Journal of Pan-Pacific Association of Applied Linguistics, 26(1), 63-85.
https://doi.org/10.25256/PAAL.26.1.4

An Analysis of Data Collection Tasks Used in Studies on Scope Ambiguity in English Involving Universal Quantifiers and Negation*

Hye-Young Kwak**

Kwak, H.-Y. (2022). An analysis of data collection tasks used in studies on scope ambiguity in English involving universal quantifiers and negation. *Journal of Pan-Pacific Association of Applied Linguistics*, *26*(1), 63-85.

This study examines types and characteristics of data collection tasks used in studies on scope ambiguity in English involving a universally quantified noun phrase and negation, and investigates any differences in comprehension patterns across studies using different tasks. Since Musolino's seminal 1998 study using a truth value judgment task, several other studies on English scope ambiguity have used the same task type or variations on it, while others have tested different tasks. Through a keyword search of relevant databases, the present study found a total of 13 studies dealing with the topic of universal quantification and negation in English published since 1998. Based on tasks' key characteristics and how they were conducted, seven different basic task types were identified. A comparison of the results of the studies found a difference in the overall comprehension patterns of English native speakers for sentences containing a universally quantified noun phrase in subject position between the truth value judgment task with one type of context story and the other tasks. The results are discussed in terms of their methodological implications, and some suggestions for further research from methodological perspectives are provided.

Keywords: universal quantifier, negation, scope ambiguity, data collection task, task effect

1 Introduction

In English, universal quantifiers such as *all* and *every* can interact with a negation word in the same sentence to produce ambiguity, as in (1).

(1) Universally quantified noun phrase (NP) in subject position and negation
Every boy didn't cut down the apple tree.

* I am very grateful to the three reviewers for their valuable comments.
** **Hye-Young Kwak**, Associate Professor, English Studies, Korea University, Sejong Campus

a. Meaning 1 (every > not reading)
   None of the boys cut down the apple tree.
b. Meaning 2 (not > every reading)
   Some of the boys cut down the apple tree but others didn't.

The sentence in (1) containing the universal quantifier *every* in subject position is ambiguous, because it can mean that none of the boys cut down the apple tree or that not every boy cut down the apple tree. The former interpretation is called the quantifier wide scope reading (i.e., every > not reading) and the latter is called the negation wide scope reading (i.e., not > every reading).

The topic of scope ambiguity involving universal quantifiers and negation in English was first systematically investigated by Musolino in a 1998 doctoral dissertation. Since then, it has been examined from various perspectives including first language (L1) acquisition, second language (L2) acquisition, and sentence processing. Musolino compared the comprehension of sentences involving quantified NPs and negation by children and adult native English speakers, exploring why children's comprehension patterns are different from those of adults and how children develop the adult-like interpretation system. Building on Musolino's study, later L1 acquisition studies (Conroy, 2008; Musolino & Lidz, 2006) investigated various factors that might influence L1 children's comprehension patterns in comparison to those of adult native speakers of English.

Following the L1 acquisition studies, several L2 acquisition studies (Chung, 2012; Kim, 2010; Lee, 2010a, 2018; O'Grady et al., 2009; Wu & Ionin, 2019) also focused on English scope ambiguity, comparing comprehension patterns of EFL or ESL learners at different English proficiency levels to those of native English speakers. These studies have examined the roles of factors such as learners' native languages, the relative processing difficulty of the two interpretations, and pragmatic ability in English as a second language. One L3 acquisition study (Jo et al., 2021) investigated the comprehension patterns of target sentences by 11- or 12-year-old children with different L1 backgrounds who were learning Korean as a second language and English as a third language, and also examined the nature of transfer in L3 acquisition. A couple of studies (Kim, 2011; Lee et al., 2011) have investigated the possibility of crosslinguistic influences on Korean-English bilingual speakers' interpretive preferences for sentences involving scope ambiguity.

It has been more than 20 years since Musolino's (1998) pioneering work inspired subsequent scholars to investigate the topic of scope ambiguity. Musolino adopted the truth value judgment task (TVJT) developed by Crain and Thornton (1998) to investigate the comprehension of meaning at the sentence level. Many of the follow-up studies on the interpretation of sentences involving quantifiers and negation have likewise employed the TVJT, with some variations. However, others have employed different types of data

collection tasks pointing out the limitations of the TVJT. One of these limitations is that the TVJT does not reveal a participant's degree of certainty regarding the selected interpretation of a target sentence because the participant must choose between true and false responses (Chung, 2012). Another is that semantic relations between the two interpretations of a target sentence may make researchers design a task in such a way that only one of the two interpretations is set up to be true (Conroy, 2008). Every task type has strengths and weaknesses; moreover, task types may influence comprehension patterns. This study therefore compares key features of the different data collection tasks used in the existing studies, and analyzes task effects in their reported results. The aim of the study is to open the opportunity to seek new and better data collection tasks for further research. The research questions are as follows:

1. What data collection tasks have been used in the studies that have investigated the comprehension of scope ambiguity in English involving a universal quantifier and negation? What are the characteristics of the different data collection tasks?
2. Do different data collection tasks influence the results of comprehension patterns of scope ambiguity? In other words, is there a difference in the comprehension patterns of scope ambiguity across studies using different tasks?

2 Background: Musolino's 1998 Study

Musolino (1998) investigated how English-speaking children interpreted sentences containing quantified NPs and negation in comparison to adult native speakers of English. He examined the reasons for the difference in interpretation patterns between the two groups and how English-speaking children mastered the adult-like interpretation system. Sample sentences used in the study are presented in (2).

