The Efficacy of Uptake Response Times as a Measure of Noticing of Corrective Feedback during Oral Interaction Feedback Episodes

MOXON, Jonathan

Abstract

Interaction is regarded as being facilitative of second language (L2) development as it provides learners with opportunities to notice their errors, become aware of the gap between their current L2 knowledge and the target language, and bring about restructuring of current interlanguage (Long, 1996). Central to interaction-driven L2 learning are *noticing* and *noticing the gap*; however, current methodologies to operationalize these cognitive processes suffer from drawbacks that limit their ability to accurately quantify incidence. In the current small-scale exploratory study three subjects took part in an interaction activity during which corrective feedback was provided in response to learners' errors, resulting in N = 42 feedback episodes. Two widely used measures—*uptake* (Lyster, 1998) and self-reports of noticing during *stimulated recalls* (Gass & Mackey, 2013)—were employed to quantify noticing. In addition, data obtained by a novel approach that examined response time between feedback provision and uptake was triangulated with uptake and stimulated recall data. Results suggest that supplementing analysis of response times to other analyses may be useful in disambiguating the incidence of noticing in feedback episodes where traditional operationalizations may result in false positives.

Introduction

A vigorous and extensive body of research in the second language (L2) learning literature is concerned with how *noticing* of language features relates to various aspects of the acquisition process. Essential to these investigations is the valid and reliable operationalization of internal processes that do not lend themselves to direct observation. In oral interaction research, two main methodologies are currently available to the researcher: *stimulated recall* and *uptake*. There are, however, drawbacks to both of these approaches that potentially impose limitations on the final accuracy of quantifications of noticing. In an attempt to address this issue, this small-scale exploratory study describes a novel approach to operationalizing noticing that employs measured response times to assist with distinguishing positive cases of noticing from false positive cases of such.

Interaction-Driven L2 Acquisition

It has been argued that participation in conversational interaction can facilitate L2 development (Gass & Mackey, 2007; Long, 1996; Pica, 1994). Interaction between a learner and a more expert interlocutor (often a native speaker), jointly focused on a task requiring the transmission or exchange of information, provides a meaning-centred environment within which feedback can be given on the learner's errors and mistakes—so-called *focus-on-form*. Long's (1996) Interaction Hypothesis proposes that interaction is a rich learning ecology because feedback episodes provide a nexus for learners to receive input, highlight deficiencies in current L2 knowledge in ways that engage their cognitive capacities and attentional resources, and produce output in response to corrective feedback. This focus-on-form "involves…an occasional shift in attention to linguistic code features…triggered by perceived problems with comprehension or production" (Long, 1998, p. 23).

For these interventions to be pedagogically successful, it is incumbent on the learner to perceive the corrective feedback and switch the focus of attention from meaning to form. In order for this switch of attentional resources to occur, learners must first *notice* when corrective feedback has been provided. Schmidt's Noticing Hypothesis (Schmidt, 1990, 1995) posits that no learning can take place unless learners become consciously aware of some language feature. For Schmidt, there can be no acquisition without awareness of form. This may in turn set in motion a deeper level of conscious processing where a learner draws comparisons between his original utterance and the errors indicated in the corrective feedback received—theoretically referred to as *noticing the gap* (Gass, 1997; Schmidt & Frota, 1986)—potentially leading to restructuring of current L2 knowledge to something that more closely approximates target language norms. At this stage of the feedback episode learners may follow up with a verbal response to the feedback. This verbal response is referred to as *uptake* (Lyster & Ranta, 1997) and is also theorized to play a key role in interlanguage development (Izumi, 2003; McDonough, 2005; Swain 1985).

Operationalizing Noticing

As we have seen, noticing and in particular noticing the gap is theorized to facilitate learners' L2 development, and in terms of this a body of research has emerged which examines learners' noticing of corrective feedback under the rationale of illuminating the processes at work in interaction-driven learning. These studies have generally applied two types of methodology to operationalize and measure incidences of noticing: *introspective protocols,* and *uptake*.

Introspective protocols offer structured opportunities for learners to self-report on their experience of the interaction. These self-reports may be given simultaneously with the interaction activity (*think alouds*) or retrospectively. Because think aloud protocols, by their nature, are difficult to use during oral interaction, a retrospective introspection protocol known as *stimulated recall* (Gass & Mackey, 2013) is the most commonly used protocol in the L2 interaction literature. In research

employing this protocol (e.g. Bao, Egi & Han, 2011; Gurzynski-Weiss & Baralt, 2015; Mackey, 2006; Mackey et al., 2000; Mackey et al, 2012; McDonough, 2006), learners report on their thoughts, recollections, and feelings as they listen to audio or watch video recordings of the interaction. Asking subjects to remember and verbalize their state of mind at the point of corrective feedback provides an opportunity to retrospectively identify incidences of noticing and noticing the gap.

Uptake refers to a learner's immediate verbal response to corrective feedback (Lyster, 1998; Panova & Lyster, 2002; Sheen, 2008), and as such is a feature present for observation in the primary data itself. It may take many forms, from 1) a simple acknowledgement of the feedback, to 2) a modification of the learner's original utterance, to 3) a form meeting target language norms. In its simplest form, this methodology identifies incidences of noticing by the presence or absence of uptake.

These two methodologies, introspective protocols and uptake, provide differing perspectives on noticing. Stimulated recall seeks to unpick the processes of noticing through offline and subjective self-assessment of past experience. In contrast, uptake analysis takes the learner's response as a performance indicator of internal processes, presumably noticing, that occurred directly before the follow-up verbal response to feedback which comprises the uptake. Both of these approaches present advantages and disadvantages, and may offer differing and even contradictory evidence of whether noticing has indeed occurred. From a psychometric point of view this presents as a problem, because if the data provided by each methodology is presumed to provide a separate or independent measure of the same underlying phenomenon, that is, the internal processes attending noticing, then a lack of correspondence between the measures represents a form of measurement error. Obviously, pinpointing the origin of the error is the critical issue.

With respect to retrospective reports, a report of receiving feedback is a good indication that the learner did in fact focus attention sufficiently on the input to notice its intent and content: and the ability of the learner to verbalize *post hoc* his experience of noticing during stimulated recall indicates quite strongly that the episode was rehearsed sufficiently to leave an accessible long-term memory representation sufficient for stimulated recall. However, these memories may be subject to decay due to the delay between the instance of noticing during the interaction and the attempt to access it during the retrospective report. Presentation of video footage of the interaction to stimulate these decaying memory traces may still not be sufficient, leading to underreporting of noticing. In addition, self-reports may be dependent on learners' metacognitive skills, potentially constraining their ability to communicate noticing which had in fact occurred.

