CrossMark

Applied Psycholinguistics **38** (2017), 1–26
doi:10.1017/S0142716416000035

The effect of implicit causality and consequentiality on nonnative pronoun resolution

WEI CHENG and AMIT ALMOR
University of South Carolina

Received: March 17, 2015 Accepted for publication: January 30, 2016

ADDRESS FOR CORRESPONDENCE
Wei Cheng, Linguistics Program, 909 Welsh Humanities Building, University of South Carolina, Columbia, SC 29208. E-mail: chengw22@gmail.com

ABSTRACT
We report two sentence-completion experiments investigating how nonnative speakers use universal semantic and discourse information, which are implicit causality and consequentiality biases associated with psychological verbs, to resolve pronouns. The results indicate that intermediate-advanced and advanced Chinese-speaking English learners show weaker implicit causality and consequentiality biases than native English speakers in pronoun resolution. Instead, nonnative speakers exhibit a general subject or first-mention bias. These findings suggest that nonnative speakers do not use semantic and discourse information in comprehension as effectively as native speakers.

One topic in second language (L2) processing research that has received considerable attention is the relationship between syntactic and nonsyntactic information in L2 comprehension. According to the shallow structure hypothesis (Clahsen & Felser, 2006), L2 speakers tend to rely on semantic and discourse information (i.e., the information outside the boundaries of a sentence) in comprehension more than native speakers, who make use of both syntactic and nonsyntactic cues. Many studies have been conducted to test this hypothesis, and some found that L2 speakers are more sensitive to semantic and discourse cues relative to syntactic cues than native speakers (e.g., Felser & Cunnings, 2012; Roberts & Felser, 2011). However, not much attention has been paid to L2 speakers' ability to use semantic and discourse information per se. If L2 speakers rely more on semantic and discourse information than on syntactic cues, can they use semantic and discourse information as effectively as native speakers, especially when such information plays a vital role in comprehension?

A few studies addressed this issue, but the findings are inconclusive. Some studies (Pan & Felser, 2011; Pan, Schimke, & Felser, 2015; Trenkic, Mirkovic, & Altmann, 2014) found that L2 speakers can effectively use semantic and discourse information in comprehension. For example, Pan and Felser investigated Chinese-speaking English learners' resolution of ambiguous prepositional phrases and found that L2 speakers are highly sensitive to discourse cues to resolve these

© Cambridge University Press 2016 0142-7164/16

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

kinds of structural ambiguity, even more so than native speakers. By contrast, other studies (Grüter, Rohde, & Schafer, 2014; Roberts, Gullberg, & Indefrey, 2008) found that this kind of information is not always easy for L2 speakers. For instance, in an online eye-tracking study, Roberts et al. found that when there are two possible referents in discourse, L2 speakers cannot use discourse information as effectively as native speakers to resolve ambiguous pronouns.

Despite the conflicting findings, a couple of recent proposals suggest that semantic and discourse information may pose difficulty for L2 speakers. For example, according to the interface hypothesis (Sorace, 2011a; Sorace & Filiaci, 2006), the information at the syntax–discourse interface is difficult to acquire and utilize for L2 speakers. Moreover, given recent claims in psycholinguistics that language comprehension is characterized by predictions about upcoming utterances (e.g., Levy, 2008), Grüter et al. (2014) proposed the RAGE hypothesis, which states that L2 speakers have a reduced ability to generate expectations during language comprehension (for similar accounts, see Kaan 2014; Kaan, Dallas, & Wijnen, 2010). Under this account, semantic and discourse information also poses a challenge for L2 speakers, who may not be able to effectively use it in predictive processing. This challenge will be more noticeable when native speakers' comprehension benefits from expectations that are driven by semantic and discourse factors.

Another factor that may potentially affect L2 speakers' ability to use semantic and discourse information is first language (L1) influence. One possibility is that L2 speakers may transfer L1-specific processing strategies to L2 comprehension. For example, Roberts et al. (2008) found that due to the pronominal system of their native language, Turkish- and German-speaking learners of Dutch interpreted pronouns differently. In Turkish, pronouns usually refer to a nontopical referent, whereas in German and Dutch, pronouns are resolved toward a local topical entity. In a forced-choice task, Roberts et al. found that while German-speaking learners patterned with native speakers of Dutch in preferring the local entity as the referent of a subject pronoun, Turkish-speaking learners showed no clear preference. The results indicate that L2 comprehension is affected by L1-specific resolution preferences. However, when learners' L1 and L2 have the same processing preferences, it is not clear whether they can benefit from this similarity.

Therefore, although L2 comprehension may primarily rely on semantic and discourse information, evidence is still lacking as to whether L2 speakers can effectively use such information in a nativelike fashion. The current study addresses this question by investigating advanced and intermediate-advanced L1-Chinese L2-English speakers' pronoun interpretation in discourse. Because referring pronouns are free variables, successful pronoun resolution is, to a large extent, dependent on context. Therefore, the ability to identify and use relevant semantic and discourse information is crucial for appropriate pronoun resolution.

In this paper, we focus on a well-studied discourse factor that influences native pronoun resolution: implicit causality and consequentiality. To briefly illustrate what implicit causality and consequentiality are, consider (1). In (1a) and (1b), where the discourse coherence relation is *Explanation* (Kehler, 2002), the pronoun is usually interpreted as referring to the causer of the eventuality described by the verb (the object in [1a] and the subject in [1b]). This reflects the opposite implicit causality biases of the verbs *fear* and *frighten*. In (1c), where the discourse relation

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

is *Result* (Kehler, 2002), the referential preference switches to the subject. This reflects the implicit consequentiality biases of the verb *fear*. The interpretation of the pronoun in (1) is determined by both verb meaning and the coherence relation as represented by the connective (i.e., the word connecting two simple clauses within a sentence). Implicit causality and consequentiality, therefore, provide a good testing ground where we can examine the role of semantic and discourse factors in L2 comprehension.

(1) a. John feared Henry$_i$ because he$_i$ . . .
b. John$_i$ frightened Henry because he$_i$ . . .
c. John$_i$ feared Henry so he$_i$ . . .

IMPLICIT CAUSALITY/CONSEQUENTIALITY AND PRONOUN INTERPRETATION

Findings in English

Various nonsyntactic factors influence pronoun interpretation, one of which is verb semantics. Garvey and Caramazza (1974) and Garvey, Caramazza, and Yates (1976) found that when presented with sentence fragments like those in (1a) and (1b) above, people are likely to interpret the pronoun as coreferential with the object *Henry* in (1a) and the subject *John* in (1b). They argued that the two types of verbs in (1a) and (1b) convey different implications for who is the cause of the eventuality expressed by the verb. Verbs such as *fear* attribute causality to the object, whereas verbs such as *frighten* attribute causality to the subject. This phenomenon is known as implicit causality.

While many verbs are associated with implicit causality, it is psychological verbs that demonstrate the most consistent and robust effect (Hartshorne & Snedeker, 2013). Based on the position of thematic roles, psychological verbs can be classified as either experiencer-stimulus (ES) or stimulus-experiencer (SE) verbs (Brown & Fish, 1983).[1] In sentences with ES verbs such as *fear*, the referent acting as the experiencer occupies the subject position (i.e., the first noun phrase [NP1]) and the referent acting as the stimulus occupies the object position (i.e., the second NP [NP2]). In contrast, SE verbs such as *frighten* have the stimulus in the subject position and the experiencer in the object position.

More important, these referential biases depend on the coherence relations between clauses. Previous studies (e.g., Au, 1986; Stevenson, Crawley, & Kleinman, 1994) found that when the connective between the first clause and the second clause changes from *because* as in (1a) and (1b) to *so* as in (2) below, the preferred reference of the pronoun reverses to NP1 in (2a) and NP2 in (2b). In other words, when the connective denotes a result rather than a cause, ES verbs have an NP1 bias and SE verbs have an NP2 bias.

(2) a. John$_i$ feared Henry so he$_i$ . . .
b. John frightened Henry$_i$ so he$_i$ . . .

This effect is termed implicit consequentiality (Crinean & Garnham, 2006; Stewart, Pickering, & Sanford, 1998), because the focus is on the consequence rather than the cause of the eventuality.

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

Multiple theories have been proposed to explain implicit causality and consequentiality (for a review, see Hartshorne, 2014). One popular view is that implicit causality and consequentiality are related to a verb's argument structure. According to Crinean and Garnham (2006), because the stimulus is the entity that gives rise to the psychological state and the experiencer is the entity that experiences that state, the stimulus is associated with the cause whereas the experiencer is associated with the consequence. Therefore, when the second clause focuses on the cause as indicated by the connective *because*, the stimulus becomes the preferred referent of the second clause, resulting in an NP2 bias for ES verbs and an NP1 bias for SE verbs. By contrast, when the second clause is about the consequence as indicated by the connective *so*, the experiencer becomes the preferred referent, leading to an NP1 bias for ES verbs and an NP2 bias for SE verbs.