(2) Musolino's (1998) sample sentences

a. NP containing *every* in subject position
Every horse didn't jump over the fence.
Meaning 1 (not > every reading):
Not every horse jumped over the fence.
Meaning 2 (every > not reading):
None of the horses jumped over the fence.

b. NP containing *every* in object position
The Smurf didn't buy every orange.
Meaning 1(not > every reading):

Not every orange was bought by the Smurf.
(Musolino, 1998, p. 86)

When testing 4- and 5-year-old children's interpretation of the target sentences, Musolino (1998) used the TVJT. He created one type of context story in such a way that one of the two meanings was true and the other meaning was false given the content of the story. For example, for a context story for the target sentence in (2a), *Every horse didn't jump over the fence*, the story was constructed in such a way that the meaning 'not every horse jumped over the fence' was true while the meaning 'none of the horses jumped over the fence' was false. The reason the story was constructed in such a particular way is that from the semantic perspective, there is an entailment relation between the two meanings conveyed by the target sentence. In all the situations that the every > not reading is true, the not > every reading is also true (Musolino & Lidz, 2006). In other words, the every > not meaning entails the not > every meaning, and therefore, if the every > not meaning is true in a particular context story, the not > every meaning is also true so that it is impossible to make the not > every meaning false. In Musolino's context story, three horses considered jumping over various objects. The horses first considered jumping over a barn but decided not to do so because it was too high. Then, the horses considered jumping over a fence, which looked lower than the barn. Two of the horses jumped over the fence while one ended up not jumping over the fence. The sample item with the context story is given in (3).

(3) Sample item: NP containing *every* in subject position
Experimenter: In this story, three horses are talking about how good they are at jumping and they decide to practice by jumping over a barn and a fence.
Horse 1: Let's start practicing by jumping over the barn there!
Horse 2: Great, I'm sure we can do it!
Horse 3: Me too! <The horses start galloping towards the barn but as they get closer they realize that it is much too high for them.>
Horse 1: Wait a minute. This barn is much taller than I thought!
Horse 2: You're right, it is much too high!
Horse 3: I think we'd better not jump, we could end up hurting ourselves.
Horse 1: How about we jump over that fence instead, it looks less tall.
Horse 2: You're right, let's do it.
Horse 3: That's a good idea! <The horses line up in front of the fence. Horse 1 goes first and does a great jump over the fence. Horse 2 follows and does a nice jump over the fence.>
Horse 3: Well, I guess it's my turn now. I don't know though. I hurt my leg the other day and I am not sure it's feeling strong enough to jump over that fence. Maybe I'd better not do it. <Horse 3 ends up not jumping over

the fence.>
Puppet: That was a great story about three horses trying to jump over things and I know what happened: Every horse didn't jump over the fence. (Musolino, 1998, p. 95)

According to the story in (3), the meaning 'not every horse jumped over the fence' is true because two of the horses jumped but one of them did not. On the other hand, the meaning 'none of the horses jumped over the fence' is false because two of the horses jumped. In Musolino's TVJT, one experimenter acted out the stories using toys and props, and the other experimenter used a puppet to deliver the target sentences. The child participants were asked to determine whether the puppet's statement was true or not according to the story they had heard.

According to the results of Musolino's (1998) study, for the negative sentences containing *every* in subject position as in (2a), the children accepted the puppet's statements 7.5% of the time while rejecting the puppet's statements 92.5% of the time. In other words, they accepted the every > not reading far more frequently than the not > every reading. As for the negative sentences containing *every* in object position as in (2b), the children accepted the not > every meaning 85% of the time. In comparison, adult native speakers accepted the same object-*every* statements 100% of the time. Based on the results, Musolino proposed that the grammar of 4- and 5-year-old children differed from that of adults.

3 Research Method

3.1 Studies used in the analysis

In order to find studies for this analysis, several databases were searched with keywords. The databases were Google Scholar, RISS, K-eArticle, and ProQuest Dissertations & Theses Global. The keywords used were "negation scope," "quantifier negation scope," "scope ambiguity," "scope interpretation English," "acquisition of scope ambiguity," "negation scope interpretation acquisition," and "negation scope English." Only studies published after Musolino's (1998) dissertation were considered, and 13 studies were selected: ten journal articles, one dissertation, one conference proceedings article, and one chapter of an edited book. The studies' topics include L1 acquisition of scope ambiguity in English, processing of ambiguous sentences by adult native speakers of English, L2 (or L3) acquisition of scope ambiguity in English, and bilingual language acquisition of scope ambiguity in English.

3.2 Procedure of analysis

The selected studies were classified into five subgroups: L1 acquisition (three studies), sentence processing by adult native speakers of English (one study), L2 acquisition (six studies), children's L3 acquisition (one study), and bilingual acquisition (two studies). All 13 studies were then analyzed in terms of type of task, with the tasks' key features and the data collection method examined in detail.

Next, a comparative analysis was conducted on 10 of the studies. The single L3 acquisition study was excluded from the comparative analysis because there was no object of comparison. The two child bilingual studies were also excluded as their participants' language backgrounds did not correspond to each other, so any comparison would not be valid. The remaining 10 studies were then analyzed in terms of the results of the participants' comprehension patterns of the target sentences. The comprehension patterns were compared across tasks for each of the two sentence positions for the quantified NP (i.e., subject and object) and for each of the three participant groups (L1 child learners of English; adult native speakers of English; and L2 learners of English, who were all L1-Korean). This procedure allowed the study to control for possible effects of native language on comprehension patterns. In other words, for target sentences involving a quantified NP in subject position, the results of L1 child learners of English in each task type were compared; then the results of adult native English speakers in each task type were compared; and then the results of L2 learners of English in each task type were compared. In the case of target sentences involving a quantified NP in object position, the results of adult native English speakers were compared across tasks, and then the results of L2 learners of English were compared. This analysis did not include L1 children because all of the studies that looked at the object position with this participant group used TVJTs.