Uptake presents different problems as an operationalization or measure of noticing. Because uptake is a performance indicator, it gives no direct information about the internal processes that resulted in the response comprising the uptake utterance. As a consequence, the meaning of uptake may be ambiguous, as the same type of uptake may signify a range of internal processes not apparent in the learner's utterance. If a learner's uptake is as simple as a "yes" in response to feedback, we have no way of ascertaining whether the learner is indicating that he or she

understands his or her interlocutor's intervention as a correction of an error, or (particularly in the case of recasts) he or she has simply understood the feedback as an utterance expressing agreement or understanding of the learner's utterance (Panova & Lyster, 2002). More pertinently, even a modified form of output that repairs the utterance to a target-like form may be a simple 'shadowing' of the recast without any deeper processing associated with noticing or noticing the gap (Gurzynski-Weiss & Baralt, 2015). It can further be remarked that uptake offers little information about the 'depth' of the learner's internal processes in the formulation and articulation of uptake. The noticing and the noticing the gap phenomena, theorized to be central to interaction-driven L2 learning, are not directly accessible through the performance data that uptake provides.

Nassaji (2016) offers a succinct summary of this issue:

> Uptake may indicate that the learners have noticed the feedback, but it does not indicate that they have learned from it or even processed it. Even successful uptake (repair) may not indicate that language acquisition has taken place. It is quite possible that such repair may be a simple mechanical repetition of the teacher's feedback. (p. 104)

As mentioned earlier, employing these two methodologies to operationalize noticing, and even to the same data, often gives differing results. Egi (2010), for example, reported cases in her data where uptake occurred without retrospective reports of noticing. Conversely, Bao, Egi and Han (2011) reported that noticing was significantly more frequent when it was measured with stimulated recall than with uptake, confirming what the average teacher already knows: learner noticing is not always concurrent with the presence of any form of uptake. Gurzynski-Weiss and Baralt (2015) reported a similar phenomenon in their investigation of the relationship between noticing and mode of uptake among learners of Spanish in face-to-face and computer-mediated conversation, where almost 10% of modified output was not accompanied by retrospective reports of noticing.

To further muddy the waters, the relationship between noticing as observed through stimulated recall and uptake may be dependent upon the kind of uptake produced. Although much research employing uptake as a measure of noticing, especially early research, makes no distinction between the types of uptake described above (Lyster, 1998; Lyster & Ranta, 2003; Panova & Lyster, 2002), more recent studies suggest that type of uptake is analytically important. Egi (2010), for example, examined the relationship between stimulated recall and type of uptake in a study of adult learners of Japanese, finding that learners' self-reports of noticing were a significant predictor of uptake that contained some form of repair of the initial utterance. In Gurzynski-Weiss and Baralt (2015), regression analysis indicated that only partially modified output (uptake containing modifications to the original utterance without repair to fully target-like form) was a significant predictor of self-reported noticing. In other words, not all uptake is equal: uptake including some modification of the original utterance appears to be associated with retrospective verbalization of

the experience of noticing.

In summary, the tools available to the researcher when investigating noticing as a phenomenon during interactionally-situated feedback episodes are, to some extent, problematic. Stimulated recall offers perhaps the clearest window on learners' online processes, but its retrospective and offline perspective make it subject to issues associated with memory decay and concomitant worries of under-reporting, reliability, and metacognitive constraints. Uptake is more convenient as it avoids an extra level of data collection and analysis in being an online performance measure of noticing, but cannot directly indicate the learner's internal processes, and is therefore potentially ambiguous in the information it provides. This ambiguity limits the usefulness of uptake as a measure of noticing because depending on the coding system employed, uptake-based quantifications may result in false positives (e.g. where automatic or shadowed repetition of feedback has not engaged conscious processes in any meaningful way, but gets counted as an incident of noticing nonetheless).

The central contribution of this paper, which follows, is to outline and examine a novel approach to disambiguating whether an instance of uptake signifies noticing by comparing the measured response time between feedback provision and learner uptake as a methodological control on false positives; with these false positives theorized to be particularly problematic in full *repair* uptake where the repair is merely an expression of shadowing or mimicking rather than cognized noticing.

Disambiguating Uptake Using Response Times

Before beginning to explain how examination of response times after feedback offers a potentially useful methodology in unpicking the ambiguity in uptake, it will be helpful to briefly consider the cognitive processes at work in noticing and relate them to stimulated recall and uptake.

Robinson describes noticing as "detection with awareness and rehearsal in short term memory" (1995, p. 318). If interaction is a metaphorical walk through the woods, noticing is a short cognitive pause to examine an unexpected or surprising species of tree, or put another way, it is the moment when we sense ourselves mentally stop to briefly focus our attention on a salient, if not remarkable, piece of detected information. If we extend this metaphor further, we may examine the tree for features in common with, or different from, trees we have been familiar with in the past. At this stage, a deeper kind of conscious processing of information is occurring in short term memory. Similarly, and with noticing in language learning, the type of conscious processing associated with the pause to take notice of a formal feature of input is linear, demanding of attentional resources, and, most importantly in the cognitive scheme of things, time consuming (Jiang, 2012). It is these aspects of the noticing process that result in the formation of the memory traces later accessible through stimulated recall.

Interestingly, and critically for the contribution of this paper, this cost in time associated with a resort to serial processing may also provide a way to decode the meaning of a learner's uptake

and have the potential to clear up ambiguity in what a particular response actually operationalizes. It could be argued that, for example, the response time between feedback and uptake for a learner who is either simply 'parroting' the feedback back to his or her interlocutor or engaging in backchannelling would be shorter than the response time of the same learner if he or she has detected and rehearsed the feedback in short term memory (i.e. noticing) —the former, in Leveltian (1989) terms, a copying and pasting of ready-formulated linguistic code into the articulator. It could additionally be argued that the latter case of noticing will become even longer if he or she has stopped to compare the feedback with existing knowledge of the language item at issue (i.e. noticing the gap). This would be the case irrespective of whether the uptake reaches full target-like form or not.

The remainder of this paper reports a study to examine the association between type of uptake and response time with a view to establishing an empirical basis for a methodology to distinguish true instances of noticing and noticing the gap from false positives cases of such.

Research questions and hypotheses

(1) Do response times between provision of corrective feedback and production of uptake differ according to the type of uptake produced?

