pronoun versus name). The multiple categorical outcome of coherence relation was simplified to a binary distinction based on the compatibility of the relation with two different event structures (completed-event-driven versus ongoing-event-driven). Binary fixed factors were centered using deviation coding (-.5, .5) in order to make interactions interpretable (Barr, 2013). Random intercepts were included for participants and items, and random slopes were included for aspect, prompt, and group when appropriate and permitted by the data. Below, we report coefficients (*b*) for all factors of interest and their interactions and *p* values obtained via model comparison. All analyses were conducted on data from 20 critical items, with 87 participants in the full dataset and 70 participants in the conservative data set.

Coreference

Continuations were coded as ‘Source’, ‘Goal’, ‘Ambiguous’, or ‘Other’, depending on the coders’ assessment of the continuation’s syntactic subject. Coders were

instructed to classify responses as ‘Source’ or ‘Goal’ only if the continuations began with a main clause and its syntactic subject most plausibly referred to the Source or Goal of the previous event. Responses were deemed ‘Ambiguous’ if both judges indicated that the continuation was ambiguous or if one judge assigned a classification of Source while the other assigned a classification of Goal; these accounted for 3.7% of the data in the L1 group, and 4.0% in the L2 group. The category ‘Other’ was used for continuations that began with subordinate clauses, had subjects expressing joint reference to the Source and Goal, or had subjects that referred to some other entity such as the theme argument. These constituted 12.3% of the L1 data and 13.0% of the L2 data. An additional 0.3% of the L1 data and 1.3% of the L2 data was excluded due to ambiguity in the coherence relation, so that both the coreference and coherence analyses we report reflect the same core dataset. Ambiguous and Other reference choices were excluded from further analysis for our primary models, which, like previous work, report outcomes for Source responses out of the combined total of Source and Goal responses.³

The coreference results are illustrated in Figure 2, and critical values from the mixed-effects models are summarized in Tables 2 and 3. There was a main effect of aspect, a main effect of group, and, for the full dataset, a reliable interaction of aspect and group (Table 2). With Source responses as the outcome measure and Other reference choices excluded from the denominator, the interaction between aspect and group was

³ Because Other responses were more common in the free prompt conditions, an additional set of analyses retained Other responses and analyzed patterns of Source and Goal choices out of all unambiguous responses. These analyses showed effects very similar to those found in our primary analyses.

not reliable in the conservative dataset ($p=.14$); however, this interaction reached full significance for both full and conservative datasets in analyses that included Other responses and used Goal responses as the outcome variable. Follow-up analyses within each group thus seemed advisable.

As shown in Table 3 for the full dataset, and Table 4 for the conservative dataset, the effect of aspect was robust in the L1 group but non-significant in the L2 group. Thus, although L2 speakers understand the relationship between aspect and completed events, as indicated by their performance on the knowledge-of-aspect task, this knowledge does not translate into significant differences in coreference for completed versus ongoing events. As in previous work with native speakers, a main effect of prompt type was also observed, which did not interact with group. Follow-up analyses within each group indicated that the predicted prompt effect was robust in both the L1 group and the L2 group, indicating that L2 speakers patterned like L1 speakers in producing more Source continuations in the pronoun conditions than the free prompt conditions. Finally, a main effect of group indicated that L2 speakers produced more Goal responses overall than L1 speakers, a finding discussed further below.