In an attempt to explain pronoun resolution in a unified theory, Kehler, Kertz, Rohde, and Elman (2008) and Kehler and Rohde (2013) put forward a Bayesian analysis as shown in (3), which accounts for how implicit causality and consequentiality influence pronoun resolution.

(3) $p(\textit{referent} \mid \textit{pronoun}) = p(\textit{referent}) \times p(\textit{pronoun} \mid \textit{referent}) \,/\, p(\textit{pronoun})$

In this formula, $p(\textit{referent} \mid \textit{pronoun})$ represents a pronoun resolution problem, that is, the probability that a pronoun is coreferential with a certain antecedent. This process is determined by $p(\textit{referent})$, the likelihood that a referent will be mentioned in subsequent discourse, as well as $p(\textit{pronoun} \mid \textit{referent})$, the probability that a particular referent will be referred to by a pronoun. Therefore, pronoun interpretation is affected by two processes: a top-down predictive process as denoted by $p(\textit{referent})$ and a bottom-up integrative process as represented by $p(\textit{pronoun} \mid \textit{referent})$. Before encountering the pronoun, language comprehenders will use contextual cues to generate a prediction of the likelihood of rementioning a certain entity. When they encounter the pronoun, the cue for integration, their interpretation will be modulated by how likely is the entity they have committed to in the predictive process to be referred to by a pronoun rather than other forms of reference such as a repeated name.

Kehler et al. (2008) and Kehler and Rohde (2013) argued that the likelihood of remention, that is, $p(\textit{referent})$, is closely connected with discourse coherence relations. As argued by Kehler (2002), coherence plays an important role in discourse interpretation because people tend to find connections between utterances. As the input unfolds, listeners or readers have a probabilistic expectation about the possible coherence relation between sentences, which in turn determines which entity will be mentioned in subsequent discourse. Many studies attempted to characterize different types of coherence relations (e.g., Hobbs, 1979; Mann & Thompson, 1988). Following Kehler's classification, implicit causality is determined by the coherence relation Explanation and implicit consequentiality by the coherence relation Result.

In contrast, $p(\textit{pronoun} \mid \textit{referent})$ reflects the probability that a given referent will be referred to by a pronoun as opposed to other anaphoric expressions. This probability is influenced by structural factors, particularly subject or first-mention preference. As found in many studies (e.g., Ariel, 1990; Gordon, Grosz, & Gilliom,

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

1993), a subject or first-mentioned antecedent is often referred to with a pronoun rather than other referring expressions.

In sum, to successfully resolve pronouns in the context of implicit causality and consequentiality, language comprehenders need to know (a) verb semantics, which entails knowledge of verb meaning and the associated argument structure; and (b) the discourse coherence relations as indicated by connectives. Moreover, they need to be able to apply these multiple sources of information in the top-down predictive process of pronoun interpretation as outlined in Kehler et al. (2008) and Kelher and Rohde (2013). These properties make pronoun interpretation in the context of implicit causality and consequentiality a good area to test the role of semantic and discourse factors in L2 comprehension.

Findings in Chinese

Given that implicit causality and consequentiality are associated with thematic roles (Crinean & Garnham, 2006), they are taken to represent universal phenomena across languages. In Chinese, psychological predicates have the same argument structure as in English (Chen, 1996). In addition, empirical evidence also supports the universality of implicit causality and consequentiality. For example, by looking at psychological verbs from eight languages including Chinese, Hartshorne, Sudo, and Uruwashi (2013) found a cross-linguistic consistency in implicit causality bias. Across the eight languages, when the connective is *because*, ES verbs such as *fear* show an NP2 bias and SE verbs such as *frighten* an NP1 bias. Studies investigating implicit causality in Chinese further confirmed the existence of this bias in this language (e.g., Jiao & Zhang, 2005; Miao, 1996; Sun, Shu, Zhou, & Zheng, 2001). As far as implicit consequentiality is concerned, we recently found that the same consequentiality biases as English also exist in Chinese: NP1 bias for ES verbs and NP2 bias for SE verbs (Cheng & Almor, 2015).

Although implicit causality and consequentiality biases are taken to be universal, psychological verbs are not. Both Chinese and English have similar ES verbs, but Chinese has a limited set of SE verbs in comparison to English (Chen, 1996; Zhang, 2002). Therefore, native Chinese speakers have much limited exposure to English-like SE verbs.

Furthermore, Chinese and English favor different ways to express causation for SE predicates. While English favors lexical causatives to express causation, such as *John disappointed Henry*, Chinese mainly uses periphrastic causatives, such as *John shi Henry shi-wang* "John made Henry disappointed" (Chen, 1996; Zhang, 2002). The major difference between the two structures is that, in lexical causatives, causation is implicitly incorporated into the psychological verb, whereas in periphrastic causatives, causation is explicitly expressed in the verb *make* in English or *shi* in Chinese.

In a recent study (Cheng & Almor, 2015), we investigated the effect of lexical and periphrastic causatives on pronoun resolution in Chinese. When the discourse relation is Explanation (i.e., implicit causality), we found an equally strong NP1 bias for SE verbs in both lexical and periphrastic causatives. When the discourse relation is Result (i.e., implicit consequentiality), there was an NP2 bias for SE

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

verbs in both lexical and periphrastic causatives. However, this bias was weaker in the latter.

PRONOUN RESOLUTION IN L2 DISCOURSE

Previous studies on L2 reference acquisition and processing primarily focused on syntactic constraints such as binding principles (e.g., Felser, Sato, & Bertenshaw, 2009; Hirakawa, 1990; Kim, Montrul, & Yoon, 2015; Lee & Schachter, 1997; Patterson, Trompelt, & Felser, 2014; Thomas, 1991; White, 1998). The few studies that investigated reference in L2 discourse focused on reflexives (e.g., Demerci, 2000; Felser & Cunnings, 2012; Lee, 2008), and only a couple investigated L2 speakers' pronoun resolution in discourse.

In an eye-tracking experiment on German- and Turkish-speaking learners of L2 Dutch, Roberts et al. (2008) found that when the context contained two possible referents for a pronoun, both Turkish and German speakers had longer reading times than native Dutch speakers, although German has the same pronoun resolution biases as Dutch. They interpreted the results as indicating L2 speakers' difficulty integrating discourse information in online pronoun resolution. In an offline experiment, they found an effect of native language on L2 pronoun resolution. Turkish has a different interpretation of subject pronoun from Dutch, and consequently, Turkish learners of Dutch did not interpret pronouns in the same way as Dutch native speakers. Thus, the similarity between L1 and L2 appears to affect L2 speakers' untimed offline judgments but not their online pronoun resolution, in which the integration of discourse information is affected regardless of their L1 background.

In an offline story-continuation study, Grüter et al. (2014) investigated L2 learners' sensitivity to event structures during pronoun interpretation. Studies on native English speakers have found that perfective and imperfective aspects in source–goal verbs, such as *hand*, express different event structures, which in turn affects reference resolution (Rohde, Kehler, & Elman, 2006). Using context sentences such as *John handed/was handing a book to Bob*, Grüter et al. found that while native English speakers showed more source references (i.e., *John*) in the imperfective (*was handing*) than the perfective (*handed*) condition, Japanese- and Korean-speaking learners of English did not show any significant continuation differences between the two aspect conditions. The results, therefore, indicated that L2 speakers cannot effectively use discourse-level information to resolve reference. However, when presented with items containing a pronoun prompt such as *John handed/was handing a book to Bob. He___*, L2 speakers patterned with L1 speakers by producing more source continuations than in the free prompt condition. Resorting to the Bayesian pronoun resolution model as proposed by Kehler et al. (2008) and reviewed in the last section, Grüter et al. argued that while L2 speakers cannot make effective predictions in the top-down predictive process, they are able to use pronouns as a cue to integrate information in the bottom-up integrative process.

To the best of our knowledge, no previous research has looked at L2 speakers' use of implicit consequentiality bias in pronoun resolution, and only one study, by Liu and Nicol (2010), included implicit causality verbs in their investigation of L2

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

pronoun interpretation. In a self-paced reading experiment, Liu and Nicol found that when pronoun gender was different from the gender of the implicit-causality biased entity, both native English speakers' and Chinese-speaking English learners' reading times were slowed. They interpreted their findings as indicating that L2 speakers are sensitive to implicit causality during online processing. However, because they used mixed genders in their design (their research focus was on gender agreement and anaphora processing rather than on implicit causality per se), their results mainly show that L2 speakers are sensitive to gender cues in online processing. It is still unclear how learners resolve pronouns in the context of implicit causality when there are multiple gender-agreeing referents.

