Reports of 2013 Studies in Japan Association for Language Education and Technology, Kansai Chapter, Methodology Special Interest Groups (SIG)
Fukuta Junya (pp. 75–99)

Measuring Attention to Linguistic Form in a Second Language During Speech Production

—A Review and Proposal for Empirical Study—

Fukuta Junya
Graduate School, Nagoya University
Research Fellow of Japan Society for the Promotion of Science

Abstract

This paper examines the methods used to measure the attention paid to linguistic form while speaking in a second language (L2). Although researchers have investigated the effects of attention on second language acquisition (SLA), the validity of the methods used in each study has been a major problem for some time. In particular we note that while there are reliable ways of, for example, carrying out performance analyses for the use of written and spoken language (e.g. T-unit, and AS-unit) there are no comprehensive, accessible and detailed guides on how to identify attention orientation in protocol data. This study reviews the methods used in previous studies. It then looks in detail at a method which employs stimulated recall prompted by disfluency-markers. The study examines the theoretical rationale behind this method and suggests that it is perhaps a good way forward for measuring attention in L2 speech production. The study also provides a sample analysis using this method to examine the validity of the procedure and to develop a coding scheme.

Keywords: attention, noticing, disfluency-marker, stimulated recall, oral production

1. Background

1.1 The Role of Attention to Form in Second Language Acquisition

The role of attention in L2 acquisition is both an old and a current problem. Schmidt (1990), for example, claims that L2 acquisition is impossible with subliminal learning, and that focal awareness on the target linguistic form (word, phrase or

sentence) is necessary to acquire L2. Schmidt argues that learners have to pay conscious attention to the formal aspects of a language via input-based learning and have to notice these formal aspects consciously in order to acquire language (the noticing hypothesis). In accordance with the awareness levels recognized in cognitive psychology, Schmidt identifies three crucial levels of awareness: perception, noticing, and understanding. Perception is the most primitive mental organization, creating internal representations of external events. It is not necessarily a conscious process, unlike awareness. Noticing involves awareness and corresponds to "focal awareness (Atkinson and Shiffrin, 1968; Kihlstrom, 1984), episodic awareness (Allport, 1979), and apperceived input (Gass, 1988)" (Schmidt, 1990: 132). Understanding is acknowledged as a higher level of conscious attention than noticing. This higher level of awareness is generally used when we engage in problem-solving, metacognition, or rule understanding. According to Schmidt, understanding is not necessary for learning, but it facilitates L2 development.

Some studies support the noticing hypothesis (e.g. Leow, 1997, 2000), while others (e.g. Leung & Williams, 2011; Williams, 2005), using empirical data, claim that learning can take place without awareness. Hama and Leow (2010) criticized the methodology used in Williams (2005) study, namely offline retrospective interviews, and partially replicated the study by adding online think-aloud protocol data. The result indicated that awareness had a significant effect on learning. This issue is still being debated (e.g. Leung & Williams, 2011a, 2011b; Faretta-Stutenberg & Morgan-Short, 2011; Rebushat, et al., in press). This discussion stresses the importance of further methodological development.

The debate on the possibility of learning without awareness may be controversial, but as to the positive effect of conscious attention, much of the research confirms that those participants who exhibited a greater level of noticing linguistic form during L2 processing demonstrated greater improvement in L2 acquisition (e.g. Godfroid, Hausen, & Boers, 2012; Philp, 2003; Robinson, 1995a, 1995c; Rosa & Leow, 2004; Rosa & O'Neill, 1999). At present, the idea that conscious attention can facilitate L2 acquisition is widely accepted in empirical studies in this area.

1.2 Speech Production Models and the Generation of Attention

This paper aims to re-examine the methods used to measure attention paid to linguistic forms when speaking in a second language. As part of this reappraisal I

briefly discuss the cognitive process that generates attention to linguistic forms in L2 learners based on previous studies that provide a blueprint for the processes underlying speech production.

Levelt (1989) postulated three stages in the primary language production system represented by three autonomous "specialists": the conceptualizer, the formulator, and the articulator. First, the conceptualizer generates the communicative intentions. Here communicative goals and intentions are elaborated (macroplanning). Speakers select the information to be encoded and decide on the order in which this information will be conveyed. At the same time, speakers decide on the perspective that they need to take in conveying the message (microplanning).

The outputs of conceptualization are sent to the formulator. The formulator is also made up of two parts: grammatical and phonological encoding. Grammatical encoding further involves lexical and syntactic encoding. During this process, a speaker accesses the mental lexicon. The mental lexicon consists of two lexical entries: lemmas that contain syntactic information about the lexical entry, and lexemes that store information about the morpho-phonological form of the lexical entry. This output is conveyed to the final phase of speech production: the articulator. The articulator converts the speech plan into actual speech.

This speech model also explains the monitoring function, which involves checking the correctness and appropriateness of production. Before and after speakers produce their speech, the monitoring process checks for and detects errors and inappropriate elements in speech production. When monitoring detects speech errors, self-repairs take place either overtly or covertly.

On the basis of this speech model, Kormos (2006) draws up a comprehensive model of L2 speech production. Kormos argues that "there is a common episodic and semantic memory for L1 and L2, a shared store for L1 and L2 lemmas and lexemes (2006: 167)." She also emphasizes the need for a specific L2 knowledge store which is a declarative memory of grammatical and phonological rules in L2. Proficient L2 learners do not rely on this separate declarative knowledge, whereas less proficient learners tend to make conscious use of the declarative knowledge of grammatical and phonological rules.

Doughty (2001) also identifies two distinct stages of the speech process related to L2 acquisition: macroprocesses (i.e. internalization of input, mapping, analysis, and restructuring) and microprocesses (i.e. selective attention, cognitive comparison, and

focus on form). She maintains that macroprocesses are more or less automatic, and microprocesses are controlled, shorter-term, and sometimes momentary.