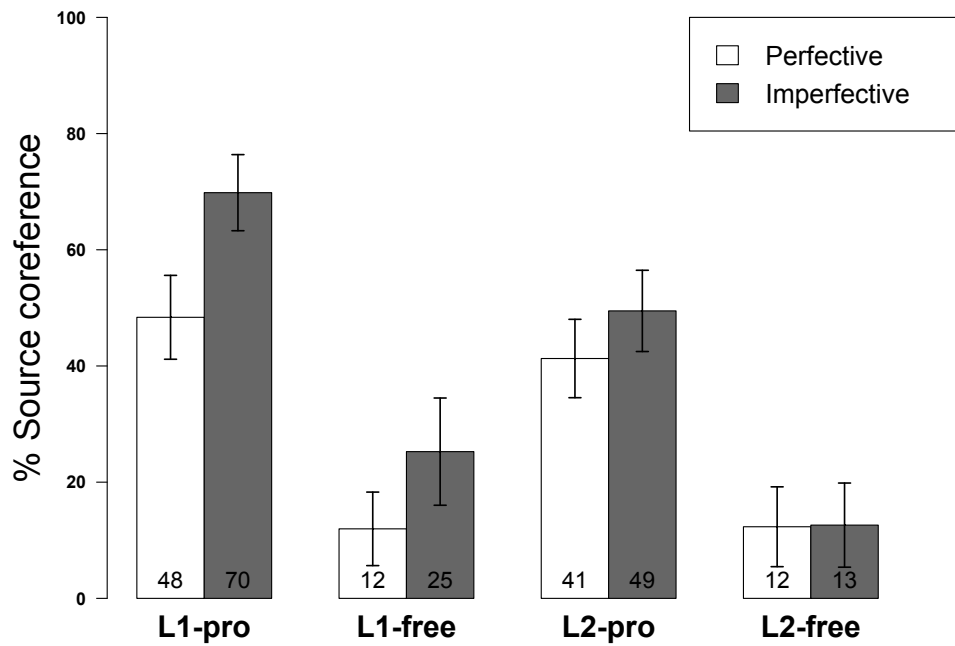


Figure 2. % Source reference by aspect, prompt type, and group. Error bars indicate 95% confidence intervals of the means adjusted for repeated measures.

Table 2

Results of mixed-effect logistic models for coreference

	Full (N=87)		Conservative (N=70)	
	<i>b</i>	p	<i>b</i>	p
Aspect	0.733	<0.001	0.736	<0.005
Prompt	-2.521	<0.001	-2.767	<0.001
Group (L1 vs. L2)	-0.703	<0.01	-0.823	<0.01
Aspect × Prompt	-0.243	0.51	-0.053	0.90
Aspect × Group	-0.890	<0.05	-0.647	0.14
Prompt × Group	0.142	0.78	-0.263	0.67
Aspect × Prompt x Group	-0.365	0.58	0.373	0.64

Note. Significant effects are in boldface.

Table 3

Reference choice models for each group (full dataset)

	L1 (N=39)		L2 (N=48)	
	<i>b</i>	p	<i>b</i>	p
Aspect	0.998	<0.05	1.074	0.14
Prompt	-2.852	<0.001	-3.296	<0.001
Aspect × Prompt	-0.365	0.65	1.176	0.42

Note. Significant effects are in boldface.

Table 4

Reference choice model for each group (conservative dataset)

	L1 (N=36)		L2 (N=34)	
	<i>b</i>	<i>p</i>	<i>b</i>	<i>p</i>
Aspect	1.021	<0.05	1.485	0.14
Prompt	-3.012	<0.001	-4.078	<0.001
Aspect × Prompt	-0.159	0.86	2.359	0.24

Note. Significant effects are in boldface.

In order to probe for potential effects of native language and proficiency, additional analyses were conducted on the full data from the L2 group by adding (i) L1 (Japanese vs. Korean) as a categorical predictor, and (ii) Proficiency (combined *z*-score, see p.14 above) as a continuous predictor. In a model including L1 as an additional fixed effect, the main effect of L1 was marginally significant ($b=.735$, $p=.06$), reflecting a higher overall proportion of Source responses among Japanese ($M=.35$, $SD=.19$) compared to Korean ($M=.28$, $SD=.16$) learners. Critically, however, L1 did not interact with Aspect ($b=.143$, $p=.85$) or Prompt type ($b=.611$, $p=.45$). Adding L1 did not improve overall model fit (as indicated by model comparison using the `anova()` function in R; $p=.39$). In a model with Proficiency as an additional fixed effect, Proficiency was not a significant predictor ($b=.120$, $p=.63$), nor did Proficiency interact with Aspect ($b=.587$, $p=.17$) or Prompt type ($b=.751$, $p=.10$). Adding Proficiency did not improve overall

model fit ($p=.16$). These additional analyses indicate that the effects reported above were not significantly modulated by proficiency or learners' native language.

Choice of referring expression

All data with Source or Goal subject reference was coded as 'Pronoun' or 'Name' to categorize its referential form. Because the syntactic subjects were necessarily pronouns in the pronoun prompt conditions, we only consider data from the free prompt conditions. Previous work has shown that when speakers re-mention a referent that has appeared as the subject of the previous clause, they favor pronominal forms, whereas non-subject referents are re-mentioned preferentially with more explicit referring expressions like names (e.g., Stevenson et al., 1994; Arnold, 2001; Miltsakaki, 2007; Kehler et al., 2008; Fukumura & van Gompel, 2010; Rohde & Kehler, 2014). We therefore anticipated that Source continuations would contain more pronouns and Goal continuations would contain more names, for both participant groups. As predicted, both L1 and L2 participants produced pronouns at a higher rate when referring back to the (subject-position) Source (L1: 65% pronouns: 34 pronouns out of 52 Source reference responses; L2: 81%, 30/37) than to the (non-subject-position) Goal (L1: 3%, 8/237; L2: 21%, 64/308). (Due to the unbalanced distribution of data across cells, no statistical analysis was performed.) The knowledge that L1 and L2 participants show in the free prompt conditions regarding the differential rates of pronominalization based on the grammatical role of the antecedent can be taken to underpin the main effect of prompt type: As shown in Figure 2, the presence of a pronoun form in the pronoun prompt conditions induced more Source continuations in these conditions compared to the free

prompt ones in both groups. These results lend support to the observation that L2 speakers can match L1 speakers' use of the discourse-level interpretive biases associated with different referential forms.