Previous observations (Gurzynski-Weiss & Baralt, 2015) indicated that episodes resulting in *partial repair* are more closely associated with noticing as measured by stimulated recall responses. Additionally, Nassaji (2016) provided observations to support the claim that full repair is potentially associated with mechanical repetition and absent conscious processing. In terms of these empirical observations reported in previous literature, and the additional theoretical reasoning offered in this paper (immediately above) that resort to conscious processing will bear a time cost, the following research hypothesis was formulated (along with the associated null hypothesis).

Research Hypothesis (H_1): Episodes involving *partial repair* (referred to also as *needs repair*) will have longer response times to uptake than episodes involving *full repair* (referred to also as *repair*).
Null Hypothesis (H_0): No difference in response times to uptake will be observed between partial repair and full repair.

Alpha for the statistical decision was set at $p < 0.05$.

(2) If yes, can response time data be used to disambiguate the intent implicit in uptake?

This question was engaged with analytically through detailed analysis of discrete episodes and their respective response times.

Method

Participants

The participants were three second year female college students at a private women's junior college in Japan ranked among the lower level of higher education institutions in the country. All had completed 6 years of compulsory secondary English education and were English majors studying on a two-year course. Their TOEIC scores ranged between 355 and 510, putting them in the low-intermediate range of proficiency. All participants were volunteers and provided their informed consent to take part in the study.

Statistical Population and Sample

Intended analyses in this study did not involve inferences about the person, and the phenomenon of interest was feedback episodes. This is a commonly adopted approach in interactionist research, using a sample of discrete episodes as a dataset for statistical analysis of, for example, the relationship between uptake and retrospective reports (Egi, 2010; Bao et al, 2011). Thus, for the purposes of this study, the statistical population was all possible real-world feedback episodes (under the constraints set for feedback in this study and elaborated below), thus comprising a theoretically infinite population specification. For generalization to this population, episodes (N = 42) were sampled from interaction generated with the three participants covered above and through the interaction activities covered below.

Interaction activity

The activity used involved two pictures of a street scene as prompts. The participants and the researcher had versions of the picture in which there were ten differences. The participants were told in Japanese that they should describe features of the scene in as much detail as possible in terms of location, shape, number, and mode of action, in order that the researcher could find the differences (number unspecified to the participant) between the two versions of the picture. They were also told that the researcher might ask questions when he deemed them appropriate in order to clarify the information in the pictures.

The interaction was recorded on video. During the activity, the researcher made notes of the timings when feedback episodes occurred in order to stop the video at the appropriate places during the subsequent stimulated recall. Doing this during the interaction activity allowed the research to proceed to stimulated recall (covered below) immediately thus helping to control under-reporting of noticing incidents due to memory decay on the part of the participant. The participants took about 15 minutes to complete the interaction activity.

Feedback

Corrective feedback was provided in response to all errors: morphosyntactic, lexical choice, and

phonological. All feedback was given in the form of recasts, reformulating the participant's utterance in a target-like form which corrected all mistakes while maintaining the researcher's perceived sense of the participant's original utterance. The recasts were declarative, meaning that the errors were not stressed in order to make the location of the error more explicit. Recasts were also unflagged, that is, they were not prefixed with phrases such as "you mean…", or "Not ____, but ____" (Lyster & Ranta, 1997). Participants were given time to modify their original utterance before moving on with the activity. Depending on the error, the original utterance could be recast by changing a single word as in example (1) from the current data below, while in other episodes more complex interventions were required such as in example (2). (Episodes are numbered by order of appearance in the paper, and do not represent the actual case number in the dataset.)

(1) Participant: Every store *have two windows.
Researcher: Every store has two windows.
Participant: Windows.

(2) Participant: *Playing guitar man.
Researcher: The man playing a guitar.
Participant: The…The man playing guitar man.

*indicates error

Stimulated recall

For the stimulated recall, the participants were shown the video recording of the interaction activity. At the beginning, participants were told that they would be shown the video recording of the interaction because the researcher was interested in understanding better what they were thinking during the activity. They were told that they should feel free to stop the video at any point and verbalize what had been in their mind at that point. They were also warned that the researcher might stop the recording and ask the same. It was stressed that they should only talk about what they had been thinking at the point the recording was stopped, rather than what they thought of their performance now or anything they might have realized about their performance at the point of viewing. They were also reassured that if they could remember nothing then they should feel free to say so, and this was done to reduce any unwanted effects from subject expectancy. The stimulated recalls were recorded using a voice recorder for subsequent transcription and analysis.

Procedure

The sessions were carried out in a bright, quiet, isolated conference room in the college where the participants study. The interaction activities were completed in a single session. Participants first completed the interaction activity before immediately taking part in the stimulated recall.

Coding

Recordings of both interaction activities and stimulated recalls were transcribed by the researcher. The transcriptions were then analyzed for (i) error type, (ii) level of uptake, (iii) reports of noticing and noticing the gap, and (iv) time intervals between the provision of corrective feedback and production of modified output. These analyses are described below.

(i) Error type

The episodes were classified into three groups according to the linguistic target of the feedback provided by the researcher for each episode: phonological, morphosyntactic, and lexical.

(ii) Uptake

Any modified output produced by the participants was coded at four levels: *acknowledgement, unmodified, needs repair* and *repair*, following Egi (2010). Uptake was coded as *acknowledgement* when participants produced no language in response to feedback except a "yes" or Japanese "はい". Uptake was coded as *unmodified* when the participant produced language that either repeated all or part of the original utterance without correcting in the direction of the feedback provided. When the participant's response included a modification of the original utterance, but was still all or partially in error, the uptake was coded as *needs repair*. Finally, when the participant's response constituted a correction to the target-like form, it was coded as *repair*. Examples of each from the current data are given below.

acknowledgement

(3) Participant: *left side *front of café …
Researcher: on the left-hand side in front of the café
Participant: Yes.

unmodified

(4) Participant: There *is four stores.
Researcher: There are four stores.
Participant: four stores

needs repair

(5) Participant: *playing guitar man
Researcher: the man playing a guitar
Participant: the man playing guitar *man

repair

(6) Researcher: Where are the tables?

Participant: On *left.
Researcher: On the left.
Participant: On the left.