In sum, there are unresolved theoretical questions concerning L2 learners' ability to use semantic and discourse information to resolve pronouns. In addition, empirically, not much is known about how L2 speakers interpret pronouns in the context of implicit causality and consequentiality. This paper addresses these theoretical and empirical issues by examining the role of implicit causality and consequentiality associated with psychological verbs in L2 English pronoun interpretation by advanced and intermediate-advanced Chinese-speaking learners. We conducted two psycholinguistic production experiments that indirectly probed learners' pronoun interpretation. Experiment 1 focused on implicit causality (i.e., with the connective *because*) and Experiment 2 focused on implicit consequentiality (i.e., with the connective *so*). In both experiments, participants produced a natural continuation to a sentence fragment with a pronoun prompt such as *John frightened Henry because he* ___. This method was used by Garvey and Caramazza (1974) and many others to investigate the effect of implicit causality and consequentiality on pronoun resolution. Both experiments contained ES and SE verbs, with the latter alternating between lexical and periphrastic causatives. As is described above, unlike English, Chinese mainly uses periphrastic rather than lexical structures to express causation in SE verbs. The inclusion of the periphrastic structure in the design therefore aims to test the effect of this cross-linguistic difference in grammatical structure on L2 pronominal comprehension.

EXPERIMENT 1

Experiment 1 investigated the role of implicit causality in L2 pronoun interpretation by placing ES and SE verbs under the Explanation discourse coherence relation elicited by the connective *because*. According to previous studies (e.g., Crinean & Garnham, 2006), when the coherence relation is Explanation, the focus of subsequent pronominal reference is on the stimulus, that is, NP2 for ES verbs and NP1 for SE verbs.

Method

Participants. Native English speakers (L1 group, $N = 77$) were recruited from the undergraduate population at the University of South Carolina in the United States. They received extra credit for participation. Although they also knew one or two other languages, none of them reported to be native or advanced speakers

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

of other languages, and all reported that English was their major language of use. Twelve of them were eliminated from analysis due to low completion rate (<70%). The data from the remaining 65 participants (mean age = 18.6 years, range = 18–27 years) were analyzed.

The L2 group was composed of 57 Chinese-speaking learners of English. They were recruited from the Guangdong University of Technology in China and received extra credit for participation. All of them were senior undergraduates majoring in English. L2 participants were asked to finish two tasks: a main task and a posttest task (see details in Procedure). Only those who took both were included in the analysis. In the end, 38 participants finished both tasks. These 38 participants (mean age = 22 years, range = 21–24 years) started learning English in school at ages between 9 and 12 years old. None of them had visited English-speaking countries at the time of the experiment. Their English proficiency was measured through a standardized test (Oxford Quick Placement Test Version 1; $M = 72.7\%$, $SD = 9.8$), which classified them as intermediate-advanced to advanced learners.

Materials and design. Following previous studies on implicit causality in English (e.g., Garvey & Caramazza, 1974), the present study used sentence fragments with a pronoun prompt. Every sentence was composed of a full clause followed by a conjunctive word *because* and a pronoun. The full clause described a two-entity past-tense emotional state represented by a psychological verb. Both arguments were referred to by common English names of the same gender. The connective *because* specified the discourse coherence relation as Explanation (Kehler et al., 2008). The pronoun was of the same gender as the antecedents.

The ES condition contained 16 ES verbs. The SE condition had 32 SE verbs. The ES and SE verbs were not semantically paired because not every SE verb has a corresponding ES verb. All ES and SE verbs were chosen from Hartshorne and Snedeker (2013), a large-scale norming study on the implicit causality bias of English verbs. Only the SE verbs with a high NP1 bias (>65%) and the ES verbs with a low NP1 bias (<35%) were selected. The English instructor of the L2 participants went over all the verbs in this study and verified that they should be familiar for the participants.

The SE verbs further alternated between two types of structures: lexical causatives and periphrastic causatives. In the lexical condition, the verb appeared in its original form, directly conjoining two arguments. In the periphrastic condition, the verb appeared in a derived form and the causative meaning was explicitly expressed in the word *make*. Thus, 64 items were constructed in the SE condition. In all items, the gender of characters was counterbalanced so that half of the items had both female characters and half had both male characters.

Sample items are shown in Example (4):

(4) Sample items in Experiment 1:
ES: Paul liked Alan because he ...
SE (lexical): Ben embarrassed James because he ...
SE (periphrastic): Ben made James embarrassed because he ...

The experiment had a mixed design with one between-item factor and one within-item factor. The between-item factor was verb type (ES vs. SE), and the

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

within-item factor was the structure of SE verbs (lexical vs. periphrastic). All participants saw the same number of items in each condition. For the L1 participants, two experimental lists were prepared with a Latin-square design for the SE verbs. That is, in each list, an SE verb appeared either in the lexical or in the periphrastic condition but not in both conditions. Each list thus contained three conditions as shown in (4) above, with 16 items in each of the two SE conditions and 8 items in the ES condition. The original intention was for each participant to see all 16 ES verbs, but due to a programming error, each list only contained one half of the ES verbs. Each list, however, contained a different half so that all ES verbs were seen by L1 participants. Because we were concerned that too many trials could be overwhelming for L2 participants, resulting in high level of incompliance in our web-based survey (see Procedure subsection), we decided to ask L2 participants to complete only half of the trials L1 participants completed. We thus divided the SE items in the two lists into two parts, producing four lists with 8 items in each SE subcondition. No change was made to the ES verbs, such that each of the four lists still contained eight ES verbs. In sum, there were two lists with 40 experimental items on each list for the L1 group, and four lists with 24 experimental items on each list for the L2 group. On each list, there were equal numbers of experimental items and fillers, which had the same structure as experimental items but different verbs (nonpsychological verbs) and connectives (e.g., *and*, *but*, etc.).

To make sure that L2 participants knew the verbs used in the experiment, a posttest consisting of an acceptability judgment task and a translation task was conducted. Previous studies (Chen, 1996; Juffs, 1996; Zhang, 2002) have shown that the argument structure of psychological verbs, especially SE verbs, poses a great challenge for L2 learners. Therefore, an acceptability judgment task was conducted to ensure that L2 participants had acquired the knowledge that SE verbs can be used in a lexical causative. A translation task was also administered to ensure that L2 participants knew the correct meanings of the verbs.

The materials for the posttest were the same as those in the main experiment (i.e., identical verbs in the same design) except that the connective and pronoun were deleted. Therefore, what participants saw in the posttest were full sentences such as *Ben embarrassed James*. For each item, participants had to first rate the acceptability of the sentence on the basis of six categories: –3 (*completely unacceptable*), –2 (*fairly unacceptable*), –1 (*slightly unacceptable*), 1 (*slightly acceptable*), 2 (*fairly acceptable*), and 3 (*completely acceptable*). Following the rating, they were required to translate the sentence into Chinese. Different fillers were used in the posttest, which were grammatical and ungrammatical sentences of various types such that overall half of the items in the posttest were grammatical and half were not.

Procedure. Both the main experiment and the posttest were conducted on Qualtrics.com, a survey website. Participants took the survey on their own computer. Many studies have shown that web-based surveys yield similar results to studies conducted in a laboratory setting (e.g., Buhrmester, Kwang, & Gosling, 2011; Gosling, Vazire, Srivastava, & John, 2004; Sprouse, 2011). Specifically, web-based methods were used in several previous studies of implicit causality and pronoun resolution, which have repeatedly yielded results that are consistent with

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

the results from studies using traditional paper-and-pencil methods (e.g., Ferstl, Garnham, & Manouilidou, 2011; Hartshorne & Snedeker, 2013). Based on this previous research, we chose to employ a web-based method because this allowed us to conveniently collect data without sacrificing data quality.

Both the L1 and L2 groups did the main experiment, in which participants were asked to type a natural continuation to the sentence fragment and were told to give their most intuitive responses and avoid humor. Before starting the experiment, they filled in a language background questionnaire and then completed a practice block of five filler items (different from the fillers in the experiment) to be familiarized with the task. The surveys were not timed, but participants were instructed to finish the survey at normal speed and complete it in one session.