4 Results

4.1 Types of data collection task

Table 1 lists the 13 studies, with acquisition types, data collection tasks, and the quantified NP position of the target sentences. The studies can be classified into seven types by task: (a) TVJT with one type of context story, (b) TVJT with two types of context story, (c) TVJT combined with self-paced reading task, (d) sentence completion task, (e) incremental verification task, (f) speeded forced choice task, and (g) contextualized acceptability judgment task.

Table 1. Background Information on Studies Involving Universal Quantifiers and Negation in English

Study	Acquisition type	Data collection task	Position of quantified NP
Musolino et al. (2000)	L1 acquisition	TVJT	*Every* in subject position and in object position
Musolino & Lidz (2006)	L1 acquisition	TVJT	*Every* in subject position and in object position
Conroy (2008)	L1 acquisition	(i) Incremental verification task; (ii) Speeded forced choice task; (iii) Sentence completion task	*Every* in subject position
Lee (2010b)	L1 adult sentence processing	TVJT combined with self-paced reading task	*Every* in subject position
O'Grady et al. (2009)	L2 acquisition	TVJT	*All* in object position
Kim (2010)	L2 acquisition	TVJT	*Every* in subject position and in object position
Lee (2010a)	L2 acquisition	TVJT combined with self-paced reading task	*Every* in object position
Lee (2018)	L2 acquisition	TVJT combined with self-paced reading task	*Every* in subject position
Chung (2012)	L2 acquisition	Contextualized acceptability judgment task	*Every* in object position
Wu & Ionin (2019)	L2 acquisition	Context-based acceptability judgment task	*Every* in subject position
Lee et al. (2011)	Child and adult bilingual acquisition	TVJT	*All* in object position
Kim (2011)	Child bilingual acquisition	TVJT	*Every* in subject position and in object position

Jo et al. (2021)	Child L3 acquisition	TVJT	*All* in object position

The following seven subsections describe the key characteristics of each task type and the way the tasks were conducted, along with a summary of one sample study using the task type.

4.1.1 TVJT with one type of context

The TVJT allows researchers to examine whether participants have access to a particular meaning of a target sentence. In the TVJT, context stories are provided before the participants judge the truth values of target sentences. Musolino (1998) and the following L1 acquisition studies adopted TVJTs with one type of context story, in which one of the possible meanings of a target sentence is true and the other is false. For example, Musolino and Lidz (2006) conducted two experiments including a TVJT. The first investigated five-year-old children's comprehension of a target sentence including a quantified NP in subject position, and the second investigated children's comprehension of a target sentence including a quantified NP in object position. For the first experiment, Musolino and Lidz created two conditions by varying the type of test sentence. In one condition, a test sentence consisted of a subject NP including *every* and negation (e.g., *Every horse didn't jump over the fence*). In the other condition, a test sentence consisted of a positive lead-in clause and a clause including a subject NP including *every* and negation (e.g., *Every horse jumped over the log but every horse didn't jump over the fence*). In the first condition, the accompanying context story was constructed similarly to those in Musolino's (1998) study. For example, as in the context story reproduced in Section 2, three horses considered jumping over a barn but did not because of its height; they then considered jumping over a fence, and two out of the three horses jumped successfully while one decided not to jump. This story sets up the not > every meaning as true because two of the horses jumped over the fence and one did not. In contrast, it sets up the every > not meaning as false because two horses did jump over the fence. As in a classic TVJT, in Musolino and Lidz's study, for the child participants, one of the experimenters acted out the context stories by manipulating props and toys, and the other experimenter used the puppet to deliver a target sentence. The control group of adult participants saw a videotaped version. The participants were asked to determine whether the puppet's statements were true or false in light of the story.

In the second condition of the first experiment, the test sentences consisted of an affirmative clause followed by a negative clause including a quantified NP in subject position (e.g., *Every horse jumped over the log but every horse didn't jump over the fence*). As in the first condition, the context story was created in such a way that, for the second part of the test sentence, the not > every meaning was true while the every > not meaning was false.

For example, in one context story, three horses tried to jump over various things to test their jumping skills; all of the horses jumped over a log, and two of the three horses also jumped over a fence.

In Musolino and Lidz's (2006) second experiment, the target sentences contained a quantified NP in object position. They examined the children's comprehension of two types of sentences. One type described what happened in the story (e.g., *The strong guy didn't put every elephant on the table*) while the second type predicted what was going to happen in the story (e.g., *The strong guy won't put every elephant on the table*). As in their first experiment, they used a TVJT in which one type of context story was provided for a test sentence. However, unlike the stories used in the first experiment, the second experiment's context stories were constructed such that the every > not reading was true. For example, in one story, a strong guy decided to put three dogs and three elephants on a large table. He succeeded in putting all the dogs on the table. However, he ended up putting none of the elephants on the table because the elephants were too heavy. In this context, both the every > not reading and the not > every reading are true, due to entailment relations between the two meanings. The reason the stories were constructed in this way was to investigate whether the children had an adult-like pragmatic ability to choose the not > every meaning based on a scalar implicature. Previous studies had shown that, with similar test sentences describing what had happened, adult native speakers of English would assign the not > every meaning instead of the every > not meaning, because they understand the speaker's choice to use the weaker *not every* over the equally possible but stronger *none* to show the speaker's intention to express the not > every meaning. That is, in the absence of the word *none*, adult native speakers would infer the not > every meaning. This type of reasoning is called scalar implicature and is a pragmatic ability (Musolino & Lidz, 2006, p. 821). Musolino and Lidz compared children's comprehension of these test sentences with their comprehension of test sentences predicting what was going to happen in a given story because adults do not calculate scalar implicatures for predictions. The child participants were told to determine either whether the puppet's description of what happened in the given story or the puppet's prediction about what would happen based on the story was correct.