Coherence relation

The final dependent measure we examined was the type of coherence relation that connected the context sentence to the continuation. Analysis was restricted to continuations with Source or Goal reference. Relations were first classified as one of six categories, following the annotation in Kehler et al. (2008): Elaboration, Explanation, Violated Expectation, Occasion, Result, and Parallel. For continuations which plausibly could be categorized as either Occasion (requiring merely the forward progression of time) or Result (requiring, in addition to forward progression, that the event in the continuation arise causally from the event in the context sentence), the more specific causal interpretation was assigned, for greater uniformity and specificity in coding. Data that did not clearly fall into one of these categories was coded as 'Ambiguous' and, as reported earlier, was excluded from all analyses. Here we also exclude Parallel since the instances were so rare (14 for L1, 25 for L2).

The distribution of coherence relations in both groups is illustrated in Figure 3, with results from the L1 group layered behind those from the L2 group, so that the two groups can be easily compared. Overall, the percentages for each relation are strikingly similar in the two groups, as are the percentages of Source/Goal choices within each relation. Most notably, both groups showed sensitivity to the aspect manipulation: Perfective context sentences (bars in Figure 3 marked with "-p") were more likely to

yield relations whose inference focuses on the end state ('Completed-event-driven' relations: Occasion, Result) whereas imperfective context sentences favored relations whose inference focuses on the start state ('Ongoing-event-driven' relations: Elaboration, Explanation, Violated Expectation).

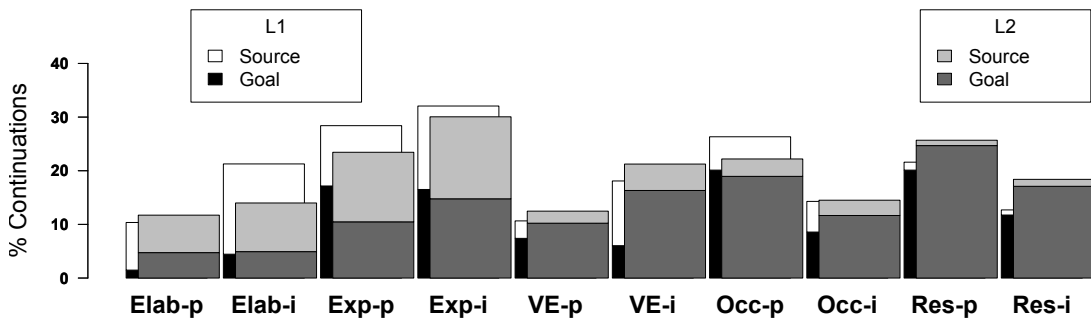


Figure 3. Distribution of coherence relations (Elaboration, Explanation, Violated Expectation, Occasion, Result) by group (L1, L2) and aspect (-p, -i) of the context sentence. (To reduce visual complexity, this graph collapses across the prompt manipulation.) The height of the bars indicates the frequency of each coherence relation within each group by aspect subset (e.g., % Elab-p refers to the proportion of Elaborations out of all continuations following perfective aspect). The shading within each bar shows the relative proportion of Source and Goal continuations instantiated within that bar.

In order to assess whether our factors of interest, event structure and referential form, would differentially affect the choice of coherence relation in the L2 versus the L1

group, we modeled the choice of coherence relation as a binary outcome (Completed-event-driven vs Ongoing-event-driven) with the same three fixed-effect factors as above (aspect, prompt, group). This grouping of coherence relations follows that in Ueno & Kehler (2010), with the addition of Violated Expectation in the Ongoing-event-driven category given its typical use in describing an unexpected outcome that prevented the transfer event from being completed. As in the coreference analysis reported above, we also conducted follow-up analyses within each group to assess whether aspect and prompt had significant effects on the production of our two groupings of coherence relations.