The posttest was administered to only the L2 group 4 to 6 days after the main experiment. It was administered after the main test to avoid its possible influence on participants' performance in the main test. In the posttest, participants were required to make an acceptability judgment for the items. If they judged the item as unacceptable, they were asked to modify this item to make it acceptable. Because the main purpose of acceptability judgment was to ensure that they had acquired the argument structure of psychological verbs, their modifications allowed us to know if their judgment was related to verb argument structure. For example, there were cases where some participants judged the sentence as unacceptable because they thought it had a tense error. Such cases were irrelevant to our research questions. After judging each item's acceptability, they also had to type the sentence's Chinese translation. L2 participants were also specifically instructed not to look up any unknown words they encountered in the survey.

Participants were randomly assigned to lists, and all items in each list were automatically randomized for each participant. In addition to the main experiment and posttest, L2 participants did a proficiency-assessment test: the Oxford Quick Placement Test Version 1. In case L2 participants might find the survey too long, this test was divided into two sections, which were administered immediately after the main experiment and the posttest, respectively.

Coding. For the data in the main experiment, the reference of the pronoun in the sentence continuation was coded as either NP1 or NP2. Cases where the reference was ambiguous, the continuation was nonsense, the character gender was misunderstood, or no continuation was given were classified as unclear. Examples of continuations are provided in Example (5).

(5) Examples of continuations:
NP1 continuation: Adam amazed Sam because he *could play piano excellently*.
NP2 continuation: Emily envied Sara because she *had a beautiful skirt*.
Ambiguous continuation: Chris frightened Stan because he *was drunk at that time.*
Nonsense continuation: Brian distrusted Richard because he *has some emergency to say*.
Gender misunderstanding continuation: Ella made Olivia exhausted because she *keeps asking him questions*.

The data were coded by the first author and two trained native English speakers unaware of the purpose of this study. One native speaker coded the L1 data and the

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

other coded the L2 data. The coders were instructed to be conservative in coding, such that any ambiguous continuations were coded as unclear. The initial coding agreement rates between the first author and the two coders were 89.5% and 91%, respectively. The coders and the first author went over and resolved most of the disagreements. Those disagreements that could not be resolved were coded as unclear.

For the data in the posttest, L2 participants' translations were coded as correct or incorrect. The data were coded by the first author, who is a native Chinese speaker, and another trained native Chinese speaker, who was unaware of the purpose of the study and had a very advanced level of English. The initial intercoder agreement rate was 96%. The same follow-up procedure was used to resolve the discrepancies. For the acceptability data, any choices of 1, 2, and 3 were coded as acceptable, and any choices of –1, –2, and –3 were coded as unacceptable.[2] In some cases, participants judged a sentence to be unacceptable because they thought it contained such errors as tense, preposition misuse, and so on, as shown in their correction of the sentence. Because these errors were not relevant to the purpose of the study, such trials were still regarded as acceptable. Types of unacceptable judgments made by participants are provided in Note 3.

For the continuation data, there were 8.3% unclear responses in the L1 group and 9.5% unclear responses in the L2 group. Chi-square tests of goodness of fit showed that the unclear responses were equally distributed across conditions in the L1 group, χ^2 (2) = 2.46, p = .293, and L2 group, χ^2 (2) = 4.04, p = .133. The data from these trials were excluded from analysis. Due to a programming error affecting a single SE verb item (*pain*), all the trials containing this item were also excluded. For the posttest data, 11.3% items were wrongly translated (including no translation). For the acceptability data, 11.1% items were judged as unacceptable. For each participant, the trials in the main experiment whose counterparts in the posttest the participant judged as unacceptable or translated incorrectly were excluded from analysis. In total, 8.3% of all responses were excluded in the L1 group and 23% in the L2 group.[4]

Results

Table 1 presents the means and standard deviations of the proportion of NP1 references relative to the total number of NP1 and NP2 references in both L1 and L2 groups. Mixed logit models were used to analyze the data. These models are more suitable than analyses of variance for analyzing categorical (here, whether the pronoun's reference was to NP1 or NP2) and unbalanced data (caused by the elimination of trials with unclear coding or translation/acceptability errors; Jaeger, 2008). Following Barr, Levy, Scheepers, and Tily (2013), all the models contained maximal random effects of subjects and items. All fixed effects were centered. The analysis was conducted using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2014).

The log odds of NP1 reference in the three conditions are presented in Figure 1. Values above zero demonstrate more references to NP1 in the first clause, and values below zero demonstrate more references to NP2.

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

Table 1. *Proportions of NP1 references by groups, verbs and structures in Experiment 1 (out of all NP1 and NP2 references)*

	M	*SD*
First language		
ES verbs	0.022	0.066
SE verbs		
Lexical	0.96	0.079
Periphrastic	0.98	0.047
Second language		
ES verbs	0.20	0.26
SE verbs		
Lexical	0.92	0.19
Periphrastic	0.99	0.045

Note: NP1, NP2, Noun phrases 1 and 2; ES, experiencer stimulus; SE, stimulus experiencer.

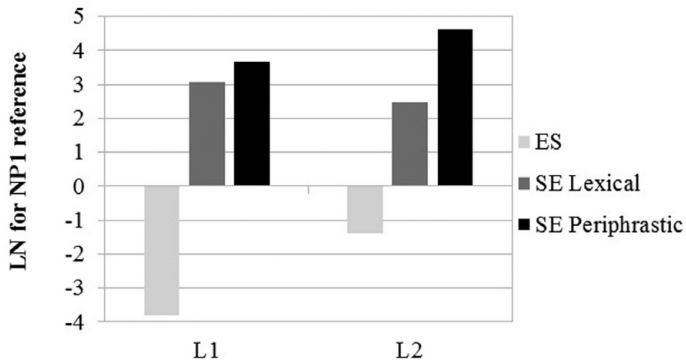


Figure 1. Log odds for noun phrase 1 (NP1) reference in Experiment 1. ES, Experiencer stimulus; SE, stimulus experiencer.

Let us consider the comparison between ES and SE verbs (excluding the periphrastic structure because it does not match with ES verbs in terms of structure). The maximal model did not converge, and was simplified to have group, verb type, and their interaction as fixed effects, and subject and item as random effects with a slope of verb type for subject. It showed a main effect of group ($\beta = 0.53$, $SE = 0.25$, $p < .05$), with the L2 group showing more NP1 references than the L1 group. There was also a main effect of verb type ($\beta = 5.06$, $SE = 0.36$, $p < .001$), such that more NP1 references were observed for SE than ES verbs. Most important, there was a significant interaction between verb type and group ($\beta = –0.93$, $SE = 0.25$, $p < .001$). Follow-up pairwise analyses were conducted within each

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

verb type condition. The maximal model was fitted with group as a fixed effect and subject and item as random effects with a slope of group for item. The results showed that in the ES condition, there were more NP1 references in the L2 group than in the L1 group ($\beta = -0.30$, $SE = 0.25$, $p < .001$). However, no significant difference was found between the two groups in the SE condition ($\beta = -0.76$, $SE = 0.65$, $p = .243$), although numerically there were fewer NP1 references in the L2 group than in the L1 group.

Turning to the two structures of SE verbs, Figure 1 shows more NP1 references in the periphrastic structure than in the lexical structure. This pattern was more prominent in the L2 group than the L1 group. The maximal mixed logit model was fitted with group, structure, and their interaction as fixed effects, subject as a random effect with a slope of structure, and item as a random effect with slopes of structure, group, and their interaction. This model showed no effects of group ($\beta = 0.41$, $SE = 0.39$, $p = .294$) or structure ($\beta = 0.05$, $SE = 0.28$, $p = .862$), but showed a significant interaction between the two factors ($\beta = -0.73$, $SE = 0.35$, $p < .05$). Follow-up analyses were conducted within each group. The maximal model had structure as a fixed effect and subject and item as random effects with a slope of structure for both. The results showed significantly more NP1 references in the periphrastic than in the lexical structure for L2 participants ($\beta = -2.43$, $SE = 1.07$, $p < .05$), but no significant difference was found for L1 participants ($\beta = 0.80$, $SE = 0.56$, $p = .157$).

To assess the possible role of L2 participants' English proficiency in their performance, we repeated the analysis with proficiency as a fixed factor but found no main effect of proficiency or interactions with other factors.

To assess the possible role of variability of the data excluded from analysis across participants (i.e., unclear responses in the L1 group, and unclear responses and data from items with translation/judgment errors in the L2 group), we ran the same statistical analyses as above with the proportion of excluded data per participant as a fixed factor. There was no moderating effect of this factor, indicating that variability of excluded data across participants did not affect the results.