4.1.2 TVJT with two types of context

The L2, L3, and bilingual acquisition studies under analysis also used TVJTs (Jo et al., 2021; Kim, 2010; Kim, 2011; Lee et al., 2011; Lee, 2010a, 2018; O'Grady et al., 2009) in order to examine the interpretive preferences of participants. These studies, unlike the L1 studies, used two different types of context story. One type was constructed in such a way that the every > not reading was true and the not > every reading was false. The second type of context story was constructed in such a way that the not > every reading was

true and the every > not reading was false. These studies assumed that in a context story where the every > not reading was set up to be true, the not > every reading could be set up as false, regardless of the semantic entailment relation between the two readings. For the test sentences involving quantified subject NPs, the context stories favoring the every > not reading employed three characters, none of whom succeeded in performing a target action. In the context stories favoring the not > every reading, one or two out of three characters succeeded in performing a target action. Similarly, for the test sentences involving quantified object NPs, in the context stories favoring the every > not reading, a main character did not conduct a target action affecting any of three objects, while in those favoring the not > every reading, the main character performed the target action on one or two of three objects. By comparing the proportions of true responses in each context, these studies investigated which interpretations the participants favored.

The L2 studies had adult participants, and rather than using puppets and acting out the context stories, they presented the context stories in various forms: as a series of pictures on a laptop computer; in written form on a laptop computer; and in written form with a prerecorded audio version of the stories, followed by summary pictures describing main events of the stories, on a paper questionnaire. At the end of the stories, the L2 learner participants were asked to determine the truth value of a target sentence based on the context story. The bilingual and L3 acquisition studies with child participants presented the context stories orally or in written form with relevant pictures on a laptop computer and had a puppet deliver the target sentences to the children. In response to the puppet's statements, the children determined the truth value of the target sentences.

For example, Kim (2010) investigated comprehension patterns of English negative sentences involving a quantified NP in subject position or a quantified NP in object position by two groups of L1-Korean learners of English. For the test sentences involving a quantified subject NP, she created two types of context story. In one of the context stories favoring the every > not reading (i.e., the every > not reading was set up to be true), three horses successfully jumped over a block, but none of the horses jumped over a fence. In contrast, in one of the context stories favoring the not > every reading (i.e., the not > every reading was set up to be true), two out of the three horses successfully jumped over the fence. The context stories were presented as a series of pictures on a computer. At the end of each story, the participants determined whether the target sentence, which was said by the experimenter, was true or not.

4.1.3 TVJT combined with a self-paced reading task

Some of the L2 studies used a self-paced reading task along with the TVJT, enabling them to examine which meaning of an ambiguous sentence causes longer reading times, which indicate more processing difficulty, and to

investigate the point at which processing difficulty occurs in real time. For example, Lee's (2018) study used both task types. For test sentences involving a quantified NP in subject position, she created two types of context story: stories favoring the every > not reading and stories favoring the not > every reading. Each story was presented as a whole on a computer screen, and then a test sentence was presented region by region, with each region appearing when the participant pressed a button on a computer keyboard. The reading times for each region (times between button presses) were measured in milliseconds. At the end of the test sentence, the participants determined the truth value of the sentence, given the context story, by pressing a key labeled "true" or "false."

As an example, one experimental item's test sentence was: *According to the story, every student didn't read the books in the library*. This sentence was presented after either a context story favoring the every > not reading or a context story favoring the not > every reading. In the context story favoring the every > not reading, any of three students did not read the book in the library. In contrast, in the context story favoring the not > every reading, only one of three students read the book in the library. The test sentence was divided into six regions, as shown in (4), and each region appeared on the computer screen when the participant pressed a button, and disappeared with the next button press, when the following region appeared.

(4) According to the story, / every / student /
(Region 1) (Region 2) (Region 3)
didn't read / the books / in the library.
(Region 4) (Region 5) (Region 6)

Lee (2018) examined participants' true responses for each type of context story and the reading times for each region of the test sentences. The analysis of reading times reveals at which region of the test sentence longer reading times occur. For example, the participants showed longer reading times at Region 5, which contained the direct object NP, in the context favoring the not > every reading than in the context favoring the every > not reading. This reading time difference indicates that the participants had more processing difficulty when they read the direct object NP after the negated verb of a test sentence in the context story favoring the not > every reading.

4.1.4 Sentence completion task

While most of the L1 acquisition studies adopted the TVJT to test children's comprehension patterns of target sentences, Conroy (2008) utilized a sentence completion task. The sentence completion task seems to enable participants to choose the interpretation of the target sentence that first comes to mind after they hear a target sentence; unlike the TVJT, it offers little or no opportunity for participants to employ contextual information to revise an initial, default

interpretation.