Tables 5 and 6 present the results of these analyses. The statistical models indicated significant main effects of aspect and prompt in the expected direction: Participants chose more Completed-event-driven coherence relations with perfective context sentences and with free prompts. (The latter effect is not depicted in Figure 3, which collapses across prompt conditions to reduce visual complexity but it is in keeping with coherence results reported in Kehler & Rohde, under review.) In contrast to the reference choice results, there was no effect of group, nor were there any reliable interactions between group and the other factors, including aspect. Follow-up analyses within each group indicated that the aspect and prompt effects were present in both the L1 group and the L2 group (Table 6). These results indicate that L2 speakers, like L1 speakers, show striking sophistication in their ability to use event structure and referential form in their decisions about the relationships that hold between sentences in a discourse. The L2 speakers' use of aspect specifically demonstrates the ability to take an event structural cue into consideration – at some point in the process of continuing the story –

in order to establish how an upcoming sentence will relate to the current sentence. We consider the implications of the L2 speakers' significant use of aspect for coherence decisions versus their reduced use of it for referential decisions in the Discussion.

Table 5

Results of mixed-effect logistic models for coherence relation

	Full (N=87)		Conservative (N=70)	
	<i>b</i>	p	<i>b</i>	p
Aspect	0.998	<0.001	1.039	<0.005
Prompt	-0.573	<0.001	-0.502	<0.001
Group (L1 vs. L2)	-0.245	0.39	-0.247	0.43
Aspect × Prompt	-0.180	0.50	-0.116	0.69
Aspect × Group	-0.345	0.26	-0.101	0.76
Prompt × Group	0.357	0.20	0.346	0.25
Aspect × Prompt x Group	0.734	0.17	1.027	0.08

Note. Significant effects are in boldface.

Table 6

Coherence relation models for each group (full dataset)⁴

	L1 (N=39)		L2 (N=48)	
	<i>b</i>	<i>p</i>	<i>b</i>	<i>p</i>
Aspect	1.396	<0.001	0.778	<0.005
Prompt	-0.789	<0.01	-0.414	<0.05
Aspect × Prompt	-0.628	0.17	0.173	0.62

Note. Significant effects are in boldface.

4. Discussion and Conclusion

The goal of this study was to explore the contribution of two discourse-level factors, event structure and referential form, to non-native speakers' referential choices. To this end, we employed an experimental paradigm that had been used in previous research with native speakers to investigate probabilistic biases that affect speakers' coreference processing across sentences. In our analyses of participants' written story continuations, we asked whether non-native speakers, like native speakers, would show

⁴ The pattern of results is the same in the conservative subset of the L1 group (N=36): main effects of aspect ($b=1.236, p<.001$) and prompt ($b=-0.676, p<.05$), interaction not significant ($b=-0.644, p=0.16$). The conservative subset of the L2 group (N=34) shows the same main effect of aspect ($b=0.942, p<.01$) and the same lack of a prompt × aspect interaction ($b=.361, p=.39$), but the main effect of prompt type is not significant ($b=-.372, p=.12$).

sensitivity to manipulations of event structure and referential form in the choices they made in their continuations. Choices of three kinds were analysed: (i) the referent of the syntactic subject of the continuation, (ii) the type of referential expression chosen for that syntactic subject (when there was a choice), and (iii) the coherence relation between the given context sentence and the continuation produced by the participant.

With regard to the referential form manipulation, our results showed significant effects for all three choices, and no substantial differences between the L1 and the L2 groups: Speakers in both groups wrote continuations with a subject referring back to the Source of the context sentence significantly more often when they were given a pronoun prompt than when they were free to choose any referential expression to start their continuation. (Recall that our study was not designed to distinguish between thematic role, grammatical role and topicality. Thus Source here is conflated with both subjecthood and topic status, both of which are likely to contribute to its status as a preferred antecedent for a pronoun.) Moreover, in the free-prompt condition, when they decided to start their continuation with a subject referring back to the Source argument, participants in both groups were more likely to choose a pronoun (vs. a name) than when they started their continuation with a subject referring to the previous Goal. Finally, participants in both groups were more likely to choose an ongoing-event driven coherence relation (Explanation, Elaboration, Violated Expectation) when they were given a pronoun prompt. All of these effects were present in both the L1 and the L2 group, and none of them interacted with group as a factor. The non-native speakers in this study thus showed remarkable awareness of subtle interpretive biases associated with pronouns in English. This is particularly noteworthy given the different status of overt

pronouns in the inventory of referential expressions in their L1s. Although the interpretive preferences associated with (rarely used) overt pronouns in Japanese and Korean are not yet fully understood, we found no evidence in this study that L2 learners of English whose L1 is a null-subject language have long-lasting difficulties converging on interpretive preferences associated with subject pronouns in English.