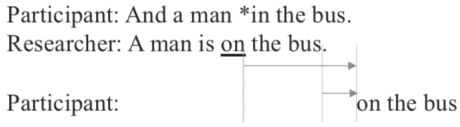
(iii) Stimulated recall

Stimulated recall data was coded at three levels based on an approach used by Egi (2010) and Bao, Egi and Han (2011). *Noticing* was adjudged to have taken place when the learner acknowledged having made a morphosyntactic, lexical, or phonological error and/or having received a target-like model of the problematic form. The episode was coded as *noticing the gap* when the participant had noticed the researcher's interlocution as feedback and reported a difference with respect to her original utterance. Episodes where the linguistic content was not referred to in any way were coded as *no noticing*.

(iv) Response times

The calculation of response times has not, as far as I am aware, previously been used in the literature in this context, and for the purposes of disambiguation of uptake as described above. Thus, the procedures are described in some detail here. Response times were calculated using phonetic analysis software package Praat (Boersma & Weenink, 2019). The package provides the researcher with graphical display of audio data with information on both the pitch and intensity (loudness) of speech production. This pitch and intensity information allows the researcher to inspect visually the start and end of words in the data. Using the cursor to highlight sections of this visual data allows the researcher to measure with an acceptable degree of accuracy the time interval between discrete points corresponding to the beginning and end of turns, such as the time between the end of one word and the beginning of another. This is illustrated in Figure 1.

One issue when calculating response times is deciding when the processes of noticing after feedback provision actually begin. On first consideration, this would appear to be the point at which the entire feedback-provision utterance ends; however, the point in the original utterance where the error occurs necessarily varies from episode to episode, as shown in the example below. There are, therefore, two alternative ways to measure response time after feedback: 1) from the point at which the entire feedback utterance/recast ends, and 2) from the point at which the correction within the interlocutor's utterance/recast occurs.



As this study represents a departure point for this methodology, and therefore no precedent exists within the literature, a decision was made to measure response time from the end of the correction in the interlocutor's feedback to the start of the participant's uptake. The rationale for this was

that this is the most plausible point at which the processes of noticing are likely to be set in motion as it marks the actual linguistic intervention from the interlocutor, and reflects methodology employed in reaction time research in linguistics, which measures reaction times from the introduction of the stimulus (Jiang, 2013).

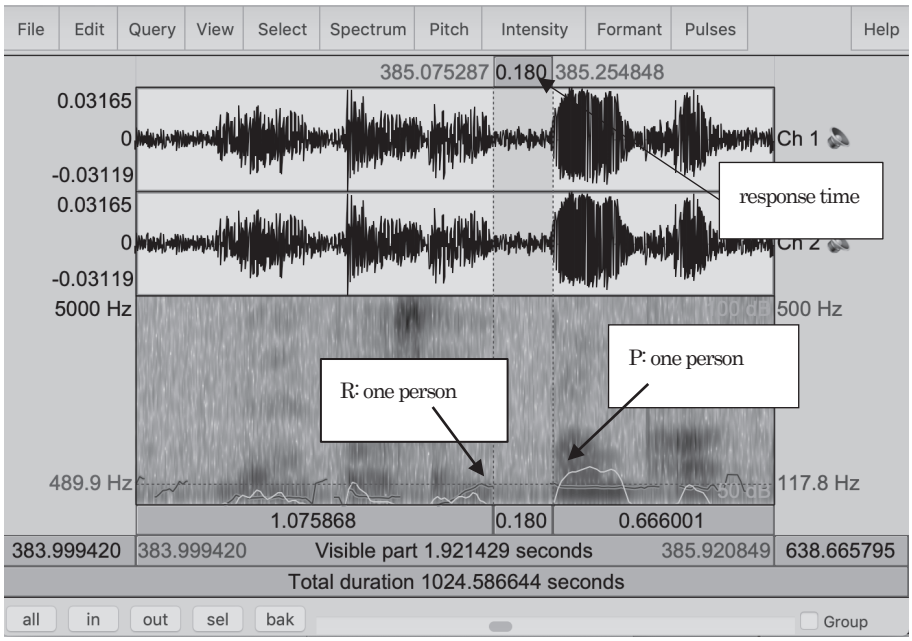


Figure 1. Example of response time analysis using Praat (Boersma & Weenink, 2019). At the bottom of the window, utterance pitch is shown by the dotted line and intensity by the light line. The participant's original utterance was "one people." The researcher's feedback is shown in R, and the participant's uptake in P. The trigger for noticing is shown at the point "person" ends. In this case, uptake begins with onset of the participant's utterance. By highlighting between these two points by reference to pitch and intensity, it is possible to arrive at an accurate measure of response time.

A second issue when sampling response times by uptake type within feedback episodes is that individual differences in overall or generalized processing speed may be present. Put another way, some respondents will tend to take more time than others to formulate uptake in response to input as a general rule. To get around this issue, raw response times were converted to *z*-scores for each participant, providing standardized scores for response times (where positive values will represent scores above the mean and negative values scores below the mean).

Data analyses

The 42 feedback episodes were combined into a single data set. After classifying the episodes by

feedback target, the data was also categorized by uptake type and by noticing. Finally, differences in response times by uptake type were investigated using a one-way analysis of variance (ANOVA). Statistical analyses were carried out using SPSS ver.25 (IBM, 2017).

Results

The results are reported in terms of descriptive statistics for error type, uptake and stimulated recall. Response time analyses are reported both by descriptive statistics and inferential statistics, namely, ANOVA.

(i) Error type

Of the 42 feedback episodes, 25 were set in motion by morphosyntactic, 16 by lexical and just one by phonological errors.

(ii) Uptake

Table 1 shows the frequencies of uptake type by feedback type. Percentages are shown in parentheses. Some form of uptake resulted from 83% of all feedback provision.

Table 1

Percentage Uptake Type by Feedback Type

		Feedback type			
		Morpho-syntactic	**lexical**	**phonological**	**total**
Uptake type	**No response**	3(15%)	4(25%)	0(0%)	7(17%)
	Acknowledge-ment	5(19%)	4(25%)	0(0%)	9(21%)
	Unmodified	6(23%)	1(6%)	0(0%)	7(17%)
	Needs repair	3(12%)	0(0%)	0(0%)	3(7%)
	Repair	8(31%)	7(44%)	1(100%)	16(35%)
		25(100%)	16(100%)	1(100%)	42(100%)

Percentages rounded off to whole numbers.

As can be seen from the data, the most common type of uptake was repair to a target-like form, accounting for 35% of the episodes. Simple acknowledgement of the feedback comprised 21%. Finally, repetition of the original utterance without modification made up 17%, while just 7% of uptake was comprised of modification that did not achieve target-like form.