Conroy (2008) presented children with a picture and an accompanying story on a computer. The story allowed both every > not and not > every meanings to be true. For example, one of the stories entailed two differently colored fire hydrants (yellow and blue) to be decorated by the characters; the characters were three differently colored dogs (i.e., purple, red, and green) that had lollipop decorations and a fireman with a hose decoration. None of the dogs decorated the blue fire hydrant, but the fireman did decorate it, using the hose decoration. Two of the three dogs decorated the yellow fire hydrant with their lollipops, but one dog did not do so. After the story was presented, the children were told to complete a sentence, *Every dog didn't decorate the fire hydrant that's_* by choosing one of the two hydrants (e.g., blue versus yellow). They were told that the puppet was trying to describe what happened in the story but did not know the names of colors well, so they should help the puppet complete the sentence. The choice of the "blue" response indicated their access to the every > not reading ('none of the dogs decorated the blue fire hydrant') whereas the choice of the "yellow" response indicated their access to the not > every reading ('some of the dogs decorated the yellow fire hydrant').

4.1.5 Incremental verification task

Conroy (2008) also created an incremental verification task to overcome limitations of the TVJT. In the TVJT, due to the entailment relations of the two target readings, some researchers used only one type of context story where the not > every reading was true and the every > not reading was false. To resolve the entailment problem, the incremental verification task enables researchers to identify which interpretation a participant had access to, even in the case where the every > not reading was true. Unlike the TVJT, the incremental verification task does not involve any context story that might help a participant accept a particular interpretation of a target sentence. In addition, the task might minimize a participant's use of discourse information that could help the participant revise a first interpretation. According to Conroy, discourse information is "information about the speaker's intentions and goals and any other pragmatic information" (p. 10).

Conroy (2008) used the incremental verification task to examine adult native English speakers' comprehension of negative sentences containing subject quantified NPs. In the test items, a participant looked at a series of four cards containing pictures that were covered by cups. After an experimenter uttered a test sentence containing a quantified NP in subject position and negation, the participant was asked to lift as few cups as possible to decide whether the test sentence was true or false by checking the pictures under the cups. In one type of test item, the pictures under the cups were such that both the every > not reading and the not > every reading were true. In the other type of test item, the pictures under cups revealed that the every > not reading was false, while the not > every reading was true. As an example, the test sentence

Every dog isn't wearing a hat was used either with the set of four pictures shown in Figure 1 or with the set of four pictures shown in Figure 2. In the figures, the rows illustrate a progression of a trial, in which one cup has been removed in (a), two cups removed in (b), and four cups removed in (c).

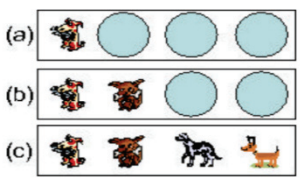

Figure 1. Sample item for incremental verification task (Conroy, 2008, p. 45) To verify that the not > every reading is true, only the first cup must be removed as in (a); to verify that the every > not reading is true, all four cups must be removed as in (c).

In Figure 1, the picture under the first cup is sufficient to verify that the not > every reading is true, because one dog isn't wearing a hat, as in (a). However, to verify whether the every > not reading is true, the participant must remove all four cups to check that none of the four dogs is wearing a hat, as in (c). In sum, on the first cup, the not > every reading can be verified, while only on the last cup can the every > not reading be verified.

The same test sentence was also presented with a series of pictures revealing that the every > not reading was false, as shown in Figure 2.

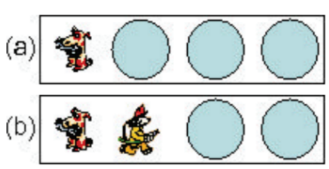

Figure 2. Sample item for incremental verification task (Conroy, 2008, p. 45) To verify the not > every reading as true, only the first cup must be removed as in (a); to verify the every > not reading as false, two of the four cups must be removed as in (b).

In the pictures shown in Figure 2, the not > every reading is verified as true on the first cup. The every > not reading is verified as false on the second cup because the picture under the second cup shows a dog wearing a hat. The participant's true or false responses, and the cup where the participant stopped were analyzed to indicate which of the two interpretations was considered and selected by the participant.

4.1.6 Speeded forced choice task

Conroy (2008) tested yet another task, the speeded forced choice task, to examine which of the interpretations adult native speakers of English obtained for test sentences involving a subject quantified NP. As in the incremental verification task, the intention of this task was again to eliminate the possibility that a participant could use discourse information to revise an initial interpretation.

In the speeded forced choice task, each participant looked at a picture on a computer screen while listening to a prerecorded audio story. The participant was then instructed to choose one of two options to complete a test sentence that described the story as quickly as possible. Figure 3 is an example picture for one of the test items. The picture in the figure was presented with a story in which two out of three dwarfs spraypainted the cow's barn, none of the dwarfs spraypainted the pig's barn, and a farmer spraypainted the pig's barn.

Figure 3. Sample item for speeded forced choice task (Conroy, 2008, p. 77)

After the story was presented, the participant saw two options on the computer screen: a picture of a cow's barn and a picture of a pig's barn. The two options stayed on the screen for 1400 milliseconds, after which a fixation point appeared for 750 milliseconds. Then, the options reappeared, and the test sentence to complete was presented orally. The test sentence to complete was *Every dwarf didn't spraypaint the* ___ or *Every dwarf didn't spraypaint the barn that belongs to the*__. The participant chose one of the two options by pressing a button on the keyboard.

This task is similar to the sentence completion task in which a participant is told to complete a sentence in the first way that comes to mind. Both tasks are intended to prevent a participant from revising an initial interpretation.

4.1.7 Contextualized acceptability judgment task

Some L2 studies have used contextualized acceptability judgment tasks to overcome some limitations of the TVJT in order to examine comprehension

patterns of sentences including negation and a quantified NP in English. Unlike the TVJT, this task asks participants to rate the acceptability of the test sentence as a description for the given context story on a scale, thus allowing the participants to indicate the degree of their certainty of their judgment.