Turning to the event structure manipulation, our results from the L1 group replicated the effect observed in previous work, showing that native speakers were more likely to start their continuation with a subject referring back to the Source argument if the event in the previous sentence was incomplete as indicated by imperfective aspect. They were also more likely to choose an ongoing-event driven coherence relation following imperfective aspect, consistent with previous findings and with expectation-based accounts of reference processing that propose a close relation between coreference and coherence choices (Hobbs, 1979; Kehler, 2002; Kehler et al., 2008). The results from the L2 group were more complex. Our analysis of coreference choices showed an interaction between aspect and group, with follow-up analyses indicating that the aspect effect was robust only in the L1 group, and did not reach significance in the L2 group (regardless of whether the analysis was based on the full or the conservative dataset). Importantly, the L2 group, especially the conservative subset, performed well on an independent task assessing knowledge of aspect on the same transfer-of-possession verbs used in the story continuation task. Thus the absence of an effect of aspect in the L2 group in the story continuation task is unlikely to be attributable simply to insufficient understanding of grammatical aspect in English. Additional analyses further indicated that neither learners' L1 nor their proficiency interacted with the effect of aspect. While

we had no principled reasons to expect Japanese- and Korean-speaking learners to perform differently, the observation that proficiency did not play a more important role was somewhat surprising given ample evidence for the role of proficiency both with regard to sensitivity to discourse-level properties in L2 (e.g., Hopp, 2009) and facilitation in sentence-level L2 processing (e.g., Chambers & Cooke, 2009). It is possible that the range of L2 proficiency represented by participants in our L2 group, in particular the sparsity of near-native L2 speakers, did not allow us to detect an effect of proficiency in this particular study.

Further support for eliminating insufficient knowledge of aspect as an explanation for the reduced effect of aspect in the L2 group comes from the results of L2 participants' coherence choices. Here we found that, just like the L1 group, L2 participants were more likely to choose an ongoing-event driven coherence relation following imperfective aspect. This effect did not interact with group, and was present in both groups independently. The differential effect of aspect on reference versus coherence choices also makes an explanation based on memory decay for the event information unlikely: If the event structure information was no longer available or active at the point when L2 participants wrote their continuations, we should have seen similarly weakened effects of aspect on both reference and coherence choices.⁵ Yet this was not the case.

⁵ The same would hold if, as suggested by a reviewer, aspect were simply a “weaker cue when constructing or retrieving information from a mental model of the discourse.” In this case, we would have expected to see reduced effects of aspect on both reference and coherence choices in the L2 group.

The question that arises is how to explain the differential impact of aspect on coreference versus coherence choices in the L2 group. To do so we draw on assumptions and evidence from two largely separate strands of previous research: the expectation-based approach to L1 reference processing within which we situated our study, and recent work on prediction in L2 sentence processing. We must emphasize, however, that our study was not designed to detect effects of prediction at the moment that they arise; more temporally sensitive measures would be required to do so. Yet in line with the expectation-based approach to L1 reference processing, we take biases that arise in written story continuations to be potentially reflective of speakers' incrementally updated discourse expectations, which can be thought of as (probabilistic) prediction beyond the sentence level, or instances of "linguistic pre-processing" (DeLong et al., 2014) more generally. We thus hypothesize that the differential impact of aspect on coreference versus coherence in our L2 group is related to the timing of the respective choices. We assume, based on previous research, that native speakers construct expectations rapidly and incrementally, and continue to update these expectations as new information unfolds. If so, the L1 participants in our task would have constructed a preliminary bias for subject reference as soon as they had processed the subject of the context sentence (since subject position is a good indicator of topichood; recall that in our materials a subject bias corresponds to a bias to the Source referent). Then, upon processing the inflected verb, they would use its aspectual information to update their expectations for coherence and coreference. Therefore, we can account for a strong effect of aspect on both reference and coherence relation within L1 speakers.

What about L2 speakers? Recent work on L2 sentence processing has indicated more limited effects of prediction in L2 vs L1 groups (Kaan, Dallas, & Wijnen, 2010; Martin et al., 2013; cf. Foucart, Martin, Moreno, & Costa, 2014; see Kaan, 2014, for review and discussion of potential causes and mitigating factors). While the focus of inquiry in these studies was on anticipation of upcoming words *within* a sentence (e.g., from the form of a determiner to the following noun), it appears reasonable that expectations *across* sentences would be similarly affected if L2 speakers' ability to generate expectations rapidly and incrementally was reduced compared to L1 speakers. In this case, we would expect L2 speakers to (often) delay decisions about the continuation until after they had processed all of the context sentence, including its sentence-final Goal argument. This delay introduces two potential repercussions. First, we might expect a main effect of group on coreference: L2 speakers could produce more Goal references than L1 speakers, given that the Goal was the most recently encountered, and hence potentially the more salient, argument at the time they were making the reference choice. This is indeed what we found, as reflected in the significant main effect of group on coreference, showing that the L2 group wrote more Goal continuations overall than the L1 group (see Kehler, Hayes, and Barner, 2011, for a similar recency effect in L1 children in a story continuation task with transfer-of-possession contexts). Second, a delay in decisions about the continuation allows a straightforward explanation of the smaller effect of aspect on referential choice for L2 than L1 speakers: Instead of creating expectations based on aspect, and then integrating the subsequently received Goal argument into the expected scenario, L2 speakers would need only to construct a continuation *compatible* with the complete context sentence. Such continuations could