As Kormos (2006) and Doughty (2001) pointed out, the speech processes of L2 learners (especially less proficient ones) are usually unlike L1 and are not fully automated. As a result L2 learners often make use of controlled processing and access their L2 specific declarative knowledge. In order to utilize such process and knowledge, selective attention controls access to awareness (Orgeta, 2009). L1 users are largely dependent on automatic processing in syntactic encoding, phonological encoding, and lexical access (according to some theories, this also applies to the monitoring process). In L2 learners, however, these processes are not usually automatic and therefore involve a largely conscious process in which attentional control is required. In order to measure attention shifts from meaning to form during speech production, we have to capture the conscious process that is required for attentional control.

1.3 Approaches to Measuring Attention during Speech Production

Many researchers have attempted to operationally define the process of attention to linguistic form with awareness or noticing during L2 processing (e.g. Jourdenais, 2001; Schmidt, 2001; Leow, 1997; Robinson, 1996, 1997; Philp, 2003; Rosa & O'Neil, 1999). The methodology used to capture participants' cognitive processes can be categorized into online and offline measures. This is based on whether the measurement is conducted "at the concurrent (online) stage of encoding the incoming information" or "at the nonconcurrent (offline) stage of retrieval of stored knowledge of the construct (Hama and Leow, 2010: 466)." In other words, it depends on whether the measurements were conducted at the same time as the attention shift occurs, or were carried out after the task was completed.

Shanks, Green, and Kolodny (1994) have suggested that offline measures are too insensitive a method to measure awareness. Their study shows that some participants who were classified as being unaware of information by the offline measure were classified as being aware by the online measure. Another problem is that participants may make different inferences about their intentions as time passes after the task used to measure attention is completed (Gass & Mackey, 2000). As a result they argue that offline measures may not accurately reflect the participants' cognitive processes. As for online measures, however, Jourdenais (2001) argues that reporting on a task at the same time as the task is being carried out may itself change the cognitive process and

the way the task is performed. This limits the validity of the online process.

These phenomena are closely related to the concepts of reactivity and veridicality. Reactivity refers to the effects (both negative and positive) on learner performance and/or on subsequent learning which results of the introspective reporting of a task are. Veridicality refers to the accuracy of participants' reporting of their thought processes. These factors are line-symmetric trade-offs which have time as an axis; in order to reflect cognitive processes accurately, offline measures should not be adopted, while online measures are not appropriate to measure attention because of the way they limit cognitive resource.

In order to measure attention to linguistic form, we cannot therefore adopt online measures except for direct analysis of a performance; it is quite difficult to produce language orally and to simultaneously elicit introspective comment. When studying speech production, the veridicality problem is also inevitable when obtaining protocol data. The next section summarizes the advantages and disadvantages of the methods described here with a particular focus on reactivity and veridicality.

1.4 Methods for Measuring Attention in Previous Studies

A questionnaire or a retrospective interview is one of the most common ways to examine participants' cognitive processes. In order to measure their attention to linguistic form, participants were asked questions about their own speech production after completing a speech task. This method of gathering data makes it possible to analyze responses qualitatively. Reactivity is not a problem because the questionnaire or interview is always completed after the task has been completed, but veridicality is an issue because of the necessary delay in completing a questionnaire or interview after the speech task is completed. Some researchers (e.g. Hama and Leow, 2010; Leow & Hama, 2013) have strongly criticized the retrospective questionnaire as a way of measuring attention.

Stimulated recall has been proposed as a solution to the veridicality problems in studies based on questionnaires. Stimulated recall is defined as "one subset of a range of introspective methods that represent a means of eliciting data about thought processes involved in carrying out a task or activity (Gass & Mackey, 2000:1)". In other words, this method combines use of retrospective interviews with reminders of the event (audio recordings, videotaped recordings, transcripts, and so on). This method does not risk reactivity because the stimulated recalls are conducted after the

tasks have been completed. In addition, the use of stimulated recall helps avoid the problem of veridicality by providing contextual information which assists the participants' ability to retrieve memories. Bloom (1954) found that if the recalls were conducted a short period of time after the task (within 48 hours), recall was 95% accurate. Accuracy declined gradually as the time between the event and the recall lengthened. A possible limitation to the stimulated recall method is that the prompts for recall can themselves be arbitrary. This study discusses the problem and suggests a possible solution.

Philp (2003) used something similar to stimulated recall - a retrospective report immediately after oral interaction - to solve the veridicality problem. Philp's study defines noticing as the learner's ability to repeat immediately after sounds (such as knocks) made during a task immediately after the task has been completed. This method is called immediate report or immediate recall. Philp noted that the immediate report method measures the amount of attention paid and, importantly, can also capture a noticing level of attention: the rehearsal of input in working memory (Robinson, 1995a). Some researchers (e.g. Norris, 1990; Egi, 2004) reported no reactivity in immediate retrospective reports on the performance of participants, but others have reported some reactivity (e.g. Russo et al. 1989). It should also be noted that this method requires participants to practice recall any number of times during a short period of time. This in itself may cause more attention to be paid to what the researcher is asking than happens with stimulated recall or other methods used in the recall process.

Performance analysis is the method most commonly used (and the sole method used in online measurement) to assess how a learner allocates attention during speech production. This method is largely based on Skehan's (1998) three-way distinction between complexity, accuracy, and fluency in learner speech production. Because of limited attention capacity, it is difficult for learners to achieve all three of these performance aspects simultaneously (Skehan & Foster, 1999). When learner accuracy and complexity are increased at the cost of fluency, it is thought that learners focus their attention on linguistic form. According to this view, fluency reflects the learner's focus on meaning, while attention to form can be observed by complexity and accuracy of speech. Some studies have measured attention levels by interactional moves: negotiation of meaning (Gilabert, Raron, & Llanes, 2009), recasts (Philp, 2003), language-related episodes (Kowal & Swain, 1994; Swain & Lapkin, 1995; Fortune,

2005; Fortune & Throp, 2001), and self-repairs (Declearck & Kormos, 2012; Kormos, 2000). The analysis of these markers can also be considered as performance analysis.