Chung (2012) investigated the comprehension patterns of L1-Korean L2 learners of English for sentences involving a quantified NP in object position. Chung pointed out that the "binary judgments used in the TVJT cannot accurately capture the certainty or gradience of judgments" (p. 295). Therefore, she used a contextualized acceptability judgment task in which participants read a short text presented with a picture and evaluated whether a test sentence was acceptable as a description of the text on a four-point scale (i.e., 1 = unacceptable; 2 = not very acceptable; 3 = somewhat acceptable; 4 = acceptable).

Chung (2012) created texts representing two types of contexts. One context favored the not > every reading, while the other favored the every > not reading. For example, for the context of one test item favoring the not > every reading, one day, a main character missed breakfast and lunch and only ate dinner. Based on the context, participants rated the acceptability of the test sentence *Mary didn't eat every meal*. In contrast, for the context favoring the every > not reading, the main character missed all three meals on the day of the story, and the test sentence *Scott didn't eat every meal* was given.

4.2 Effect of type of data collection task on comprehension patterns

4.2.1 Quantified NP in subject position

Studies examining L1 children's interpretation of quantified NPs in subject position have used the TVJT (three studies) and the sentence completion task (one study). These studies have reported similar overall comprehension patterns in both tasks (see Table 2): The children accepted the every > not reading more frequently. However, the acceptance rates of the not > every reading varied between the studies, even between the three using the TVJT.

Table 2. Quantified NP in Subject Position: L1-English Children's Comprehension Results

Study	Task	Every > not reading	Not > every reading
Musolino et al. (2000): Experiment 2	TVJT (one type of context story)	Far more accepted (92.5%)	Less accepted (7.5%)
Musolino & Lidz (2006): Experiment 1; Condition 1	TVJT (one type of context story)	Far more accepted (85%)	Less accepted (15%)

Conroy (2008)	TVJT (one type of context story)	More accepted (55%)	Less accepted (45%)
Conroy (2008)	Sentence completion task	More accepted (58%)	Less accepted (42%)

Six studies investigated adult native English speakers' comprehension patterns for sentences with quantified NPs in subject position. These include studies focusing on L2 adults, but which have L1 adult control groups. These studies used seven different task types: the TVJT with one type of context story, the TVJT with two types of context story[1], the TVJT with a self-paced reading task, the sentence completion task, the incremental verification task, the speeded forced choice task, and the context-based acceptability judgment task (see Table 3).

The TVJT with one type of context story elicited different comprehension patterns than the other tasks (see Table 3). In the TVJT with one type of context story, the not > every reading was fully or far more accepted. In contrast, in the TVJT with two types of context story, the sentence completion task, the incremental verification task, the speeded forced choice task, and the TVJT with a self-paced reading task, the every > not reading was more accepted. Similarly, the results of the context-based acceptability judgment task revealed that the every > not reading was judged as somewhat acceptable and that the not > every reading was judged as not very acceptable.

In addition, the acceptance rates of the not > every reading were lower in the sentence completion task, incremental verification task, speeded forced choice task, and TVJT combined with a self-paced reading task than in the TVJT alone. This finding indicates that the tasks with time pressure might prevent participants from utilizing discourse information that they could use to revise their initial interpretation.

[1] In the TVJT studies involving two types of context story, the participants' comprehension patterns were analyzed by comparing the proportions of true responses in two types of context: one favoring the every > not reading (i.e., the every > not reading is set up to be true) and another favoring the not > every reading (i.e., the not > every reading is set up to be true). In the subsequent tables summarizing the results from the TVJT studies with two types of context story, the percentage rate in parenthesis in the column for the every > not reading was converted from the percentage of true responses in the contexts favoring the every > not reading, and the percentage rate in the column for the not > every reading was converted from the percentage of true responses in the contexts favoring the not > every reading.

Table 3. Quantified NP in Subject Position: L1-English Adults' Comprehension Results

Study	Task	Every > not reading	Not > every reading
Musolino et al. (2000): Experiment 2	TVJT (one type of context story)	Not accepted (0%)	Fully accepted (100%)
Musolino & Lidz (2006): Experiment 1; Condition 1	TVJT (one type of context story)	Less accepted (7.5%)	Far more accepted (92.5%)
Conroy (2008)	TVJT (one type of context story)	Less accepted (25.9%)	Far more accepted (74.1%)
Conroy (2008)	Sentence completion task	More accepted (60%)	Less accepted (40%)
Conroy (2008)	Incremental verification task	Far more accepted (77.1%)	Less accepted (22.9%)
Conroy (2008)	Speeded forced choice task	Far more accepted (81.5%)	Less accepted (18.5%)
Kim (2010)	TVJT (two types of context story)	More accepted (100%)	Less accepted (87.5%)
Lee (2010b)	TVJT with a self-paced reading task (two types of context story)	Far more accepted (70.8%)	Less accepted (36.9%)
Wu & Ionin (2019)	Context-based acceptability judgment task	Somewhat acceptable (about 3.4)	Not very acceptable (2.5)

As for the results from the two studies with adult L1-Korean L2 learners of English, the TVJT and the TVJT combined with the online self-paced reading task elicited similar comprehension patterns (see Table 4). Overall, the every > not reading was accepted more than the not > every reading. In addition, the L2 learners with higher English proficiency tended to accept the not > every reading more frequently than the L2 learners with lower English proficiency.