Discussion

Experiment 1 was conducted to examine Chinese-speaking English learners' pronoun interpretation in the context of implicit causality. Consistent with previous research (e.g., Hartshorne & Snedeker, 2013), we found an NP1 bias for SE verbs and an NP2 bias for ES verbs in the L1 group. This suggests that when the discourse relation is Explanation as indicated by the connective *because*, the two types of psychological verbs produce referential biases in the opposite directions.[5] Nonnative speakers also showed similar patterns, indicating that their pronoun interpretation was guided by implicit causality to some extent.

However, we also observed an interaction between verb type (ES vs. SE) and group (L1 vs. L2). When following ES verbs, the pronoun elicited significantly more NP1 references by L2 than L1 speakers, and when following SE verbs, the pronoun showed less NP1 references by L2 than L1 speakers, although this difference did not reach significance. These results suggest that although both L1 and L2 speakers showed implicit causality biases in the same direction, L2

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

speakers had a weaker bias than L1 speakers. Although they had the knowledge about the difference between ES and SE verbs in terms of causality bias, they did not apply this semantic and discourse information as effectively as L1 speakers.

One remaining question is why L2 speakers showed a significant difference in reference from native speakers in the ES condition but not in the SE condition. One explanation is that L2 speakers' ability to calculate implicit causality is reduced for ES verbs but not for SE verbs. However, this is unlikely given previous findings that ES verbs are much easier to acquire than SE verbs (Juffs, 1996) and the fact that ES verbs rather than SE verbs are used similarly in English and Chinese. If L2 speakers could perform at the native level for SE verbs, they should have been able to do so for ES verbs.

Another possibility is that L2 speakers' seemingly superior performance for SE verbs reflects a general subject or first-mention bias. Subject or first-mention preference is one of the most important factors influencing pronoun interpretation (e.g., Crawley, Stevenson, & Kleinman, 1990; Gernsbacher, Hargreaves, & Beeman, 1989; Järvikivi, van Gompel, Hyönä, & Bertram, 2005). Therefore, the subject/first-mention bias may compensate learners' inefficiency in calculating implicit causality biases in SE verbs, resulting in an increased NP1 reference in the SE condition. Such an NP1 bias was also reflected in the main effect of group, with more NP1 references in the L2 than in the L1 group. This possibility will be further explored in Experiment 2.

Considering the difference between Chinese and English psychological verbs, the present experiment also manipulated the structure of SE verbs. We found an interaction between group and structure, such that significantly more NP1 references were observed in the periphrastic than in the lexical condition in the L2 group, but no significant difference was observed in the L1 group. This difference may result from L1 influence. Although Chinese lexical and periphrastic causatives have similar strength of referential bias in the context of implicit causality (Cheng & Almor, 2015), Chinese favors periphrastic causatives to express causation in SE verbs. Therefore, our Chinese learners of English likely found it easier to process causation in English periphrastic than lexical causatives.

EXPERIMENT 2

Following the same design as in Experiment 1 and changing the discourse relation to Result, Experiment 2 examined the effect of implicit consequentiality on Chinese-speaking English learners' pronoun resolution.

Previous studies (e.g., Stevenson et al., 1994) showed that changing the connective from *because* to *so* leads to reverse referential biases for both ES and SE verbs. When the connective is *so*, the clause that follows typically describes the consequence, and the focus shifts to the consequence of the psychological state, that is, the experiencer (NP1 for ES verbs and NP2 for SE verbs; Crinean & Garnham, 2006). Experiment 1 showed a reduced effect of implicit causality on Chinese-speaking learners' pronoun interpretation. The question remains as to whether these findings generalize to the context of implicit consequentiality.

In addition, Experiment 1 found no significant difference between the L1 and L2 groups in the SE verb condition. We argued that this could reflect a general

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

subject/first-mention bias for L2 speakers rather than a native-level ability to compute implicit causality biases for these verbs. Because SE verbs have an NP2 consequentiality bias, Experiment 2 provides an opportunity to tease apart the two possible accounts. If L2 speakers have no difficulty computing referential biases associated with English SE verbs, they should exhibit nativelike performance in Experiment 2 as well, showing a strong NP2 bias. However, if they do resort to a general subject/first-mention bias, they should not show as strong an NP2 bias as native speakers in this experiment.

Method

Participants. New participants were recruited from the same population as in the first experiment. Forty-one native English speakers were recruited. The data from 3 participants were eliminated due to low completion rate (<70%). The data of the remaining 38 native English speakers (mean age = 18.6 years, range = 18–22 years) were included in the final analysis. None of them reported to be native or advanced speakers of other languages, and all reported that English was their major language of use.

Forty-five Chinese-speaking learners of English were recruited from the same population as in the first experiment. In the end, 28 participants (mean age = 22.1 years, range = 21–24 years) finished both the main test and the posttest. They started learning English in school at ages between 9 and 12 years old. None of them had been to English-speaking countries before the time of the experiment. Their English proficiency was measured through the same standardized test as used in Experiment 1 ($M = 73.5\%$, $SD = 10.7$), which classified them as intermediate-advanced to advanced learners. An independent-samples t test was conducted to compare the English proficiencies of L2 participants in Experiments 1 and 2. The results did not show a significant difference, $t\ (64) = -0.32$, $p = .376$, indicating that they had equivalent proficiency.

Materials and design. The materials (see (6) below) were the same as in Experiment 1, except that the connective was changed to *so*, which elicited a Result discourse relation.

(6) Sample items in Experiment 2:
ES: Paul liked Alan so he . . .
SE (lexical): Ben embarrassed James so he . . .
SE (periphrastic): Ben made James embarrassed so he . . .

The design of the second experiment was the same as that of the first experiment, except that more items were included in this experiment. Both L1 and L2 groups read two lists. Each list had 16 items in each of the three conditions: ES, SE lexical, and SE periphrastic. On each list, every SE verb only appeared in either the lexical or the periphrastic condition but not in both. Forty-eight fillers were included on each list.

Procedure. The procedure was the same as that in Experiment 1.

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

Table 2. *Proportions of NP1 references by groups, verbs, and structures in Experiment 2 (out of all NP1 and NP2 references)*

	M	*SD*
First language		
ES verbs	0.95	0.11
SE verbs		
Lexical	0.14	0.21
Periphrastic	0.15	0.24
Second language		
ES verbs	0.91	0.14
SE verbs		
Lexical	0.63	0.26
Periphrastic	0.82	0.16

Note: NP1, NP2, Noun phrases 1 and 2; ES, experiencer stimulus; SE, stimulus experiencer.

Coding. The same data coding method as in Experiment 1 was used. The data from the main experiment were coded by the first author and two new trained native English speakers unaware of the purpose of this study. One native speaker coded the L1 data and the other coded the L2 data. The intercoder agreement rates were 84% and 86%, respectively. The translation data was coded by the first author and the same bilingual coder as in Experiment 1. The coding agreement rate was 95%. All disagreements were resolved following the same procedure as in Experiment 1.

For the continuation data, there were 19.8% unclear responses in the L1 group and 22.2% in the L2 group. Chi-square tests of goodness-of-fit showed that the unclear responses were equally distributed across conditions in the L1 group, χ^2 (2) = 2.54, p = .279, and the L2 group, χ^2 (2) = 0.88, p = .643. In the posttest, 13.5% trials were wrongly translated (including no translation), and 6.8% were judged as unacceptable. In total, 19.8% of all responses were excluded in the L1 group and 32% in the L2 group. Due to a programming error, all the trials containing the SE verb *pain* were further excluded.

Results

Table 2 presents the means and standard deviations of the proportion of NP1 references relative to the total number of NP1 and NP2 references in both the L1 and L2 groups. The same analysis procedure as in Experiment 1 was employed to analyze the data in Experiment 2.

Let us first focus on ES verbs and SE verbs in lexical causatives (excluding periphrastic causatives). As Figure 2 shows, there was a clear referential difference between ES and SE verbs in the L1 group, with more references to NP1 following

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

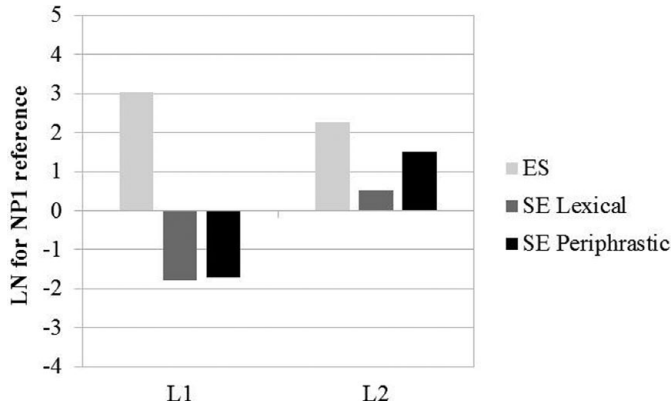


Figure 2. Log odds for noun phrase 1 (NP1) reference in Experiment 2. ES, Experiencer stimulus; SE, stimulus experiencer.