As can be seen in the summary above, the problem of reactivity and veridicality creates a dilemma. That is, the offline elicitation method does not cause reactivity, but the veridicality problem remains because of the distance of time between the task being completed and the retrospective interview. However some offline methods such as stimulated recall and immediate report avoid the veridicality issue.

1.5 Representative Studies related to Attention and Speech Production

Foster and Skehan (1996) investigated the impact of pre-task planning on the way L2 learners allocated attention by using three different tasks (a personal information exchange task, a narrative task, and a decision-making task) and three different implementation conditions (no planning, broad-brush planning, detailed planning; the detailed planning group received guidance on how they might use planning time to consider linguistic forms, but the broad-brush planning group received no guidance and were simply told to plan). Statistical analysis showed that this method demonstrated the importance of trade-off effects between complexity and accuracy.

On the other hand, Robinson (2001a, 2001b, 2003, 2005) disputes Skehan's trade-off hypothesis. Robinson argues that task complexity along the cognitive/conceptual dimension (resource-directing dimensions: e.g. +/- here and now, +/- few elements, and +/- reasoning demands) facilitates learner attention to linguistic form. That is, cognitive demands can raise the degree of accuracy and complexity simultaneously at the cost of fluency (note that Robinson does not disagree with the trade-offs between complexity and accuracy scores, and fluency scores).

Gilabert (2007) analyzed the effects of manipulating the cognitive demands on L2 speech performance. He focuses on a measure of accuracy and self-repair to measure learner attention to form. In this study, 42 below-intermediate students performed three different task types and manipulated task demands in terms of Robinson's resource-directing dimensions. Results showed an overall effect of task demands on self-repairs and a measure of accuracy across task types. This implies that resource-directing task demands encourage self-repairs to take place. Based on the result, Gilabert argues that task demands have the potential to foster the interlanguage through fostering awareness of gaps between what they want to say and the target language of the learner.

Gilabert, Raron, and Llanes (2009) investigate the impact of cognitive demands

along task-directing dimensions in three different types of oral tasks (a narrative reconstruction task, an instruction-giving map task, and a decision-making task) on learners' oral interaction. 60 Spanish EFL learners participated in the study. The study focused on interactional moves. They concluded that increasing cognitive demands along the resource-directing dimension generates more interactional moves even though there were some differences in each task type. They argued that this provides evidence for the multiple-resource model, and these kinds of cognitive demands have the potential to draw learner attention to the lexical and syntactic aspects of a language.

1.6 The Methodological Issues

For the reasons enumerated below, however, this series of studies has not resolved certain issues. Firstly, earlier studies did not identify attention being directed to different formal aspects (e.g. syntactic structure, lexical choice, or phonological aspects), and treated them as one package. This distinction is crucial. Izumi (2003) maintained that learners rely more heavily on the use of semantic information or contextual cues than on syntactic information because of their restricted L2 knowledge. In the area of psycholinguistic processing, Clahsen and Felser (2006) and Clahsen et al. (2010) argued that L2 learners have been found to under-use syntactic information, and that they depend on lexical-semantic and pragmatic information because of a shortage of working memory capacity (the shallow-structure hypothesis). From this point of view, attention to semantic, lexical, and syntactic aspects of language might be influenced in different degrees by task-related factors. In order to measure attention to linguistic form in detail, analysis of complexity, accuracy and fluency is too coarse-grained to measure the direction in which attention is oriented. It is widely accepted that fluency reflects a learner's focus on meaning, and the commitment of attention to form can be observed by the complexity or accuracy of the learner's speech production. Hulstijn and Hulstijn (1984), for example, show that attention to form has a positive effect on learner accuracy. However, there is very limited evidence to prove that learner accuracy guarantees paying attention to form. In an empirical study, Fukuta and Yamashita (2013) show the differences between the performance analysis (in terms of complexity, accuracy, and fluency) and analysis of protocol data extracted by stimulated recall. They indicate the difficulty of identifying attention orientation by performance analysis. Fukuta and Yamashita argue that it is impossible to determine solely by performance analysis whether the increase of accuracy scores is the result of

attention to linguistic form (Robinson, 1995b), or is due to avoiding what is not well-known and so likely to lead to errors (Foster & Skehan, 1996).

Secondly, the analysis of markers of interactional moves such as self-repairs, recasts, and language-related episodes (e.g. Gilabert, Raron and Llanes, 2009) provide solutions to these problems. However, this method may only capture quite high levels of attention (possibly the understanding level of attention), which is used only for self-repairs and other interactional moves. That is, it is only the attention that is used for overt repairs. Uptake against recasts can be observed by counting the numbers of the markers used in Gilabert (2007) and Gilabert, Raron and Llanes (2009). L2 learners usually shift their attention from meaning to form without using these markers. Therefore, a more inclusive method is needed to capture attention shifts, or, with respect to the perspective, a methodological triangulation is required.

1.7 Stimulated Recall Prompted by Disfluency-markers

To solve the methodological issues, this study proposes stimulated recall prompted by disfluency-markers(1). The theoretical rationale behind this method is described below. Attention towards linguistic form has often been highlighted by complexity and accuracy scores in the learner's speech production and few researchers disagree with the trade-offs between complexity or accuracy scores, and fluency scores. On the basis of this idea, disfluency-markers are probably produced by learners when conscious syntactic processing overrides semantic processing. In addition, the analysis of self-repair has been regarded as one of the most frequently-used ways of measuring attention to linguistic form, and self-repair is one of the disfluency-markers. Moreover, disfluency-markers as prompts can also minimize the arbitrary nature of the timing of the prompt for stimulated recall. By using this method, inclusive analysis of attention will be achieved via a methodological triangulation of the performance analysis and the protocol analysis. In addition, the direction to which attention is oriented can be analyzed in detail (e.g. syntax, lexicon, or phonology).