draw on aspectual information, consulted retroactively when deciding on a coherent continuation, but its influence on reference choice would plausibly be weaker at this time point when other constraints known to affect reference are concurrently at play, such as referential form and a tendency to connect to a recent argument. More specifically, if L2 speakers incrementally construct their continuations, we would expect them to first choose the referent of the subject, and then move on to construct the predicate. At the time they process the subject, they would be expected to activate a search for salient entities in their representation of the discourse, and thus connect to the Source and Goal arguments (see, e.g., Lewis, Vasishth & Van Dyke, 2006, for a cue-based model of working memory in sentence processing). Such a search would not directly access event structural information, and so the aspectual form would be less likely to influence the choice of referent when the choice is initiated at this time point. However, once the L2 speakers have moved on to the construction of the predicate of the target sentence, the event structure of the previous sentence becomes relevant for multiple decisions about the target sentence, such as its tense and aspect and the type of predicate that will maintain coherence. We postulate that it is at this time point that the L2 speakers consistently consider the information encoded in the event structure. Once the event structure has been consulted, the L2 speakers then use the contrast between ongoing and completed events to construct an appropriate coherence relation, leading to the significant effect of aspect on Completed-event-driven vs Ongoing-event-driven coherence relations. Our overall pattern of results thus suggests that L1 and L2 speakers both use grammatical aspect to build appropriate event structures, but that L1 speakers use the event structure proactively to predict a coherence relation and a subject referent prior to the completion

of the context sentence, whereas L2 speakers tend to wait to build a coherence relation until farther downstream, after the end of the context sentence and after they have posited a referent for the subject of the continuation.

The account we presented relies on the assumption that linguistic pre-processing, or prediction, constitutes a critical component of language comprehension and processing. This assumption is now well-established for L1 processing, yet it has only recently become a topic of interest in the SLA literature, where heretofore the focus has been primarily on processing as information *integration*, that is, (incremental) meaning construction as incoming information is *received*. The SLA literature on prediction in the sense of linguistic pre-processing is still very limited.⁶ The results to date suggest that, overall, non-native speakers' ability to predict upcoming information at the sentence level based on lexical and morphosyntactic cues is much more variable, and often reduced, when compared to native-speakers (see Kaan, 2014, for review). We suggest that the findings from the present study reflect non-native speakers' similarly variable or reduced reliance on *expectations at the discourse-level*. This is notable because the expectations examined here, unlike those in the previous L2 literature, do not require an obvious

⁶ Note that the term 'prediction' has been used in the SLA literature, primarily in the context of L2 reading, to refer to inference generation, or guessing (e.g., in fill-the-gap tasks), more generally (e.g., McLaughlin, 1987). This usage does not specify the temporal aspect of this process, i.e., *when* such inference generation takes place during the incremental construction of meaning as we read/listen. As such, it is compatible with both (retroactive) information integration and prediction in terms of (proactive) linguistic pre-processing.

accumulation of experience with lexical or morphosyntactic detail. Rather, they draw on associations between ongoing/completed events and coherence relations that seem quite similar in the native and second languages of our participants – i.e. associations that we might expect to be robust and easy to draw on. If this account is on the right track, it points to a broader observation, namely that L2 processing is generally less reliant on linguistic pre-processing. In other words, it raises the possibility that L2 speakers have *Reduced Ability to Generate Expectations* (RAGE). We state this observation as a hypothesis – the RAGE hypothesis – for future research to address. We choose the broadest possible formulation of this hypothesis here because we see value in the SLA field exploring how generally and consistently non-native speakers construct and use expectations. That is, to what extent do they do so, or fail to do so, across different linguistic domains, proficiency levels, L1-L2 typologies, situations of language use, and other factors? We expect that some or all of these dimensions will interact with non-native speakers' tendency to engage in linguistic pre-processing. The goal will be to understand the specific contexts in which the RAGE hypothesis does and does not hold. As Federmeier (2007, p. 495) remarked in the context of reduced effects of prediction observed in aging native speakers, “predictive processing may not be the best – or even a viable – strategy for all individuals at all phases of the lifespan and/or in all processing situations”. *Why* certain circumstances might lead to the reduced involvement of prediction and expectations in language processing, and what exactly those circumstances are, will be important questions to explore in future research, potentially shedding light not just on L2 processing but on language processing across populations.