ES and more references to NP2 following SE verbs. However, the L2 group showed a different pattern. There were more NP1 references following both ES and SE verbs, although the probability was higher for the former. The maximal model did not converge, and was simplified with group, verb type, and their interaction as fixed effects, and subject and item as random effects with a slope of group for item. A logit model showed that there was a main effect of group ($\beta = 0.62$, $SE = 0.23$, $p < .01$), with the L2 group showing more NP1 references than the L1 group, a main effect of verb type ($\beta = –2.47$, $SE = 0.15$, $p < .001$), such that the ES condition had a higher probability of NP1 reference than the SE condition, and a significant interaction between group and verb type ($\beta = 0.95$, $SE = 0.14$, $p < .001$). Follow-up pairwise analyses were conducted within each verb type condition. The maximal model contained group as a fixed effect and subject and item as random effects with a slope of group for item. In the SE condition, L2 speakers showed a significantly higher probability of NP1 reference than L1 speakers ($\beta = 1.57$, $SE = 0.22$, $p < .001$); and in the ES condition, they showed a slightly lower probability of reference to NP1 than L1 speakers, which did not reach significance ($\beta = –0.43$, $SE = 0.37$, $p = .244$).

Focusing on SE verbs only, the maximal mixed logit model was fitted with group, structure, and their interaction as fixed effects, subject as a random effect with a slope of structure, and item as a random effect with slopes of structure, group, and their interaction. There was also a main effect of group ($\beta = 1.97$, $SE = 0.21$, $p < .001$), with the L2 group showing more NP1 references than the L1 group, an effect of structure approaching significance ($\beta = –0.21$, $SE = 0.11$, $p = .071$), with the periphrastic structure having more NP1 references than the lexical structure, and a significant interaction between group and structure ($\beta = –0.32$, $SE = 0.09$, $p < .001$). Pairwise analyses were conducted within each group. The maximal model had structure as a fixed effects and subject and item as random effects, with a slope of structure for both. The results revealed that L2 participants made significantly more NP1 references in the

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

periphrastic condition than in the lexical condition ($\beta = -0.48$, $SE = 0.17$, $p < .01$), whereas the L1 group did not show a significant difference ($\beta = 0.10$, $SE = 0.13$, $p = .443$).

Like in Experiment 1, we examined the role of proficiency in L2 participants' performance by repeating the analysis with proficiency as a fixed factor. The results did not show any modulating effect of proficiency. In addition, to assess whether the variability of excluded data across participants influenced our results, we repeated the analyses with the proportion of excluded data per participant as a fixed factor and found no moderating effect either.

Discussion

Experiment 2 examined Chinese-speaking L2 English learners' pronoun interpretation in the context of implicit consequentiality, that is, when the discourse coherence relation focuses on the individual associated with the consequence of a psychological change (i.e., the experiencer). As expected, when the connective was *so*, L1 speakers showed a strong NP1 bias for ES verbs and an NP2 bias for SE verbs. The direction of this bias was the inverse of that in Experiment 1 when the connective was *because*. The contrast between the two experiments is consistent with previous findings on L1 pronoun resolution (e.g., Stevenson et al., 1994), providing validity to the test items used in the current study.

Although L2 speakers made more references to NP1 in the ES condition than in the SE condition, they did not show an NP2 bias in the SE condition. Rather, they exhibited an NP1 bias in both structures of SE verbs, particularly in the case of periphrastic structures. Because implicit consequentiality biases the reference to the experiencer (i.e., NP2 for SE verbs), the results indicate that L2 participants had reduced sensitivity to implicit consequentiality biases.

These results are different from L2 speakers' performance in Experiment 1, which showed no significant difference between the L1 and L2 groups in the SE condition. Although on the surface, that finding may seem to indicate that our L2 participants in Experiment 1 had no difficulty computing causality bias associated with SE verbs, we argued in the discussion of Experiment 1 that it could actually reflect a general subject/first-mention bias for L2 speakers. Because L2 participants in both experiments were drawn from the same population and had similar English proficiencies, their non-nativelike performance with SE verbs in the current experiment supports this claim. When the discourse bias associated with SE verbs is not NP1 anymore, as in the case of implicit consequentiality, L2 speakers' resort to subject/first-mention bias would result in their nonnativelike performance.

Finally, compared with the L1 group, the L2 group showed a significantly stronger NP1 bias in periphrastic causatives than in lexical causatives. Because the implicit consequentiality bias is toward NP2, L2 participants seemed to perform worse in the periphrastic condition. Their different performances between the two structures are probably due to L1 influence. As was mentioned in the literature review, in Chinese when the discourse relation is Result, periphrastic causatives lead to more NP1 references in subsequent discourse than lexical causatives (Cheng & Almor, 2015). Therefore, L2 participants' performance with the two types of

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

causatives in this experiment may be influenced by this feature of their native language.

GENERAL DISCUSSION

One central question motivating this study is whether L2 speakers use semantic and discourse information effectively in comprehension. Although L2 comprehension has been argued to be primarily based on semantic and discourse information (Clahsen & Felser, 2006), whether it is easy for L2 learners to use this information is still unclear. We address this question by focusing on intermediate-advanced and advanced L1-Chinese L2-English learners' pronoun interpretation in the context of implicit causality and consequentiality. By embedding two types of psychological verbs (ES and SE verbs) in different discourse coherence relations (Explanation and Result), we found that although L2 speakers were to some extent aware of the different causality and consequentiality biases of ES and SE verbs, they could not apply this information as effectively as L1 speakers in pronoun resolution. In both experiments, when the discourse-biased referent was NP2 (ES verbs in Experiment 1 and SE verbs in Experiment 2), L2 speakers showed more NP1 references than L1 speakers, exhibiting a subject or first-mention bias. L2 speakers' subject or first-mention bias was not only observed in the lexical condition of SE verbs but also in the periphrastic condition.

Because implicit-causality and consequentiality-biased pronoun interpretation involves information about verb semantics and discourse coherence relations (Kehler et al., 2008), the present study shows that L2 speakers, at least those at the intermediate-advanced and advanced levels, cannot use semantic and discourse information such as implicit causality and consequentiality as effectively as L1 speakers to resolve pronouns. This finding has implications for L2 processing theories. The shallow structure hypothesis argues that, unlike native speakers, L2 speakers' comprehension is guided less by syntactic information and more by semantic and pragmatic information (Clahsen & Felser, 2006). However, if L2 speakers rely more on semantic and discourse information for successful comprehension, does it follow that semantic/discourse information is easy for them to use? Previous research on L2 processing focused on the L1–L2 difference in relative availability of different types of information, making it difficult to answer this question. In contrast, our results indicate that semantic/discourse information may not be very easy for L2 speakers to use. Even though verb implicit causality/consequentiality is employed in their native language processing to resolve pronouns, L2 speakers in our study still could not use this type of information very effectively. Therefore, although L2 comprehension may be primarily guided by semantic and discourse information, this does not mean that this type of information can be used efficiently, at least for learners who have not reached nativelike proficiency. This is an important fact that needs to be addressed by any theory, such as the shallow structure hypothesis, which argues for an overall advantage of semantic and pragmatic information over syntactic information in L2 processing.

The remaining question is what underlies L2 speakers' difficulty with implicit causality and consequentiality. One possible cause is that L2 speakers are less

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

efficient than L1 speakers when processing multiple levels of information. Implicit causality and consequentiality not only involve lexical semantic information but also discourse coherence relations. Therefore, resolving pronouns in the context of implicit causality and consequentiality requires the integration of both semantic and discourse information. As our results show, this integration poses a challenge for L2 speakers. In both experiments, the different causality biases encoded in ES and SE verbs provided native speakers strong cues about the potential referent. However, L2 speakers showed a weakened sensitivity to implicit causality and consequentiality biases, resulting in less robust referential preferences than native speakers. These findings are therefore consistent with previous studies showing that L2 speakers have difficulty integrating multiple sources of information (Hopp, 2010; Roberts et al., 2008; Sorace & Filiaci, 2006).

Such difficulty may result from nonlinguistic factors such as executive control of general cognitive resources. For example, Sorace (2011a, 2011b) argued that the need to inhibit the language they are not using at a particular instant may lead bilinguals to spend less attentional resources than monolinguals on linguistic tasks. Under this account, as late bilinguals, L2 speakers in our study needed to inhibit their native language when performing the English task. Therefore, they may have been left with insufficient cognitive resources to effectively compute implicit causality and consequentiality biases to resolve pronominal reference.