2. The Study

To develop a coding scheme and appropriate analytical procedures, I analyzed protocol data elicited by Japanese EFL learners who carried out speech production tasks. Through the qualitative analysis, I attempted to develop an analytical procedure using stimulated recall prompted by disfluency-markers with a coding scheme to

measure learner attention to linguistic form during the tasks.

2.1 Operational Definition of Attention to Linguistic Form

In accordance with the Philp and Iwashita (2013) definition, which is adopted from Mackey et al. (2001), this study defines noticing as a "learner's articulation of response to the input, or to their own output, indicative of a perception of form, without distinguishing the degree of understanding involved (Philp & Iwashita, 2013, p.358)". As a result, the study focuses on "verbalization" as a criterion of a higher level of awareness (the noticing level), and classified verbalized reports to identify the focus of awareness. Since it is remarkably difficult to distinguish the noticing level from the understanding level of awareness (Schmidt, 1990), the present study treats them without distinction [(2)].

2.2 Data Sources

The study re-analyzed the data used in Fukuta (2013a, 2013b). Both studies conducted meaning-focused, narrative tasks (speech production monologues). The participants of the Fukuta (2013a) study were 36 Japanese students at both under- and post-graduate level. Fukuta stated that their proficiency levels were considered to be around the B2 level of the Common European Framework of Reference for Languages (CEFR), estimated from their recent Test of English for International Communication (TOEIC) scores and with an average score of 819.17. TOEIC was developed by the Educational Testing Service. The participants of the Fukuta (2013b) study were 28 Japanese graduate school students. Their TOEIC score was 821.75 on average. All participants in both studies reported that English was their strongest second language. In other words, learners with a high proficiency level in L2 other than English were not included in this study.

Three four-frame cartoons (Fukuta, 2013a) and two six-frame cartoons (Fukuta, 2013b) were used for the speech production task. All participants took part in each study and engaged in the narrative task which required them to describe a cartoon. They were not allowed to take any notes. The speech was audio-recorded[(3)]. Immediately after the speech production task, participants were asked to recall what they had thought during the task while listening to what they had actually said (stimulated recall). While the participants were listening to their own speeches, the researcher halted the audio-recording whenever a participant seemed hesitant, and

asked what they were thinking at that moment. This retrospective interview in the main study was carried out largely on the basis of the guidelines set up by Ericsson and Simon (1980), Gass and Mackey (2000). These retrospective interviews were conducted in their native language, Japanese. The retrospective interviews were audio-recorded and transcribed by the researcher.

The participants in the Fukuta (2013a) study narrate a story of three different tasks respectively. In the Fukuta (2013b) study, 14 participants were asked to tell the story of the same cartoon twice while the other 14 participants were given the same type of task with a different picture and were directed to carry out the task twice. This experiment was conducted twice with a one-week interval.

2.3 Analysis

The retrospective interviews were also transcribed as text files. In order to find the patterns of the learners' underlying mental processes, the transcripts of speech data and protocol data were carefully inspected and qualitatively analyzed.

First, all protocol data are segmented by each prompted trial. As a result, 934 segments from the Fukuta (2013a) study and 434 segments from the Fukuta (2013b) study were collected. The researcher coded and allocated themes to each segment with respect to linguistic features and the concepts they mentioned. These themes were categorized into more theoretical and abstract concepts in terms of their similarities and differences from the data and relevant literature (e.g. de Bot, 1992; Levelt, 1989; Kormos, 2006). In accordance with the categories and the themes, a tentative coding scheme was developed.

3. Findings

3.1 Capturing Attention Orientation

First, five disfluency-markers (false starts, repetition, pauses, use of filler words, and self-repair) were identified from the transcripts of learners' speeches. These markers are shown in Table 1. From the retrospective interview data corresponding to each disfluency-marker, each processing component or "specialist" (Levelt, 1989) was also identified. As described earlier, the specialists consist of conceptualizing, formulating, and monitoring processes. Examples are given below.

Example 1: er one day on the train … ah …

[Retrospective comment]: I intend to explain the situation written in this picture, but err, at first I tried to say "the man sitting center of the chair," without any explanation [of the situation], then I tried to say this was the picture in a train, so I was stuck for something to say (Conceptualizing Process).

Example 2: And he started to ta…he started talking, and …

[Retrospective comment]: I thought I have to say "started talking". (Formulating Process)

***Example 3*:** about the … about his work … or something like that.

[Retrospective comment]: I tried to say "work" as usual, but at first I thought I tried to say other words, but I reconsidered that was not appropriate. So, yeah, this was the problem of expression, I mean, of word. (Monitoring Process)

Table 1

Identified Disfluency-markers

	Definition	Example
False starts	Repeated or restarted speaking just after the beginning of the speech	Ah … on the … on the train … er one day on the train…
Repetition	Repeated parts of the speech without making any correction	One day, **a woman like … a woman like** to … likes to … see the cherry blossoms in the spring.
Pause	Period during which participants say nothing	And finally, the government of the city **… (1.5)** decided to construct the new road.
Filler	Filled pauses using *um* or *err*	that is why the salesman **err** ashamed in **err ah** sorry,
Self-repair	Repeated parts of the speech with corrections	That is why he had to speak, err **speak to speak with** him…

The participant in Example 1 was considering how to describe the picture in the task. This is what Levelt (1989) called *microplanning*. At first the speaker wanted to describe the people in the picture, but then elaborated on this goal by selecting other information which might explain what was happening. In this case, the speaker was not thinking of the choice of lexis, syntax or phonetic representation, but of how to convey the information in the picture. This protocol showed that the speaker focused on

meaning.