(iii) Stimulated recall

Table 2 shows noticing (as inferred by stimulated recall) by type of error. It is immediately obvious that the participants in this study verbalized recollections of noticing or noticing the gap in

retrospective introspections in only 10% of cases. The most common of these self-reports were concerning feedback on morphosyntactic errors.

Table 2

Noticing as Measured by Stimulated Recall (by Type of Error)

		Feedback type			
		Morpho-syntactic	**lexical**	**phonological**	**total**
Noticing	no noticing	21(88%)	15(100%)	0(0%)	36(90%)
	noticing	1(4%)	0(0%)	0(0%)	1(3%)
	noticing the gap	2(8%)	0(0%)	1(100%)	3[1](7%)
		24(100%)	15(100%)	1(100%)	40[2](100%)

[1]Noticing the gap is, of course, contingent upon noticing, but is presented separately here.
[2]Two episodes were not subjected to stimulated recall.
Percentages rounded off to whole numbers.

(iv) Response times

Table 3 shows descriptive statistics for standardized response times to feedback by uptake type.

Table 3

Descriptive Statistics for Standardized Response Times

Uptake type	**Mean**	**SD**	**Skewness**		**Kurtosis**	
			Statistic	**Std. Error**	**Statistic**	**Std. Error**
Acknowledge-ment	0.49	0.96	1.356	0.752	1.963	1.481
unmodified	0.26	0.56	0.526	0.794	0.378	1.587
needs repair	0.95	1.29	1.443	1.225	–	–
repair	−0.63	0.66	0.358	0.654	0.257	1.091

Note: mean values based on standardized scores, and thus both positive and negative in value

As can be seen from the data, response times when *needs repair* uptake was produced were by far the longest. Kurtosis statistics for the *needs repair* data could not be calculated due to the small number of data points. All other statistics were within acceptable ranges. Because episode subsamples (*acknowledgement, unmodified, needs repair* and *repair*) were relatively unequal in size, Levene's test was executed to test for equality of variances, an assumption necessary for ANOVA under such limitation. The result was not significant, $p > 0.05$, thus indicating equality of variance. Additionally, the Shapiro-Wilk test was executed to test for normality, and again the result was not significant, $p > 0.05$, indicating that scores for each group did not deviate significantly from normality.

(RQ1) Do response times between provision of corrective feedback and production of uptake differ according to the type of uptake produced?

To answer the first research question (and test the associated research hypothesis), a one-way between groups ANOVA was conducted to compare the effect of uptake type (i.e. the acknowledgement, unmodified, needs repair, and repair conditions) on standardized response times. There was a significant effect of uptake type on standardized reactions times at the $p < .01$ level for the four conditions [F (3, 30) = 5.984, p = 0.003]. Post hoc comparisons using the Hochberg test (used because of the large differences in group size [Field, 2013]) indicated that response time for the repair group (M = -0.63, SD = 0.66) was significantly shorter than the acknowledgement group (M = 0.49, SD = 0.96). Mean response time for the repair group (M = -0.63, SD = 0.66) was also significantly shorter than the needs repair group (M = 0.95, SD = 1.29). This latter finding led to rejection of the null hypothesis (H_0) formulated for this study, and therefore acceptance of the research hypothesis (H_1) that response times for the needs repair group (partial repair) would be longer than those for the repair group (full repair). No other group means were found to be significantly different.

(RQ2) If yes, can response time data be used to disambiguate the intent implicit in uptake?

As ANOVA results indicated significant between-groups differences in standardized response times, *post hoc* examination of the response time data was carried out. Two groups of data were considered: 1) discrete episode response times when noticing had been reported (via stimulated recall) and 2) discrete episode response times where full *repair* resulted from feedback provision. These are shown in Tables 4 and 5 respectively. As can be seen from Table 4, two of the response times for uptake where noticing was reported were longer than the mean (which would be a value of zero in the standardized scores presented in the table), while two fell below the mean. In the case of the two feedback episodes that resulted in utterance repair (i.e. full repair), one response time sits below the mean and one above. As is clear from Table 5, all standardized response times for *repair* episodes bar one fall below the mean.

Table 4
Analysis of Feedback Episodes Where Noticing Was Reported by Standardized Reaction Time, Linguistic Target, Noticing Level, and Uptake Type

Episode	Interlocutor	Transcript	Reaction Time *z*-score	Linguistic target	Noticing level	Uptake type
	Participant:	Playing guitar man				
(7)	Researcher:	The man playing a guitar	2.41	Morpho-syntactic	noticing the gap	needs repair
	Participant:	The man playing guitar man				
	Participant:	Go out a man				
(8)	Researcher:	A man is coming out	0.47	Morpho-syntactic	noticing the gap	repair
	Participant:	A man is coming out				
	Participant:	And on the fish...				
(9)	Researcher:	Above the fish	−0.28	Lexical	noticing the gap	Acknowledgement
	Participant:	ah yes				
	Participant:	She look like child [*i*]				
(10)	Researcher:	like a child [*ai*]	−0.89	Phono-logical	noticing	repair
	Participant:	child [*ai*]				

Table 5
Standardized Reaction Times during Feedback Episodes which Resulted in Repair during Uptake

Episode	Interlocutor	Transcript	Reaction Time *z*-score	Linguistic target
	Participant:	There is one people.		
(11)	Researcher:	One person.	−1.33	Morphosyntactic
	Participant:	One person.		
	Participant:	And on left...		
(12)	Researcher:	On the left.	−0.87	Morphosyntactic
	Participant:	On the left.		
	Participant:	Here, people talking and eat lunch.		
(13)	Researcher:	People are talking and eating lunch.	−0.75	Morphosyntactic
	Participant:	Eating lunch.		
	Participant:	There are two desks.		
(14)	Researcher:	Tables.	−0.70	Lexical
	Participant:	Tables.		
	Participant:	Café and flower shop has parasol.		
(15)	Researcher:	Awning.	−0.58	Lexical
	Participant:	Awning.		
	Participant:	There is ***chuousen.***		
(16)	Researcher:	A centre line.	−0.66	Lexical
	Participant:	Centre line.		

Discussion

The purpose of this study was to explore the potential usefulness of using uptake response times as a means of disambiguating possible false positives of noticing in uptake. It was argued that the shift from parallel to linear processing associated with the detection and registering of feedback in short term memory, and the comparing of that information with current imperfect linguistic knowledge, would result in a measurable cost in time as compared with instances where these processes of noticing were not engaged. The exploratory nature of this study as a preliminary attempt to arrive at a working methodology precludes the drawing of clear conclusions. However, the data do provide some encouraging signs, and a direction for further travel in continuing research.