In addition, the present findings are consistent with the RAGE hypothesis (Grüter et al., 2014), which holds that L2 speakers have a reduced ability to generate expectations. The Bayesian analysis of pronoun resolution (Kehler et al., 2008; Kehler & Rohde, 2013) maintains that pronominal reference is, to a large extent, dependent on language comprehenders' prediction about upcoming referent in the top-down predictive stage, which is determined by their computation of local coherence relations based on available semantic and discourse information, that is, implicit causality and consequentiality in our case. Therefore, our findings that L2 speakers cannot effectively use implicit causality and consequentiality to resolve pronouns may be attributed to their inadequate ability to generate appropriate expectations about upcoming referent in the predictive stage of pronoun resolution.

Our results also show more NP1 references in the L2 group than in the L1 group. In our discussions within the two experiments, we suggest that this may be due to L2 speakers' subject/first-mention preference in pronoun resolution. Their adoption of subject/first-mention as a default strategy could be related to their reduced ability to use semantic and discourse information. As argued above, using the information of implicit causality and consequentiality to establish coreference may be challenging for L2 speakers because it involves the integration of various kinds of semantic and discourse information. By contrast, it may be easier for them to use the link between pronouns and subject/first-mention coreference, which is only concerned with either an antecedent's grammatical role or surface position. Therefore, given the challenge of using semantic and discourse information, L2 speakers may prefer the easier subject/first-mention strategy, resulting in increased NP1 references.

According to the RAGE account (Grüter et al., 2014), this NP1 bias could also be related to L2 speakers' reduced ability to make predictions. Subject/first-mention

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

bias is one of the major biases in pronoun resolution (e.g., Crawley et al., 1990; Gernsbacher et al., 1989). In the Bayesian approach to pronoun resolution (Kehler et al., 2008; Kehler & Rohde, 2013), this bias is associated with the bottom-up integrative process in which the probability that a given referent can be referred to by a pronoun rather than other types of referential forms is calculated. Studies with monolinguals have established that subject or first-mentioned antecedents are often referred to by pronouns rather than other referring expressions (e.g., Gordon et al., 1993; Järvikivi et al., 2005). Thus, the presence of a pronoun in the sentence fragment may increase the chance of its reference to the subject or first-mentioned entity. If L2 speakers' ability to make predictions is reduced, they cannot engage effectively in the top-down predictive process in pronoun interpretation. As a result, they need to rely more on the bottom-up process to achieve resolution, demonstrating a subject/first-mention bias in their final pronominal interpretation. By contrast, because native speakers' ability to make predictions does not suffer, their final resolution is primarily based on the calculation in the top-down process, thereby showing a bias strongly induced by implicit causality and consequentiality rather than whether the antecedent is the subject or first-mentioned entity. Of course, it needs to be pointed out that although our findings are consistent with the RAGE hypothesis, we did not directly test it in this study. The current results provide a basis to explore this possibility in future studies.

In addition to a general inefficiency in processing semantic and discourse information, another possible reason for L2 speakers' difficulty in computing implicit causality and consequentiality biases is L1 influence. As noted earlier, although implicit causality and consequentiality are taken to be a universal phenomenon, Chinese and English differ in the syntactic representations of psychological verbs. Specifically, Chinese favors periphrastic causatives over lexical causatives to express causation in SE verbs. This grammatical difference may influence Chinese learner' performance such that they find it difficult to compute the causality and consequentiality biases associated with SE verbs as represented in lexical causatives. Therefore, L2 participants' performance with SE lexical causatives could result from not only their reduced sensitivity to semantic and discourse information but also L1 influence.

Although the two factors are confounded with regard to SE lexical causatives, L1 influence cannot explain the results related to ES verbs and SE periphrastic causatives. In Experiment 1, where the causality bias for ES verbs is NP2, L2 participants showed more NP1 references than L1 participants. In Experiment 2, where the consequentiality bias for SE verbs is NP2, L2 participants showed an NP1 bias for periphrastic causatives. Because both ES verbs and SE periphrastic causatives are frequently used psychological predicates in Chinese, there should be no interference from learners' L1. Therefore, our results indicate that the presence of corresponding structures in their L1 does not help L2 speakers use semantic and discourse information in a nativelike way. Even though learners' L1 and L2 share the same grammatical structures and comprehension biases, due to a reduced ability to use semantic and discourse information, they may still not be able to effectively use relevant information to arrive at nativelike comprehension.

Another possible source of L1 influence is that Chinese has both null and overt pronouns whereas English only has overt pronouns. As was found in Roberts

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

et al. (2008), if null pronouns and overt pronouns in learners' L1 are used in distinct discourse contexts, their interpretation of pronouns in L2 may be affected. However, we do not think that this difference in pronominal system provides a convincing account for our participants' non-nativelike performance. Although no studies have looked at the resolution of Chinese null pronouns in the context of implicit causality and consequentiality, given the universality of these biases, it is likely that Chinese null pronouns are resolved in the same way as overt pronouns. Unlike some other pro-drop languages such as Turkish, Italian, and Spanish, in which null and overt pronouns are used in distinct contexts, Chinese pronouns are frequently used in discourse and may not differ from null pronouns in terms of referent discourse accessibility (Yang, Gordon, Hendrick, & Wu, 1999). According to several native Chinese speakers we interviewed, for the Chinese sentences equivalent to the English items used in this study, replacing the overt pronouns with null pronouns does not lead to a change in interpretation or well-formedness. Therefore, L2 speakers' non-nativelike performance cannot be simply attributed to the fact that null pronouns are used in their native language.

Although this is the first study to show that L2 speakers have reduced sensitivity to the information of implicit causality and consequentiality when resolving ambiguous pronouns, it clearly has some limitations that will need to be addressed in future studies. We collected data via web-based sentence-completion surveys. Although web surveys have been proven to be a sound instrument to study pronoun reference resolution (Ferstl et al., 2011; Hartshorne & Snedeker, 2013), one potential disadvantage of this method is that we were unable to monitor the completion process. Therefore, we could not ensure that all participants completed the survey under similar conditions. Another limitation of our web-based survey method is that it only provided us with offline rather than online measures. Although such offline completion measures are common in studies of preferential interpretation in general, and in studies of implicit causality and consequentiality in particular, these measures only assess comprehenders' final interpretation of pronouns instead of the time course of this process. In addition, compared with online tasks, when finishing offline tasks, participants are under less cognitive pressure and therefore may resort more to strategic metalinguistic processing that is based on explicit rather than their implicit linguistic knowledge. Thus, as found in previous studies (Roberts et al., 2008; Zuffery, Mak, Degand, & Sanders, 2015), L2 speakers' performance may differ between the two types of tasks. However, as Sorace (2011a) pointed out, this does not mean that offline data cannot tell us anything about how people process language. Rather, we believe that the offline data as reported here provide a basis for future studies to examine L2 speakers' online pronoun resolution in the context of implicit causality and consequentiality.

Finally, although this study took L2 participants' English proficiency into consideration by focusing on intermediate-advanced and advanced learners, we did not specifically test the role of this factor. It may be possible that the ability to resolve pronouns in discourse is a developmental issue, with more advanced learners such as those at the end state of acquisition performing more nativelike. The potential influence of proficiency level in the processes we investigate here is an important question that future research will need to address.

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation Grant BCS0822617 (to A.A.). We extend our sincere thanks to the following people for their help with the study: Xiaojing Yang who recruited the Chinese participants, the undergraduate research assistants in the Language and Cognition aLab at the University of South Carolina (David Ziegler, Karen Shebusky, Martina Nelson, and Sheida Abdi), and Xianmin Shen who helped with coding the data.