It is very difficult to distinguish clearly between monitoring and other processes, and it does not make too much sense to worry about the distinction when measuring attention orientation. In this study, I did not distinguish conceptualizing and formulating from the monitoring process. Instead, both the formulating process and the monitoring of the formulation process were regarded as being attention to form and were analyzed together.

Then monitoring and formulation were classified in detail. As a result, we could see that attention was oriented towards three components: namely syntactic encoding, phonological encoding, and lexical choice. The participant in Example 4 referred to a lexicalized phrase, "a loud voice", and then analyzed the phrases and changed the word "voice" into "noise". This could be regarded as an example of the lexical encoding process within the formulator (Levelt, 1989). On the other hand, the participant in Example 5 further analyzed the speech and drew attention to morpho-syntactic features. This is regarded as syntactic encoding in Levelt's speech production model. The participant in Example 6 referred to pronunciation. Japanese, the first language of all the participants, does not have onset clusters such as /str/, so it is reasonable that the participant encountered difficulties in the phonological/phonetic encoding process.

Example 4: he turn the music with a with a … loud … kind of loud noise…
[Retrospective comments]: I usually and frequently use the expression "loud voice" when I use the word "loud". But this is not about voice but music, so, I confused. (Lexical choice)
Example 5: So she refu er she refused refused to the plan …
[Retrospective comments]: This was, well, the issue of grammar. I thought when I use the word "refuse", I confused whether I should use "ing" or "to+verb", but I cannot remember and I get lost for words. (Syntactic encoding)
Example 6: Constr … constructing …
[Retrospective comment]: I thought I should say it with much better pronunciation. (Phonological/Phonetic encoding)

Finally, the following episodes were coded as *Other*.

Example 7: She thought … umm she thought she … err have to park her bicycle.

[Retrospective comment]: I could not concentrate to speak because of it (finger tapping).

Example 8: One morning … a girl was going to a park.

[Retrospective comment]: umm … sorry I cannot remember what I thought at that time.

Example 7 illustrates the case of a learner's paying attention to an interference task (finger-tapping), not the main task (cartoon narration). In Example 8 the learner cannot recall what they thought because too much time had passed since the task was completed. Since those episodes cannot be categorized as attention to form or meaning, I coded them as *Other*. By this method of analysis, the direction in which a learner's attention was oriented was identified and analyzed (i.e. the lexical, phonological, or syntactic aspect).

The result of this analysis is summarized as a conceptual diagram in Figure 1. We can observe the disfluency-markers in the learner performances (observable level). By analyzing the protocol data corresponding to each disfluency-marker, we can distinguish where attention is directed to the conceptualizing and formulating processes (and the monitoring of each process). Then, closer analysis reveals the direction to which attention is oriented in detail (namely, attention to the syntactic, lexical, and/or phonological process).

I propose a possible coding procedure as follows (see also Figure 2). The order of the coding procedure largely follows the Levelt speech model. First, the protocol segment is examined to see whether it demonstrates attention to conceptual aspects without attention to form. The formal aspect of attention is identified as a lexical item or lexicalized phrase, and so on. If the learners analyzed their language syntactically, the segment is coded as "syntactic encoding". If the segment refers to a phonological- or phonetic-related episode, this segment is coded as "phonological/phonetic encoding". If the segment escapes all classification, it is coded as "other".

3.2 Inter-rater Agreement and Analysis of Problematic Data

To ensure the reliability of this tentative coding scheme and procedure, 10% of the examples in the Fukuta (2013a) study were analyzed both by the author and a second coder (a major in Applied Linguistics in a Masters course in Graduate School). The result was a match of 88.4% (Cohen's Kappa = .807). The author re-coded 30% of all

the data a month later, and the intra-rater agreement was then 94.2 % (Cohen's Kappa = .883). Disagreements and areas where the second coder encountered difficulties were discussed further. The following section shows the results of those discussions.

Prepositions

The participants in the following extracts from Examples 8 and 9 refer to prepositions. However, the attention of each participant is oriented to a different aspect.

Example 8: One day, a man is in the, at the station.
[Retrospective Comment]: I think usage of "at" and "in" is really confusing!

Example 9: Therefore, the man give the guy… give the baggage to the guy.
[Retrospective Comment]: On the way to produce the sentence, I repair what I said because the sentence using "give something to someone" is more suitable.