(RQ1) Do response times between provision of corrective feedback and production of uptake differ according to the type of uptake produced?

Before considering response times, a better understanding of the data can be reached by first examining stimulated recall reports and uptake in some detail. Participants' self-reports of their conscious thought processes during stimulated recalls provided little information for analysis, but some brief comments can be made. Only 12% of feedback episodes resulted in reports of some kind of noticing, far fewer than in previous recast-based studies such as Egi (54%: 2010), Bao et al. (37%: 2011), and Gurzynski-Weiss and Baralt (75%: 2015).

This paucity of self-reported noticing raises two issues. First, the nature of the errors may have precluded noticing on a level deep enough to result in formation of long-term memory representations. Many of the errors were of the kind that might be expected of beginners or false beginners, and as such may be better characterized as *mistakes* rather than true errors (Ellis, 1989). This is an important distinction and is returned to later in this discussion.

Second, the data may reflect a problem with stimulated recall methodology inherent in the open-ended nature of the questioning it employs (Egi, 2010). A self-report of some level of noticing indicates just that: some degree of attention having been focused on the learner's error as a result of feedback on the erroneous nature of the original utterance. However, as that self-report may not be exhaustive of everything that was in the participant's mind at the time, no indication of noticing may not be conclusive proof of no noticing: a 'false negative.' Clearly, this is related in some way to how 'deep' the level of noticing actually was during the episode, but as the protocol widely employed in stimulated recalls (Gass & Mackey, 2013; MacDonough, 2005; Mackey et al. 2002; Mackey et al., 2010) does not make any indication of those areas of his recollections that the participant should be considering, the possibility remains that he or she is not verbalizing that part of his thought processes which is relevant to the object of research interest.

This raises an interesting related issue regarding the ability of participants in stimulated recall protocols to verbalize their thought process. This ability could be constrained by individual

differences in a number of affective and cognitive factors, such as anxiety in the laboratory environment, motivation to take part in the research, self-imposed expectations of what the researcher wants to hear, or unwillingness to appear to have been in error on the one hand, and ease with which long term memory representations are formed, or meta-awareness of thought processes on the other. These issues appear to warrant further investigation, as they potentially impact on the quality of the data that can be collected if the researcher is relying solely on stimulated recall to measure noticing.

As was seen from the results, some form of uptake resulted from feedback provision in almost 83% of episodes, comparing favourably with previous studies (14%: Bao et al., 2010). *Acknowledgement* and *repair* accounted for around half of this uptake. When this data was put side-by-side with the data from stimulated recalls, the online observation data quantified seven times more instances of 'noticing' than retrospective reports. That is, in the current study these two methodologies offered widely differing quantifications of what is notionally the same process: either stimulated recall is under-reporting, or uptake is over-reporting.

This is problematic because noticing, either as stimulated recall reports or uptake, has been used in a body of research investigating the relationship between noticing and a range of other measures, such as L2 development (Ammar & Spada, 2006; Goo, 2012; Lyster, 2004; Mackey et al, 2002; Mackey et al., 2010; Yang & Lyster, 2010), individual differences in working memory (Goo, 2012; Mackey et al, 2002; Mackey et al., 2010; Trofivimich et al., 2007), language anxiety (Sheen, 2008), and type of feedback provided (Ammar & Spada, 2006; Goo, 2012; Lyster, 2004; Yang & Lyster, 2010). Clearly, no quantification of internal cognitive processes can be perfect in every way, but in order to maximise the validity of our results it is clearly important that we employ measures that are as valid as possible.

It was in an attempt to refine and improve on the reliability of uptake as a measure of noticing that this exploratory study was carried out. Rejection of the null hypothesis and acceptance of the research hypothesis (with respect to response-time differences between episodes involving partial repair versus full repair) indicated support for the theoretical reasoning which informed their formulation; which, of course, was that the turn to conscious processing would extract a time cost. Thus, while acknowledging the limitation imposed by the size of the dataset, the results reported here are encouraging, because these results provide an empirical rationale for use of response times as one further discriminative criterion in identifying real cases of noticing versus false cases. The results are even more encouraging when one considers that the small sample size is limiting because small sample size offers low statistical power; that is, a higher possibility of a false negative (or Type II error). It is therefore important to note that the null hypothesis was rejected in this study even though the sample size was small and lacked power, and confirmation of the research hypothesis could be characterized, in statistical logic, as having been achieved despite the limitations of sample size rather than because of them.

As previously noted, Gurzynski-Weiss and Baralt (2015) reported that partially modified output

significantly predicted reports of retrospective noticing. As they argue, partial repair of an erroneous utterance may be indicative of learners' having to some degree or other unpacked and decoded the utterance and attempted to put it back together in a different form, with this obviously requiring a stronger focusing of attention on linguistic form. Such a characterization of *needs repair* uptake is consistent with the longer response time observed for this type of feedback in the current data. As was seen from the results of ANOVA, the mean standardized response time when participants produced *needs repair* uptake was found to be significantly longer than full-repair episodes. Indeed, full repair was also found to be significantly shorter than simple acknowledgement. These results lend support to the inference that these instances of *needs repair* uptake exhibit the difference in degree of conscious serial processing that was hypothesized to occur when noticing and noticing the gap occur.

Further, results in the current study showed that the mean response time for full *repair* uptake was the shortest of the four types. Intuitively, *repair* would appear to be indicative of noticing as a reformulation of the learner's error in response to feedback provided. However, this may be dependent upon the kind of feedback given to the learner. In the case of indirect feedback, such as metalinguistic feedback, prompts, or elicitations (Lyster 2004), this is likely to be the case, as this type of feedback requires the learner to actively engage with the linguistic building blocks to formulate a new version of the utterance. A body of research proposes that this indirect feedback is more effective for this very reason (Ammar & Spada, 2006; Lyster, 2004; Yang & Lyster, 2010). However, in the current study all errors were recast in a target-like form, that is direct feedback, providing the learner with a 'final version' to mimic. Seen in this light, in some cases full *repair* uptake may indeed represent little more than a simple repetition of the provided exemplar (Nassaji, 2016). In a Leveltian (1989) sense, this kind of uptake may often be a kind of 'copy-and-paste' of verbal text encoded in short term memory on to the formulator for subsequent articulation. The observed shorter-than-mean response time is consistent with this kind of utterance formulation in which little conscious processing is involved. Different patterns may be observed when feedback provision is indirect rather than the direct feedback, namely recasts, employed here.