NOTES

1. Note that Brown and Fish's (1983) classification of psychological verbs (ES vs. SE verbs) is different from that in the literature of theoretical linguistics, where psychological verbs are named according to the position of the experiencer. That is, *fear*-type verbs are experiencer–subject verbs whereas *frighten*-type verbs are experiencer–object verbs. To be consistent with previous studies on implicit causality and consequentiality, this paper follows the terminology of Brown and Fish (1983).
2. Note that our use of the 6-point scale in the acceptability judgment task was not the same as other grammaticality judgment studies. We used a 6-point scale in the judgment task because acceptability judgments are usually subtle, such that a dichotomy of choices may not accurately capture participants' grammatical knowledge. In our analysis, we dichotomized the scale into acceptable and unacceptable rather than calculating the judgment score as grammaticality judgment studies normally do, because we were not interested in what participants scored in this task, but rather whether they knew that the verb could be used in a certain way. Therefore, what was of importance to us is whether they considered the sentence acceptable or unacceptable instead of the strength of the judgment.
3. All of the target verbs in the judgment task were the same ES and SE verbs as in the main test. The design was also the same as the main test, such that ES verbs appeared in the subject–verb–object structure and SE verbs appeared in either the subject–verb–object lexical structure or the periphrastic structure. All the target items were meant to be acceptable. Sometimes, participants judged the sentences as unacceptable because they thought that (a) the verb cannot be used in the periphrastic structure, (b) the verb cannot be used in the lexical structure, and (c) human entity cannot appear in the subject or object position. These judgments indicate that participants did not have the right knowledge about the argument structure of these verbs. Therefore, their performance on these items in the main test was not reliable, and the corresponding data in the main test were excluded from analysis. Participants sometimes judged the sentences as unacceptable because they thought that (a) the tense was not correct and (b) there lacked a preposition after the verb. These unacceptable judgments were not relevant for our purpose, and the corresponding data in the main test were not excluded.
4. Note that the percentage of the excluded data in the L2 group does not equal the total of the percentages of "unclear" coding, wrong translation, and unacceptability judgment. This is because some trials had both wrong translation and judgment, and the counterparts of some wrongly translated or judged trials had "unclear" coding in the main experiment.
5. As pointed out by one reviewer, in Experiment 1 (and also in Experiment 2), native speakers showed a stronger referential bias in the ES condition than in the SE lexical

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

condition. This pattern is consistent with previous findings (e.g., Ferstl et al., 2011). We do not have an explanation for this currently, nor did previous studies. Future research is required to better understand the mechanism behind this pattern. However, as far as the present study is concerned, the point is that despite a difference in bias strength, native English speakers did show opposite directions of reference between the two types of verbs.

REFERENCES

Ariel, M. (1990). *Accessing NP antecedents*. London: Routledge.

Au, T. K. F. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language, 25*, 104–122.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (R package version 1.1-7) [Computer software]. Retrieved from http://CRAN.R-project.org/package=lme4

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, *14*, 237–273.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5.

Chen, D. (1996). *L2 acquisition of English psych predicates by native speakers of Chinese and French*. Unpublished doctoral dissertation, McGill University.

Cheng, W., & Almor, A. (2015). *The effect of lexical and periphrastic causatives on pronoun resolution: Evidence from Chinese*. Poster session presented at the 28th CUNY Conference on Human Sentence Processing, Los Angeles, March.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3–42.

Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research, 19*, 245–264.

Crinean, M., & Garnham, A. (2006). Implicit causality, implicit consequentiality and semantic roles. *Language and Cognitive Processes, 21*, 636–648.

Demerci, M. (2000). The role of pragmatics in reflexive interpretation by Turkish learners of English. *Second Language Research, 14*, 325–353.

Felser, C., & Cunnings, I. (2012). Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics, 33*, 571–603.

Felser, C., Sato, M., & Bertenshaw, N. (2009). The on-line application of binding Principle A in English as a second language. *Bilingualism: Language and Cognition, 12*, 485–502.

Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in English: A corpus of 300 verbs. *Behavior Research Methods*, *43*, 124–135.

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry, 5*, 459–464.

Garvey, C., Caramazza, A., & Yates, J. (1976). Factors influencing assignment of pronoun antecedents. *Cognition*, *3*, 227–243.

Gernsbacher, M. A., Hargreaves, D. J., & Beeman, M. (1989). Building and accessing clausal representations—The advantage of 1st mention versus the advantage of clause recency. *Journal of Memory and Language*, *28*, 735–755.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*, 311–347.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59*, 93–104.

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

Grüter, T., Rohde, H., & Schafer, A. (2014). The role of discourse-level expectations in non-native speakers' referential choices. In W. Orman & M. J. Valleau (Eds.), *BUCLD 38: Proceedings of the 38th Annual Boston University Conference on Language Development* (pp. 179–191). Somerville, MA: Cascadilla Proceedings Project.

Hartshorne, J. K. (2014). What is implicit causality? *Language, Cognition and Neuroscience*, *19*, 804–824.

Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes, 28*, 1474–1508.

Hartshorne, J. K., Sudo, Y., & Uruwashi, M. (2013). Are implicit causality pronoun resolution biases consistent across languages and cultures? *Experimental Psychology, 60*, 179–196.

Hirakawa, M. (1990). A study of the L2 acquisition of English reflexives. *Second Language Research*, *6*, 60–85.

Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science, 3*, 67–90.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*, 901–931.

Jaeger, F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Järvikivi, J., van Gompel, R. P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science, 16*, 260–264.

Jiao, J., & Zhang, B. (2005). The effect of the implicit causality of Chinese verbs on pronoun resolution. *Psychological Science, 28*, 1082–1085.

Juffs, A. (1996). Semantics-syntax correspondences in second language acquisition. *Second Language Research, 12*, 177–121.

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-prediction in second language processing. In J. W. Zwart & M. de Vries (Eds.), *Structure preserved: Studies in syntax for Jan Koster* (pp. 207–213). Amsterdam: John Benjamins.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics, 25*, 1–44.

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, *39*, 1–37.

Kim, E., Montrul, S., & Yoon, J. (2015). The on-line processing of binding principles in second language acquisition: Evidence from eye tracking. *Applied Psycholinguistics*, *36*, 1317–1374.

Lee, D., & Schachter, J. (1997). Sensitive period effects in binding theory. *Language Acquisition, 6*, 333–362.

Lee, K.-Y. (2008). The role of pragmatics in reflexive interpretation by Korean learners of English. In M. Bowles, R. Foote, S. Perpiñán, & R. Bhatt (Eds.), *Selected Proceedings of the 2007 Second Language Research Forum* (pp. 97–112). Somerville, MA: Cascadilla Proceedings Project.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177.

Liu, R., & Nicol, J. (2010). Online processing of anaphora by advanced English learners. In M. T. Prior, Y. Watanabe, & S. K. Lee (Eds.), *Selected Proceedings of the 2008 Second Language Research Forum* (pp. 150–165). Somerville, MA: Cascadilla Proceedings Project.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text, 3*, 243–281.

Miao, X. (1996). A study of semantic and grammatical factors influencing pronoun processing. *Acta Psychological Sinica*, *28*, 352–358.

Pan, H.-Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, *121*, 221–236.

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press

Pan, H.-Y., Schimke, S., & Felser, C. (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism, 19*, 298–313.

Patterson, C., Trompelt, H., & Felser, C. (2014). The online application of binding condition B in native and non-native pronoun resolution. *Frontiers in Psychology, 5*, 147. doi:10.3389/fpsyg.2014.00147

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden-paths in second-language sentence processing. *Applied Psycholinguistics, 32*, 299–331.

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. (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. 26–29). Austin, TX: Cognitive Science Society.

Sorace, A. (2011a). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism, 1*, 1–33.

Sorace, A. (2011b). Cognitive advantages in bilingualism: Is there a "bilingual paradox"? In P. Valore (Ed.), *Multilingualism: Language, power, and knowledge* (pp. 335–358). Pisa: Edistudio.

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research, 22*, 339–368.

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods, 43*, 155–167.

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, *9*, 519–548.

Stewart, A. J., & Pickering, M. J., & Sanford, A. J. (1998). Implicit consequentiality. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 1031–1036). Mahwah, NJ: Erlbaum.

Sun, Y., Shu, H., Zhou, X., & Zheng, X. (2001). The effect of implicit verb causality on pronoun processing. *Psychological Science, 24*, 39–41.

Thomas, M. (1991). Universal grammar and the interpretation of reflexives in a second language. *Language, 67*, 211–239.

Trenkic, D., Mirkovic, J., & Altmann, G. (2014). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition*, *17*, 237–257.

White, L. (1998). Second language acquisition of Binding Principle B: Child/adult differences. *Second Language Research, 14*, 425–439.

Yang, C. L., Gordon, P. C., Hendrick, R., & Wu, J. T. (1999). Comprehension of referring expressions in Chinese. *Language and Cognitive Processes, 14*, 715–743.

Zhang, J. (2002). *The acquisition of English psych predicates by Chinese-speaking learners*. Unpublished doctoral dissertation, Guangdong University of Foreign Studies.

Zufferey, S., Mak, W., Degand, L., & Sanders, T. (2015). Advanced learners' comprehension of discourse connectives: The role of L1 transfer across on-line and off-line tasks. *Second Language Research, 31*, 389–411.

https://doi.org/10.1017/S0142716416000035 Published online by Cambridge University Press