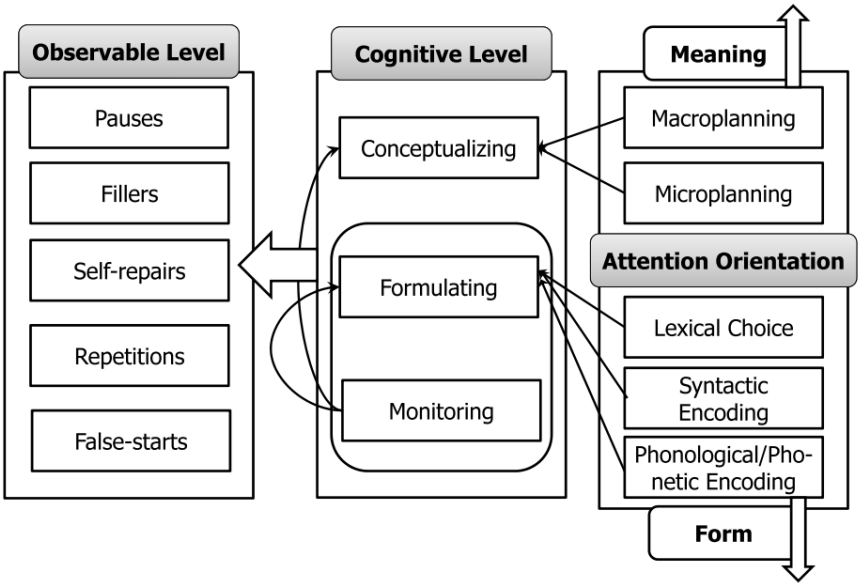

Figure 1. Conceptual Diagram of the Proposed Coding Scheme

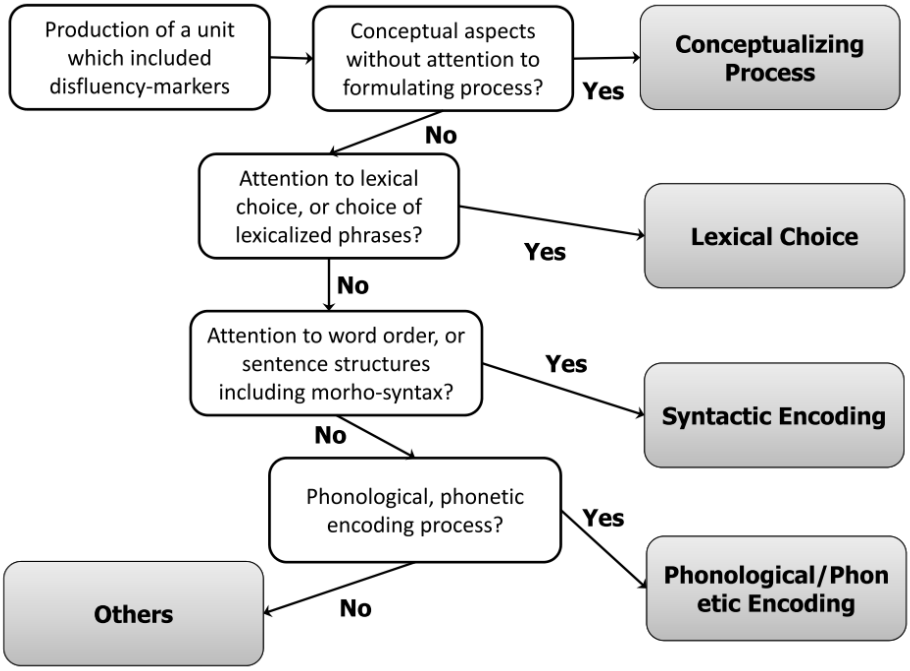

Figure 2. Proposed Coding Procedure

The participant in Example 8 referred to the choice of an appropriate word to express the place of a man in the picture. On the other hand, the participant in Example 9 referred to phrase structure in the subcategory "give". Phrase structure is considered as a syntactic issue; therefore Example 8 can be regarded as lexical encoding whereas Example 9 can be regarded as syntactic encoding.

Conjunctions

The participants in Examples 10 and 11 also refer to the same linguistic items: conjunctions. However, they also pay attention to different aspects.

> ***Example 10:*** It's too… when the man notice that, it's too late.
> [Retrospective Comment]: At first, I start saying "It's too late", but I planned to use "when". So I thought I have to say the clause earlier. So I reproduce the sentence.
> ***Example 11:*** Turn that music down if… when you listen to music.
> [Retrospective Comment]: I thought "when" is more suitable than "if", according to the meaning of connection between the words and what I wanted to say.

The participants in those episodes both referred to adverbial clauses (i.e. "when"). However, the participant in Example 10 draws attention to word order constrained by syntactic rules. Example 11 shows the participant's attention focused on choosing the appropriate word to express the meaning. Therefore, Example 10 can be regarded as attention to syntactic encoding where Example 11 shows attention being paid to lexical encoding. Following on, Examples 12 and 13 also refer to conjunctions (i.e. "and", "but" and "so"). Example 12 is in essence the same as Example 11. In both, the learners paid attention to an appropriate lexical choice. On the other hand, the participant in Example 13 was concerned with whether what was said could be understood. This could be considered as attention to the discourse level of processing in contrast with microplanning, attention to linguistic form.

> ***Example 12:*** He noticed that his bicycle disturb the way, but… and so, he was ashamed.
> [Retrospective Comment]: I said "and", but I noticed it's a mistake, so I said it again.
> ***Example 13:*** He afraid that the park was destroyed, so… destroyed. Therefore, he

decide to resist that.

[Retrospective Comment]: I intended to connect the two sentences by “so”, but I thought I should cut the sentence into two pieces because It’s easier to understand.

3.3 Summary of the Analysis

These results suggest that the learner’s focus on form can be identified by protocol analysis. When eliciting retrospective data by stimulated recall, disfluency-markers acting as prompts help reduce issues caused by the interval that has passed since the task was performed. The qualitative inspection showed that analysis of the attention to linguistic form can be classified in terms of attention orientation (to syntactic encoding, phonological encoding, and lexical choice). As for problematic data, the study proposes some solutions and coding examples.

The study also shows an element of inter-coder disagreement. As I noted earlier, it is difficult to distinguish between the conceptualization process and other linguistic processes. These difficulties are not so often apparent when learners engage in narrative tasks, and therefore it can fall within an acceptable error range for statistical analysis. However, if this was to appear more frequently (it was less than five percent in this study), it would be preferable for researchers to state how it was treated in the analysis.

4. Conclusion

When researchers interpret their data it is important to be sure that everyone is using the same language. To achieve this, the present study used a two-tiered approach. First, this paper summarized the methods that can be used to measure attention to linguistic form as used in previous studies. The paper then tried to create a comprehensive, accessible and detailed guide on how to identify attention orientation in protocol data. In the critical review of previously-used methods, the paper debated the issues they raised and then proposed the usefulness of stimulated recall prompted by disfluency-markers outlining its theoretical rationale. Secondly, the study aimed to provide a useful resource for the researcher. The coding scheme used in the present study was achieved by qualitative analysis of protocol data. Problematic data which did not easily fit the definitions of the coding scheme were also discussed.