(RQ2) If yes, can response time data be used to disambiguate the intent implicit in uptake?

At this stage, however, while these interpretations are consistent with the data, response times still represent a statistical trend rather than a fool-proof—or even useful—measure of incidence of noticing. With this in mind, we can take a closer look at some of the episodes which might be disambiguated by employing measured response times to see if the methodology has true utility. Closer examination of the data suggests that the approach may have particular validity with respect to morphosyntactic errors.

As a first step, I consider the small number of episodes where both noticing (stimulated recall) and uptake occurred. These are shown in Table 4. Two of these response times were longer than

mean, and two shorter. Interestingly, the ones longer than the mean both resulted from feedback targeting morphosyntactic error, while those shorter than the mean resulted from feedback focused on issues of phonology and lexis. The longest standardized response time (more than two standard deviations from the mean) occurs in (7) where the participant is attempting to form an adjectival phrase after self-reported noticing the gap; however the final result is still in need of further repair, and partial repair is a type of uptake that is theorized to be especially indicative of noticing (Gurzynski-Weiss & Baralt, 2015). The participant's retrospective report, " ギターを弾いている男と言いたかったけど、組立が変になっていてと思いました " (*I wanted to say "The man playing a guitar" but I thought that I had got the combination [of words] wrong*), is consistent with this interpretation of the response time. The stimulated recall response given to (8), " わかるのをとりあえず言って、正しいのを言ってくれたのでそれを使いました " (*I said what I could say, but you gave me a correct version so I used that*) indicates both noticing and noticing the gap. Thus, despite concerns about full *repair*, the uptake measure, being truly indicative of noticing, in this case the response time does appear to confirm an occurrence of noticing.

Response times in episodes (9) and (10), however, both fall below the mean despite retrospective reports of noticing. Episode (9) results in simple *acknowledgement*, while the stimulated recall was "above ということ、あ！それがあった、って単純に思いました。あ！なるほど、そういう ah です " (*Of course – above – that's the word. It was as simple as that. That's what I thought*). As can be seen from this retrospective report, the learner appears to have made *mistake* rather than a true *error* which, of course, should not be unexpected as mistakes occur even in native-speaker output. The short response time is consistent with little effortful conscious processing. Finally, it can be seen that (11), a correction of a mispronunciation of the word *child* results in the shortest response time, as we would expect in the case of a simple mistake rather than a full-blown error. Stimulated recall results appear to confirm this interpretation: "Children と言ったけど、一人しかいないと思って、言いなおさなきゃと思いました (*I said children, but I realized there was only one and I thought I should correct what I said*). These, of course, represent only four data points, but the data appear to be consistent with the hypothesis that true noticing is consistent with the elevated response times expected when conscious processing comes online.

So, what happens if we apply response time analysis in attempt to disambiguate false positives? ANOVA analysis indicated that full repair episode response times were significantly shorter than those in which partial repair was observed, and it is to a *post hoc* consideration of this type of uptake that I turn next. Table 5 shows data for the feedback episodes which fit these criteria. The first three of these episodes target morphosyntactic errors, and the remaining three lexical.

In (11) the error is one of plural form. In the stimulated recall, the participant responded " バスの特徴として後ろに人が乗っているかもしれないけど、「一人」と言った " (*in describing the bus, it occurred to me that there might be other people in the back of the bus, but I said "one person*"). This response suggests that the participant's attention was focused on message formulation (*meaning*) rather than *form*, but alternatively she may have simply failed to verbalize some noticing that did, in fact,

occur. However, the response time is between two and three standard deviations longer than those observed in episodes (7) and (8) above where noticing was reported, and it seems unlikely that the participant engaged in conscious processing of the input. A researcher examining this data point *post hoc* might be fairly confident in dismissing this as a false positive based on reaction time evidence. Similar differences from the noticing data is observed for episodes (12) and (13). The error in (12) is one of article usage. The stimulated recall, "本当は left じゃなくて手前" (*it wasn't left, it was in front*), again suggests utterance formulation, rather than grammar, to be the main target of the participant's attention, and the response time indicates that little conscious processing has been allocated to the former. Again, we might see this as a mechanical repetition of the feedback, or phonological data 'cut and pasted' into the formulator to produce her uptake. Considering the next episode (13), the participant describes her recollection of it as "このパラソルは何っていったらいいんだろう" (*I was wondering what English word to use for this parasol*" [in the picture]). The response gives no indication of any noticing of the feedback, and the short response time appears to confirm this judgement and the repair might be dismissed as a false positive.

Looking next at episodes where lexical feedback was provided, standardized response times were similarly shorter than the mean. In these two cases, the participant reported retrospectively that her focus of attention had been message formulation or completion of the task itself, and standardized response times indicate that no conscious processing – and by extension no noticing – has occurred. Episode (14) was retrospectively reported as "ここで最初にこのパラソルの人のこと言おうと思っていました" (*at first I was going to talk about the man under the awning*) which is clearly message formulation. In episode (15), the participant's account was "あぁ、違うんだと思って、先生のにもあるんだって" (*I thought I got it wrong as it was in your picture, too*). Both of these episodes could be rejected as incidences of noticing based not just on the stimulated recall data, but also on the strength of the observed response times.

Indeed, the only episode (16) in which the standardized response time is positive is in an episode where the participant implicitly asks for a translation of the Japanese. Responding "自分の写真にある乗り物について全部触れていたんで言いました" (*I'd already talked about all the vehicles in my picture, so I said this*), again the participant's stimulated recall is overridingly concerned with message formulation. However, in this case the researcher might conclude that the learner did indeed notice the gap if the measured standardized response time were taken into account, and hence this approach potentially represents one means by which the true noticing 'wheat' could be sorted from the mechanical repetition 'chaff'.

Conclusions and limitations

This paper examined the efficacy of response times as a measure of noticing in uptake produced in response to corrective feedback during interaction. As has been outlined above, current methodologies available to the interactionist researcher, stimulated recall and uptake, offer

different perspectives on the processes of noticing, and present both advantages and disadvantages. Stimulated recall appears to shine the brightest light on the learner's internal processes, and is more sensitive to different degrees of noticing, that is, noticing and noticing the gap. However, several studies have reported episodes where uptake has not been accompanied by retrospective reports of noticing; indicating that relying solely on stimulated recalls may result in an under-reporting of instances of noticing. The data in the current study also raised the possibility that stimulated recall may, on occasion, result in the reverse problem, which is *under-reporting* of noticing, something which we might term *false negatives*. At the same time, employing only uptake as an operationalization of noticing risks excluding cases where the processes of noticing have been present but did not result in a learner producing any language. Additionally, uptake, especially when that uptake results in complete repair of the utterance, can be difficult to interpret, and researchers may conversely be including data in their quantifications of noticing that in reality are not real noticing at all.