With regard to discourse-level expectations in L2 processing, we have argued that the results from the story continuation study presented here fit well with both discourse-driven accounts of (L1) referential processing and with an account that assumes reduced reliance on expectations in L2 processing. Although our account is speculative, we believe it provides a compelling explanation for the data we have presented here and raises many intriguing questions for future research. Further support for this account will require capturing these predictive effects as they arise incrementally, using more temporally sensitive methods. Predictive use of discourse-level information in L1 reference processing has been shown successfully in previous studies using the visual-world eye-tracking paradigm (e.g., Pyykkönen & Järvikivi, 2010). Experiments using this paradigm to explore the predictive effects of grammatical aspect with L1 and L2 English speakers are currently in progress in our lab.

Acknowledgments

Many thanks to Michelle Adams, Alexis Toliva, and Mónica Vidal for assistance with data collection and annotation, and to Aya Takeda for assistance with data analysis.

Earlier versions of this research were presented at the 26th Annual CUNY Conference on Human Sentence Processing, GASLA 12, and BUCLD 38; a preliminary report appeared in the proceedings of BUCLD 38. We are grateful for the many helpful comments from these audiences, the anonymous reviewers of this manuscript, and in particular Carrie Jackson. This research has been supported by a Standard Research Grant from the National Science Foundation to the first and the third authors (BCS-1251450), for which we are grateful. Any opinions, findings, and conclusions or recommendations expressed

in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 247-264.
- Ariel, M. (1990). *Accessing noun phrase antecedents*. London: Routledge.
- Arnold, J. E. (2001). The effects of thematic roles on pronoun use and frequency of reference. *Discourse Processes*, *31*, 137-162.
- Arnold, J. E. (2010). How speakers refer: the role of accessibility. *Language and Linguistics Compass*, *4*, 187-203.
- Asher, N., & Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.
- Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavior Research Methods, Instruments & Computers*, *28*, 584-589.
- Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning and use*. Oxford: Blackwell.
- Barr, D. (2013, March 27). Coding categorical variables when analyzing factorial experiments with regression. Retrieved from <http://talklab.psy.gla.ac.uk/tvw/catpred/>

- Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language and Linguistic Theory*, 25, 657-689.
- Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *Modern Language Journal*, 64, 311-317.
- Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of Anaphoric Pronouns. *Journal of Verbal Learning and Verbal Behaviour*, 16, 601–609.
- Carminati, M. N. (2002). *The processing of Italian subject pronouns*. (Unpublished doctoral dissertation). University of Massachusetts at Amherst.
- Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 35, 1029–1040.
- DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistic Compass*, 8, 631-645.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117-1121.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44, 491-505.

- Ferretti, T. R., Rohde, H., Kehler, A., & Crutchley, M. (2009). Verb aspect, event structure, and coreferential processing. *Journal of Memory and Language*, *61*, 191-205.
- Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in English: a corpus of 300 verbs. *Behavioral Research Methods*, *43*, 124-135.
- Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can Bilinguals See It Coming? Word Anticipation in L2 Sentence Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1461-1469.
- Fukumura, K., & van Gompel, P. G. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language*, *62*, 52-66.
- Gabriele, A. (2005). *The acquisition of aspect in a second language: A bidirectional study of learners of English and Japanese*. (Unpublished doctoral dissertation). City University of New York Graduate Center.
- Gabriele, A. (2009). Transfer and transition in the SLA of aspect: A bidirectional study of learners of English and Japanese. *Studies in Second Language Acquisition*, *31*, 371-402.
- Gatt, A., Krahmer, E., van Deemter, K., & van Gompel, R. P. G. (2014). Models and empirical data for the production of referring expressions. *Language, Cognition and Neuroscience, Special Issue on the Production of Referring Expressions: Models and Empirical Data*, *29*, 899-911.
- Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, *21*, 203-225.

- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307.
- Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, 3, 67-90.
- Hopp, H. (2009). The syntax-discourse interface in near-native L2 acquisition: Off-line and online performance. *Bilingualism: Language and Cognition*, 12, 463–483.
- Hwang, M. O. (1983). Topic continuity/discontinuity in Korean narrative. *Korean Linguistics*, 3, 47-80.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, 4, 257–282.
- Kaan, E., Dallas, A., & Wijnen, F. (2010). Syntactic predictions in second-language sentence processing. In J.-W. Zwart & M. de Vries (Eds.), *Structure preserved. Festschrift in the honor of Jan Koster* (pp. 207-213). Amsterdam: John Benjamins.
- Kaiser, E. (2011). Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus. *Language and Cognitive Processes*, 26, 1625-1666.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.
- Kehler, A., Hayes, E., & Barner, D. (2011). Pragmatically-driven biases in children's pronoun interpretation. Poster presented at the 24th annual CUNY conference on human sentence processing, Stanford, CA.

- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25, 1-44.
- Kehler, A. & Rohde, H. (under review). Pronoun Interpretation in a QUD Model of Discourse.
- Kilborn, K. (1992). On-line integration of grammatical information in a second language. In Harris, R. (Ed.), *Cognitive Processing in Bilinguals* (pp. 337-350). Amsterdam: Elsevier.
- Kim, H. (1997). Referential choice for third persons in Korean conversation. *Discourse and Cognition*, 4, 121-158.
- Kim, K., Grüter, T., & Schafer, A. J. (2013) Effects of event-structure and topic/focus-marking on pronoun reference in Korean. Poster presented at the 26th annual CUNY conference on human sentence processing, Columbia, SC.
- Lee, E. H. (2006). Stative progressives in Korean and English. *Journal of Pragmatics*, 38, 695-717.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10, 447-454.
- Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory & Cognition*, 31, 663-672.
- Magliano, J. P., & Schleich, M. C. (2000). Verb aspect and situation models. *Discourse Processes*, 29, 83-112.
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8, 243-281.

- Martin, C. D., Thierry, G., Kuipers, J., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*, 574-588.
- McKoon, G., Greene, S., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology*, *19*, 1040-1052.
- McLaughlin, B. (1987). Reading in a second language: studies with adult and child learners. In S. R. Goldman & H. T. Trueba (Eds.), *Becoming literate in English as a second language* (pp. 57-70). Norwood, NJ: Ablex Publishing Corporation.
- Miltsakaki, E. (2007). A rethink of the relationship between salience and anaphora resolution. In A. Branco (Ed.), *Proceedings of the 6th discourse anaphora and anaphor resolution colloquium* (pp. 91–96). Lagos, Portugal: Discourse anaphora and anaphora resolution colloquium.
- Moens, M., & Steedman, M. (1988). Temporal Ontology and Temporal Reference. *Computational Linguistics*, *14*, 15–28.
- Obana, Y. (2003). Anaphoric choices in Japanese fictional novels: The discourse arrangement of noun phrases, zero and third person pronouns. *Text*, *23*, 405-443.
- Pearson (2011). *Versant English Test: test description and validation summary*. Palo Alto, CA: Pearson Knowledge Technologies. (<http://www.versanttest.com>)
- Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, *57*, 5-16.

- Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition, 30*, 333-357.
- Rohde, H., Kehler, A., & Elman, J. L. (2006). Event structure and discourse coherence biases in pronoun interpretation. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 617-622). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition, and Neuroscience, 29*, 912-927.
- Shirai, Y. (2000). The semantics of the Japanese imperfective *-teiru*: An integrative approach. *Journal of Pragmatics, 32*, 327-361.
- Sorace, A. (2011). Pinning down the concept of 'interface' in bilingualism. *Linguistic Approaches to Bilingualism, 1*, 1-33.
- Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research, 22*, 339-368.
- Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focusing and the representation of events. *Language and Cognitive Processes, 9*, 519-548.
- Ueno, M., & Kehler, A. (2010). The interpretation of null and overt pronouns in Japanese: Grammatical and pragmatic factors. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 2057-2062). Austin: Cognitive Science Society.

APPENDIX 1: List of experimental stimuli in the Story Continuation Task

Emily [brought/was bringing] a drink to Melissa. (She)

Donald [brought/was bringing] a book to Steve. (He)

Brenda [fed/was feeding] a bowl of soup to Anne. (She)

Max [fed/was feeding] a piece of chocolate to Luke. (He)

Maria [gave/was giving] a program to Emma. (She)

Patrick [gave/was giving] a towel to Ron. (He)

Elizabeth [passed/was passing] the bible to Diana. (She)

Tom [passed/was passing] the salt to Nick. (He)

Katherine [pushed/was pushing] a shopping cart to Barbara. (She)

Daniel [pushed/was pushing] a chair to Ken. (He)

Lily [rolled/was rolling] a ball to Betty. (She)

Matt [rolled/was rolling] a suitcase to Bill. (He)

Molly [mailed/was mailing] a letter to Alice. (She)

Adam [mailed/was mailing] a package to Fred. (He)

Rebecca [served/was serving] a piece of chicken to Amanda. (She)

David [served/was serving] a beer to Paul. (He)

Nina [took/was taking] the teapot to Margaret. (She)

Henry [took/was taking] a report to Larry. (He)

Sue [threw/was throwing] a hat to Laura. (She)

Mike [threw/was throwing] the keys to Roger. (He)