The coded data can be analyzed both qualitatively and quantitatively. For example, qualitative analysis can reveal how learners react to form-focused feedback from

teachers by analyzing the underlying awareness-related cognitive process. Quantitative analysis can tell us to what extent learners direct their attention to form under certain task conditions, and whether the attention to form triggers language acquisition. These issues are of central importance both to research into the theory of language acquisition and to teaching L2.

The proposed coding outlined in this study does not pretend to offer a complete set of distinctions and explanations, nor is it offered as a perfect guide for treating problematic data with sufficient examples for this purpose, although I have tried to supply as many examples as possible. The coding scheme and procedure should be validated further and extensive examples should be provided in future studies. I hope the present study will stimulate further research that may lead to a more sophisticated and accessible method for capturing attention shifts during L2 speech production. I also hope that the method presented in this study will add to the empirical data that is related to attention shifts, and further, will shed light on the role played by attention, one of the key concerns of L2 acquisition theory and practice.

Acknowledgment

An earlier version of this research was presented at the conference of the Japan Association for Language Education and Technology, Kansai Chapter, Methodology SIG (The consortium of universities in Akita, 2013). I am grateful to Professors Junko Yamashita and Masatoshi Sugiura, of Nagoya University, for their valuable comments. I also acknowledge the insightful comments on an earlier draft of this paper by Mr. Yuichi Suzuki, of The University of Maryland. Any remaining errors, of course, are my own. This research was partially supported by the Japan Society for the Promotion of Science.

Notes

1. While I was writing this paper, another empirical research paper using stimulated recall prompted by disfluency-markers (Philp & Iwashita, 2013) was published. It describes the following implementation procedure: “The interviewer paused the recording at points in which either partner modified output or demonstrated any difficulty with expressing themselves. The participant could also choose to pause the DVD at any time. Once the DVD was paused, the interviewee was then asked to verbalize what he or she was thinking at the time of the interaction” (p. 358). This paper, however, did not present a more detailed implementation

procedure including a coding scheme. At this point, I believe that my study can still provide useful perspectives for further empirical study.

2. There are a number of ways to address awareness levels from protocol data. Leow (2001), for example, distinguished two levels by the use of think-aloud protocol through the following criteria: +/- cognitive change (whether participants show some behavioral or cognitive change due to the experience), and +/- meta-awareness (whether they report awareness of the experience and/or describe the underlying linguistic rules). The stimulated recall, unlike think-aloud, has much potential to distort the quality of reporting of awareness levels because of the veridicality problem. Therefore, this study focused only on attention orientation, and not on the level of awareness.
3. Stimulated recall usually utilizes audio- or video-recorded data as a stimulus for the support of memory retrieval. The two previous studies used audio-recorded stimuli because they did not want to focus on non-verbal behavior during the learners' tasks.

References

Bloom, B. (1954). The thought process of students in discussion. In S.J. French (Ed.), *Accent on teaching: experiments in general education* (pp. 23-46). New York: Harper & Brothers.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3–42. doi:/10.1017/S0142716406060024

Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological Structure in Native and Nonnative Language Processing. *Language Learning, 60*, 21–43. doi:10.1111/j.1467–9922.2009.00550.x

De Bot, K. (1992). A Bilingual Production Model: Levelt's "Speaking" Model Adapted. *Applied Linguistics, 13*, 1–24.

Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language and Cognition, 16,* 1–15. doi:10.1017/S1366728911000629

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and Second Language Instruction.* (pp. 206‐257). Cambridge: Cambridge University Press.

Egi, T. (2004). Verbal reports, noticing, and SLA research. *Language Awareness*, *13*, 243–264. doi:10.1080/09658410408668810

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data (Revised Edition). Cambridge, MA: MIT Press.

Faretta-Stutenberg, M., and Morgan-Short, K. (2011). 'Learning without awareness reconsidered: A replication of Williams (2005).' In G. Granena et al. (eds.): *Selected Proceedings of the 2010 Second Language Research Forum*, 18-28.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*, 299‐323. doi:10.1017/S0272263100015047

Fortune, A. (2005). Learners' use of metalanguage in collaborative form‐focused L2 output tasks. *Language Awareness, 14*, 21‐38. doi:10.1080/09658410508668818

Fortune, A. & Throp, D. (2001). Knotted and entangled: New light on the identification, classification and value of language related episodes in collaborative output tasks. *Language Awareness, 10*, 143‐160. doi:10.1080/09658410108667031

Fukuta, J. (2013a). *Effects of Task Demands on Attention toward Linguistic Form in L2 Speech Production*. Unpublished Master's Thesis, Nagoya University.

Fukuta, J. (2013b) *Effects of task repetition on learners' attention shifts in L2 speech production.* Paper presented at 39th Annual Conference of Japan Society of English Language Education, Hokkaido, Japan.

Fukuta, J., & Yamashita, J. (2013). *Effects of Simple versus Complex Tasks on L2 Learners' Attention to Form in Oral Production.* Paper presented at the annual conference of the Second Language Research Forum (SLRF) 2013, Utah, the United States.

Gass, S. (1988) Integrating research areas: A framework for second language studies. *Applied Linguistics, 9*, 198–217.

Gass, S., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum.