Table 4. Quantified NP in Subject Position: L1-Korean L2-English Adults' Comprehension Results

Study	Task	Proficiency level	Every > not reading	Not > every reading
Kim (2010)	TVJT (two types of context story)	Intermediate	Far more accepted (91.67%)	Less accepted (5%)
		Advanced	More accepted (100%)	Less accepted (63.33%)
Lee (2018)	TVJT with a self-paced reading task (two types of context story)	Low	Far more accepted (92%)	Less accepted (34%)
		High	More accepted (88%)	Less accepted (69%)

4.2.2 Quantified NP in object position

In the four studies that examined the interpretations of adult native speakers of English for quantified NPs in object position (both L1 and L2 studies), the TVJT (three studies) and contextualized acceptability judgment task (one study) showed similar comprehension patterns (see Table 5). In the TVJTs, the not > every reading was accepted far more than the every > not reading, showing a strong preference toward the not > every reading. Similarly, in the contextualized acceptability judgment task, the target sentences in the contexts favoring the not > every reading were rated as acceptable while the target sentences in the contexts favoring the every > not reading were rated as not very acceptable; target sentences were interpreted more readily as having the not > every reading than the every > not reading.

Table 5. Quantified NP in Object Position: L1-English Adults' Comprehension Results

Study	Task	Every > not reading	Not > every reading
Musolino & Lidz (2006): Experiment 2	TVJT (one type of context story)	Less accepted (20%)	Far more accepted (80%)
Kim (2010)	TVJT (two types of context story)	Less accepted (22.5%)	Far more accepted (97.5%)
Lee (2010a)	TVJT with a self-paced reading task (two types of context story)	Less accepted (45.14%)	Far more accepted (90.63%)

Chung (2012)	Contextualized acceptability judgment task	Not very acceptable (2.36)	Somewhat acceptable (3.49)

The four studies with L1-Korean L2 learners of English as participants each had a different method of measuring participants' English proficiency, making it difficult to directly compare the results of the studies. Overall, however, learners with lower proficiency chose the every > not reading more frequently than the not > every reading in the TVJT and the TVJT combined with a self-paced reading task. Similarly, in the contextualized acceptability judgment task, learners with lower proficiency rated the target sentences as somewhat acceptable in the context favoring the every > not reading and as unacceptable or not very acceptable in the context favoring the not > every reading.

Table 6. Quantified NP in Object Position: L1-Korean L2-English Adults' Comprehension Results

Study	Task	Proficiency level	Every > not reading	Not > every reading
Kim (2010)	TVJT (two types of context story)	Intermediate	More accepted (91.67%)	Less accepted (83.33%)
		Advanced	More accepted (98.33%)	Less accepted (86.67%)
Lee (2010a)	TVJT with a self-paced reading task (two types of context story)	Low	Far more accepted (86.58%)	Less accepted (37.51%)
		High	Accepted[2] (78.53%)	Accepted (75.17%)
Chung (2012)	Contextualized acceptability judgment task	Low intermediate	Somewhat acceptable (3.53)	Unacceptable (1.59)
		High intermediate	Somewhat acceptable (3.36)	Not very acceptable (2.06)
		Advanced	Not very acceptable (2.07)	Not very acceptable (2.87)
O'Grady et al. (2009)	TVJT (two types of context story)	Intermediate	Far more accepted (93%)	Less accepted (28%)

[2] Lee (2010a) reported no significant difference in the rates of acceptance between the two readings.

5 Discussion and Conclusion

This study has analyzed the results of the seven types of data collection task employed in studies focusing on the comprehension patterns of scope ambiguity involving universal quantifiers and negation in English. The TVJT presents context stories in which one of the interpretations of a target sentence is true and the other interpretation is false. The task allows a researcher to examine whether a participant does or does not have access to a certain interpretation of a target sentence or to examine which interpretation of a target sentence is preferred. The sentence completion task and the speeded forced choice task encourage participants to select the interpretation of an ambiguous sentence that comes first to mind after hearing the sentence. The incremental verification task makes two interpretations of a target sentence available and allows a participant to verify the interpretations incrementally in real time. It may indicate which interpretation of a target sentence a participant is considering without any context provided. The contextualized acceptability judgment task makes a participant determine whether a target sentence is an acceptable description of the event described in an accompanying context story by using a four-point scale. It is similar to the TVJT in terms of the use of context stories, but it allows a participant to rate acceptability, rather than having to select between true and false responses. The self-paced reading task combined with the TVJT makes it possible to examine which interpretation of an ambiguous sentence causes longer reading times, which indicate more processing difficulty, and to investigate at which part of the sentence processing difficulty occurs in the process of interpreting the sentence in real time.

The results also revealed differences in the comprehension patterns of sentences containing universal quantification in subject position among different data collection tasks. In the TVJT with one type of context story, native English speakers assigned the not > every reading to test sentences more frequently. However, in the two-context-story TVJTs, whether alone or combined with the self-paced reading task, native English speakers assigned the every > not reading to test sentences more frequently. Similarly, in the sentence completion task, the incremental verification task, and the speeded forced choice task, native English speakers assigned the every > not reading to test sentences more frequently. In the context-based acceptability judgment task, native English speakers also judged test sentences as more acceptable descriptions of the contexts when the sentences were presented with contexts favoring the every > not reading than when the contexts favored the not > every reading.