The current exploratory study sought to address the last of these issues and ANOVA results obtained did indeed indicate a significant difference in response times between *repair* uptake and *needs repair* uptake. By considering response times during episodes in which stimulated recall indicated that noticing or noticing the gap had indeed occurred, the results suggested that genuine noticing resulted in longer response times as expected. Further close examination of full *repair* episodes suggested that almost all of these cases could be argued to be 'false positives' of noticing when response time evidence was taken into account. In the light of these discussions, response time data may provide a means by which the ambiguities of uptake data can be addressed as an empirically-grounded methodology.

These encouraging signs should be tempered by the limitations of the current study which is exploratory and excursive in its scope. The data consisted of only 42 episodes and while this comes close to other studies of this type in the literature (e.g. Bao et al., 2011), a larger data set would have increased confidence in the conclusions drawn from these results; especially given that the prospects for a new methodology for engaging with interactional data is being outlined. In addition, the current study was limited to one type of feedback provision—recasts—and future research should examine whether other types of feedback result in different uptake response time patterns. Finally, although *post hoc* inspection of response times in specific episodes suggest that response times may indicate the presence of false positives, in the absence of inferential statistics it is impossible to declare a cutoff where noticing has not taken place. Again, further research is necessary to address this issue.

References

Ammar, A., & Spada, N. (2006). One size fits all?: Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition, 28*(04), 543-574.

Bao, M., Egi, T., & Han, Y. (2011). Classroom study on noticing and recast features: Capturing learner noticing with uptake and stimulated recall. *System, 39*(2), 215-228.

Boersma, Paul & Weenink, David (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.05, retrieved 16 October 2019 from http://www.praat.org/

Egi, T. (2010). Uptake, modified output, and learner perceptions of recasts: Learner responses as language awareness. *The Modern Language Journal, 94*(1), 1-21.

Ellis, R. (1989). *Understanding second language acquisition* (Vol. 31). Oxford: Oxford university press.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

Gass, S. M., & Mackey, A. (2007). Input, interaction, and output in second language acquisition. In VanPatten, B & Williams, J (Eds), *Theories in second language acquisition: An introduction* (pp. 175-199). New York: Routledge.

Gass, S. M., & Mackey, A. (2013). *Stimulated recall methodology in second language research*. New York: Routledge.

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition, 34*(03), 445-474.

Gurzynski-Weiss, L., & Baralt, M. (2015). Does type of modified output correspond to learner noticing of feedback? A closer look in face-to-face and computer-mediated task-based interaction. *Applied Psycholinguistics, 36*(6), 1393-1420.

IBM Corp. (2017). IBM SPSS Statistics for Macintosh, Version 25.0. Armonk, NY: IBM Corp.

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics, 24*(2), 168-196.

Jiang, N. (2013). *Conducting reaction time research in second language studies*. New York: Routledge.

Levelt, W. J. (1989). *Speaking: From intention to articulation*. Cambridge, Massachusetts: MIT press.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In Ritchie, W. C. & Bhatia T. K. (Eds.), *Handbook of second language acquisition* (pp. 413-468). New York: Academic.

Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning, 48*(2), 183-218.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition, 26*(03), 399-432.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition, 19*(01), 37-66.

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27*(3), 405-430.

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning, 60*(3), 501-533.

Mackey, A., Philp, J., Egi, T., & Fujii, A. (2002). Individual differences in working memory, noticing of interactional feedback, and L2 development, in P Robinson (Ed), *Individual Differences and*

Instructed Language Learning (pp. 181-209). Amsterdam: John Betjemans.

McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition, 27*(1), 79-103.

McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning, 56*(4), 693-720.

Nassaji, H. (2016). *The interactional feedback dimension in instructed second language learning: Linking theory, research, and practice*. London: Bloomsbury.

Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly, 36*(4), 573-595.

Pica, T. (1994). Research on Negotiation: What Does It Reveal About Second-Language Learning Conditions, Processes, and Outcomes? *Language Learning, 44*(3), 493-527.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45*(2), 283-331.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In Schmidt, R (Ed), *Attention and awareness in foreign language learning* (pp. 1-63). Hawaii: University of Hawaii Press.

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In Day, R (ed), *Talking to learn: Conversation in second language acquisition* (pp. 237-326). Rowley, MA: Newbury House.

Sheen, Y. (2008). Recasts, language anxiety, modified output, and L2 learning. *Language Learning, 58*(4), 835-874.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In Gass, S & Madden, C (Eds.), *Input in second language acquisition* (pp. 165-179). Rowley, MA: Newbury House.

Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms *Studies in Second Language Acquisition, 32*(2), 235-263.

口頭インタラクションにおけるフィードバックエピソードでの「気づき」
―「気づき」を計る方法としてアップテイクまでの反応時間の効能性について―

モクスン・ジョナサン

学習者にとって相互交流（インタラクション）は自分の発信の中にあるエラーに気づく機会であり、現在の言語知識と目標とする言語知識の違いに気づく機会、そして現状の言語知識を修正する機会として、第二言語発達に有効であるとされている（Long, 1996）。インタラクションによる第二言語発達において中心となるのは noticing と noticing the gap である。しかし、この認知プロセスを数値化する現在の方法は、正確な数量化が行われているのかが不明である。この小規模な研究では、３人の被験者のインタラクション活動において、エラーに対するフィードバックが行われ N＝42のフィードバックエピソードのデータが収集された。Noticing を数量化する方法は、先行研究において標準的に使用されている uptake（Lyster, 1998）と stimulated recall（Gass & Mackey, 2013）が使用された。さらに、フィードバックを得てから uptake を発信するまでの反応時間を測る新規の方法を用いて uptake と stimulated recall からのデータを照らし合わせ、uptake を発信した被験者の意図の明確化を試みた。その結果、uptake による noticing の数値化においてこの新規の方法を加えることは、偽陽性の疑いがあるエピソードをより明確に数量化することに有効であることが示唆された。