Gilabert, R. (2007). Effects of manipulating task complexity on self–repairs during L2 oral production. *IRAL – International Review of Applied Linguistics in Language Teaching, 45,* 215–240. doi:10.1515/iral.2007.010

Gilabert, R., Raron, J., & Llanes, A. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance : *IRAL – International Review of Applied Linguistics in Language Teaching, 47*, 367‐395. doi:10.1515/iral.2009.016

Godfroid, A., Housen, A., & Boers, F. (2010). A Procedure for testing the noticing hypothesis in the context of vocabulary acquisition. In M. Putz & L. Sicola (Eds.), *Cognitive Processing in Second Language Acquisition*, pp. 169–197. Amsterdam: John Benjamins Publishing Company.

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition, 32*, 465‐491. doi:10.1017/S0272263110000045

Hulstijn, J. H., & Hulstijn, W. (1984). Grammatical errors as a function of processing constraints and explicit knowledge. *Language Learning, 34*, 23-43. doi:10.1111/j.1467-1770.1984.tb00994.x

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics, 24*, 168–196. doi:10.1093/applin/24.2.168

Jourdenais, R. (2001). Cognition, instruction and protocol analysis. In P. Robinson (Ed.), *Cognition and Second Language Instruction.* (pp.354‐376). Cambridge: Cambridge University Press.

Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning, 50*, 343–384. doi:10.1111/0023–8333.00120

Kormos, J. (2006). *Speech production and second language acquisition*. London: Routledge.

Kowal, M. & Swain, M. (1994). Using collaborative language production tasks to promote students' language awareness. *Language Awareness*, 3, 73‐93

Leow, R. P. (1997). Attention, awareness, and foreign language behaviour. *Language Learning.* 47, 133‐59. doi:10.1111/j.1467–1770.2001.tb00016.x

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: aware vs. unaware learners. *Studies in Second Language Acquisition. 22*, 577‐584.

Leow, R. P., and Hama, M. (2013). Implicit learning in SLA and the issue of internal validity. *Studies in Second Language Acquisition, 35*. 545-557.

Leung, J. H. C., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition, 33*, 33‐55. doi:/10.1017/S0272263110000525

Levelt, W. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.

Norris, S.P. (1990). Effect of eliciting verbal reports of thinking on critical thinking test performance. *Journal of Educational Measurement, 27*, 41–58.

Ortega, L. (2009). *Understanding Second Language Acquisition.* London: Hodder Education.

Philp, J. (2003). Constraints on "noticing the gap." *Studies in Second Language Acquisition, 25*, 99‐126. doi: /10.1017/S0272263103000044

Philp, J. & Iwashita, N. (2013). Talking, tuning in and noticing: exploring the benefits of output in task-based peer interaction. *Language Awareness, 22*, 353-370. doi:10.1080/09658416.2012.758128

Russo, J.E., Johnson, E.J. & Stephens, D.L. (1989). The validity of verbal protocols. *Memory and Cognition, 17*, 759–69.

Rebuschat, P., Hamrick, P., Sachs, R., Riestenberg, K., and Ziegler, N. (in press). 'Implicit and explicit knowledge of form-meaning connections: Evidence from subjective measures of awareness.' In J. Bergsleithner, S. Frota, and J. K. Yoshioka (Eds.): *Noticing: L2 studies and essays in honor of Dick Schmidt*. Honolulu, HI: University of Hawai'i at Manoa, National Foreign Language Resource Center.

Robinson, P. J. (1995a). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45,* 283–331. doi:10.1111/j.1467–1770.1995.tb00441.x.

Robinson, P. J. (1995b). Task complexity and second language narrative discourse. *Language Learning, 45*, 99–140. doi:10.1111/j.1467–1770.1995.tb00964.x

Robinson, P. J. (1995c). Aptitude, awareness and the fundamental similarity of implicit and explicit second language learning. In R. Schmidt (Ed.), *Attention and Awareness in Foreign Language Learning*, (pp.303-358). University of Hawai'i at Manoa, Second Language Teaching and Curriculum Center, Technical Report no 9: University of Hawai'i Press.

Robinson, P. J. (1996). Learning simple and complex second language rules under implicit, incidental, rule‐search, and instructed conditions. *Studies in Second Language Acquisition. 18*, 27‐67.

Robinson, P. J. (1997), Generalizability and automaticity of second language learning under implicit, incidental, enhanced and instructed conditions. *Studies in Second Language Acquisition. 19*, 223‐247.

Robinson, P. J. (2001a). Task complexity, task difficulty, and task production: exploring interactions in a componential framework. *Applied Linguistics, 22*, 27–57. doi:10.1093/applin/22.1.27

Robinson. P. J. (2001b). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 287–318). Cambridge: Cambridge University Press.

Robinson. P. J. (2003). Attention and memory during SLA. In C. Doughty and M. Long (Eds.) *Handbook of Second Language Acquisition* (pp. 631–678). Oxford: Blackwell.

Robinson. P. J. (2005). Cognitive complexity and task sequencing: A review of studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching, 43*: 1–33. doi:10.1515/iral.2005.43.1.1

Rosa, E., & Leow, R. P. (2004). Awareness, different learning conditions, and second language development. *Applied Psycholinguistics, 25*, 269–292. doi:/10.1017/S0142716404001134

Rosa, E., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness. *Studies in Second Language Acquisition, 21*, 511–556. doi:/10.1017/S0272263199004015

Russo, J. E., Johnson, E. J., and Stephens, D. L. (1989). The validity of verbal protocols. *Memory and Cognition, 17*, 759‑769.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158. doi:10.1093/applin/11.2.129

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3‑32). New York: Cambridge University Press.

Shanks, D. R., Green, R. E., & Kolodny, J. A. (1994). A critical examination of the evidence for unconscious (implicit) learning. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 837‑860). Cambridge, MA: MIT Press.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17*, 38–62. doi:10.1093/applin/17.1.38

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*, 93–120. doi:10.1111/1467–9922.00071

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*, 371–391. doi:10.1093/applin/16.3.371

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*, 269‑304. doi:/10.1017/S0272263105050138.