These results suggest that there may be an effect of task type on the comprehension patterns of sentences containing universal quantification and negation. In the TVJT with one type of context story, the not > every reading was set up as true and the every > not reading was set up as false. As Conroy

(2008) pointed out, the "principle of charity" of the TVJT, which assumes that people will assent to the interpretation of ambiguous sentences that is made true in the context, may have led to higher acceptance rates of the not > every reading, whether or not this reading was the participant's preferred interpretation. In the TVJT with two types of context story, each interpretation was set up as true in different context stories, and the results revealed that native English speakers preferred assigning the every > not reading to target sentences. In the TVJT with one type of context story, the inclusion of context stories where the every > not reading is true is avoided due to the entailment problem. However, it seems that it is not realistic to expect lay persons to interpret target sentences as having the not > every interpretation at first, and then to consider a context story where the every > not interpretation is designed to be true, and further to think that in all situations that the every > not reading is true, the not > every reading is always true and therefore to judge the target sentence as true. For example, in Kim's (2010) L2 acquisition study testing the comprehension of target sentences including quantified object NPs, English native speakers provided "true" responses 22.5% of the time and "false" responses 77.5% of the time in the context stories where the every > not reading was designed to be true; and they provided "true" responses 97.5% of the time and false" responses 2.5% of the time in the context stories where the not > every reading was designed to be true. These results suggest that they may not have thought of the logical relationship between the two interpretations when they judged the truth values of the target sentences presented with the context stories where the every > not reading was true. They may have simply thought that given that the main character performed an action on none of the objects in the context stories, the not > every interpretation (i.e., that the main character performed an action on some of the objects, but not on others) was false. It seems that for lay persons, the entailment problem may not matter when they are engaged in the TVJT. In order to examine which interpretation is both possible and preferred, it would be better to include two types of context stories in the TVJT. One way to make sure that a participant has access to the not > every reading in the context stories where the every > not reading is designed to be true would be to conduct a follow-up interview where participants are asked to justify their responses; if participants explained that the main character performed an action on some of the objects and not on others after choosing "false" responses, it would indicate they have access to the not > every interpretation.

The existing data collection tasks differ in terms of allowing participants to use discourse or pragmatic information in determining the interpretation of a target sentence. For this reason, combining two or more data collection tasks is a more effective way of investigating how participants use various types of information, including discourse information, in comprehending a target sentence and how they finally choose between two interpretations of a target sentence. In addition, the data collection tasks analyzed in this study do not

reveal how participants interpret sentences in real time (i.e., while hearing a target sentence from beginning to end). In order to examine whether and how participants consider an initial interpretation that they then revise in the course of the comprehension process, future research should consider employing online processing measures, such as eye-tracking.

This study focused on the analysis of key features of different data collection tasks and the possibility of task effects on the results of the interpretation of scope ambiguity in English, not on factors that might influence the comprehension patterns of a particular participant group or the differences in the comprehension patterns between different participant groups. The overall findings highlight the importance of selecting the right type of data collection task(s), which depends on the nature of the factors investigated and the characteristics of particular participant groups. It is hoped that this overview of key characteristics of different data collection tasks, the way each task has been conducted, and the results from the studies using different data collection tasks will provide valuable insights as to the choice of data collection tasks in further studies.

References

Chung, E. S. (2012). Sources of difficulty in L2 scope judgments. *Second Language Research*, *29*(3), 285-310.

Conroy, A. M. (2008). *The role of verification strategies in semantic ambiguity resolution in children and adults* (Publication No. 3339372) [Doctoral dissertation, University of Maryland, College Park]. ProQuest Dissertations & Theses Global.

Crain, S., & Thornton, R. (1998). *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics.* MIT Press.

Jo, K., Kim, K., & Kim, H. (2021). Children's interpretation of negation and quantifier scope in L3 English. *Bilingualism: Language and Cognition*, *24*, 427-438.

Kim, H. (2011). Access to inverse scope in Korean-English bilingual children. *Korea Journal of English Language and Linguistics*, *11*(2), 415-435.

Kim, M. (2010). Korean EFL learners' interpretation of quantifier-negation scope interaction in English. *English Language and Linguistics*, *16*(1), 163-183.

Lee, M., Kwak, H.-Y., Lee, S., & O'Grady, W. (2011). Processing, pragmatics, and scope in Korean and English. *Japanese Korean Linguistics*, *19*, 297-314.

Lee, S. (2010a). Interpretation of scope by Korean L2 learners of English: A self-paced reading study. *English Teaching*, *65*(4), 59-78.

Lee, S. (2010b). Quantifier scope in English: The view from sentence

processing. *English Language and Linguistics*, *16*(3), 163-186.
Lee, S. (2018). Semantic processing of scope ambiguity by Korean L2 learners of English. *The Journal of Studies in Language*, *33*(4), 695-713.
Musolino, J. (1998). *Universal grammar and the acquisition of semantic knowledge: An experimental investigation into the acquisition of quantifier-negation interaction in English* (Publication No. 9908995) [Doctoral dissertation, University of Maryland, College Park]. ProQuest Dissertations & Theses Global.
Musolino, J., Crain, S., & Thornton, R. (2000). Navigating negative quantificational space. *Linguistics*, *38*(1), 1-32.
Musolino, J., & Lidz, J. (2006). Why children aren't universally successful with quantification. *Linguistics*, *44*(4), 817-852.
O'Grady, W., Lee, M., & Kwak, H.-Y. (2009). Emergentism and second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 69-88). Emerald Group Publishing Limited.
Wu, M.-J., & Ionin, T. (2019). L1-Mandarin L2-English speakers' acquisition of English universal quantifier-negation scope. In M. B. Brown & B. Dailey (Eds.), *Proceedings of 43rd Annual Boston University Conference on Language Development* (pp. 716-729). Cascadilla Press.

Hye-Young Kwak, Associate Professor
English Studies, Korea University, Sejong Campus
2511 Sejong-ro, Sejong, 339-700, Korea
Phone: +82-44-860-1237
E-mail: hyeyoungku@korea.ac.kr

Received: May 31, 2022
Revised: July 4, 2022
Accepted: July 14, 2022