Second Language Spoken Fluency in Monologue and Dialogue

Steven John Kirk, BSc, MA

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

July 2016

Abstract

Although second language spoken fluency has long been recognized as a major component of language proficiency, it has never been clearly defined. It has been shown that fluency is a complex phenomenon, with a host of relevant factors, and it has been suggested that it might be better separated into multiple concepts, such as cognitive fluency and utterance fluency. There is also evidence that fluency has a dialogic aspect, that is, that the fluency of a conversation is a co-construction of the two speakers, rather than simply alternating monologues. This can be observed in the confluence created by smooth turn exchanges, which results in minimizing gaps and avoiding overlap.

The present study seeks to examine the co-construction of dialogic fluency through a parallel case study of two Japanese learners of English. One learner was of lower-intermediate proficiency, and the other was of higher proficiency, but both were able to create good impressions of fluency in conversations with native speakers of English. The case study design was semi-experimental in that it involved a story-retelling task done in monologue and dialogue, which was repeated to take into account the effect of practice. The case study allowed the close examination of the construction of fluency in the story-retelling task moment-by-moment through the course of the retellings, taking into account all relevant factors. The semi-experimental, parallel case study design allowed the findings to be compared (1) between monologue (where the learner recorded herself telling the story alone) and dialogue (where the learner told the story to a native speaker interlocutor), and (2) between the two learners of differing proficiency. This study was also mixed-methods in that it combined a qualitative, grounded theory approach to data analysis involving discourse analytic techniques, with quantitative comparisons of

temporal variables of fluency. It was also multi-modal in that video was employed to take into account gaze, gesture, and head nods.

Results of quantitative analyses revealed that the dialogues were comparatively more fluent than the monologues in terms of speech rate, articulation rate, and length of silences, for both speakers, although the higher-proficiency subject had faster speech and articulation rates than the lower-proficiency learner. This implies that narrative in dialogue is not just a listener occasionally backchanneling while the speaker delivers a monologue. The qualitative analyses revealed that the co-construction of smooth conversation was facilitated by the alignment of rhythm between the speaker and listener, supported by gaze, gestures, and head nods. The learners in these case studies were able to employ different fluency techniques for stressing words in phrases to create rhythm in spite of lower speech rates, and were able to adjust those techniques to maintain rhythm with even lower speech rates at difficult points of the story.

These results confirm previous research that some apparent "dysfluencies" in speech should be considered as speech management phenomena, that positively contribute to the co-construction of fluent conversation. They also suggest that alignment between the speakers in terms of rhythm of speech and gaze are important in conversation, confirming previous research showing alignment at these and other levels of interaction. Finally, it appears that fluency is a multi-level construct, and that dialogic fluency should be considered a separate construct from cognitive fluency, of equal or more importance. This has implications for language testing, such that fluency may not be able to be captured with single test types, and for language teaching and learning more generally.

Acknowledgements

There are many people who have helped and supported me while working on this thesis. First and forement, I would like to thank my supervisor, Ron Carter. Over the past several years, he has always been positive and motivating, especially through those difficult times of self-doubt that many people experience while writing a PhD dissertation, and without him, I couldn't have done it. He was always willing to give me time to discuss my writing and ideas, and whether before or after retirement, always got back to me with comments in record time.

Secondly, I would like to thank Mike McCarthy, my unofficial extra supervisor. Not only did he help me and inspire me with my research, but also he often reminded me not to forget the other important things in life.

Next, I must thank Chris Candlin for recommending Nottingham to me. He said that it would be a perfect fit for my research interests, and he was right. Also, I owe a lot to Zoltán Dörnyei, whose advice helped me to stay true to myself throughout this project.

There are many other friends and family who helped and supported me through this project in various ways—too many to name them all. But finally I would like to thank to Mike Handford, Natsuko Ohguri, Mika Yamaguchi, Alex Gilmore, Dawn Knight, Phoebe Lin, and of course, my family.

Finally, a special thank you to my mother, Marlene, who would have been so happy to see me finish this.

Table of Contents

1 Introduction

1.1 Why fluency?

Anyone who has studied a foreign language would probably agree that a major goal is to be able to speak the language “fluently”. Although in common parlance this is sometimes synonymous with *nativelike proficiency*—particularly as used in the phrase “to be fluent in (a language)”, *fluently* generally implies flow or smoothness, especially in spoken language production. In my own experience as a second language learner of Japanese, the fluency that I hope to attain means being able to say what I want to say easily and immediately, that is, without the pain and frustration that second language learners often feel as they struggle to express themselves. In other words, fluency is mainly concerned with being able to *speak* easily under the time pressures of everyday conversation. I can take as much time as I need to craft an email in Japanese, but especially in conversations involving multiple Japanese native speakers, I often cannot produce the language in a manner that is timely enough to catch those fleeting openings in the conversation. On the other hand, in conversations involving interaction with a single Japanese speaker, I generally feel at ease, and although I still have moments of frustration with expressing what I want to say, this does not generally cause problems with the conversation. The interlocutor might wait patiently for a moment, or offer suggestions to help me complete my thought. That said, the success of the conversation also varies with the particular individual—some I feel are more supportive than others. These experiences led me toward the idea that my success in a conversation might not only depend on my ability, but also on the interlocutor's contribution.

Being able to keep up your end of the conversation and contribute to a comfortable interaction also seems to be an important aspect of fluent speech (Ejzenberg, 2000). In Japanese, I often find myself trying hard to make a good showing of fluency when I first start talking with someone, so that they do not talk down to me or treat me like someone who cannot speak Japanese. Then, I do my best to maintain the conversation, even if it means that I say what I can rather than what I really think. Using discourse markers and nativelike hesitation devices also helps to convey a positive image of fluency even when I don't feel so confident with the language myself. That said, my ability to maintain a smooth conversation with a native speaker depends to a great degree on many aspects of the context. It depends on the other person, our relationship, and our familiarity with each other. It also depends on the topic of the conversation and whether the topic is focused on shared knowledge or whether it involves asking about me.

The study described in this thesis focuses on a language learner not dissimilar to myself, who I refer to as Asami. Like me, she is not really what you would consider a high-proficiency speaker, but like me, she is able to keep up her end of conversations with native speakers very well. Again, like me, this involves a degree of confidence, or at least a showing of confidence, as well as disregard for the consequences of making mistakes with the language. In my teaching experience, students like Asami are wonderful in class, because they have plenty to learn, but they also have the confidence and willingness to speak that can help them improve. They are able, when they have the opportunity, to participate in conversations with proficient speakers that both participants can enjoy. Despite the presence of students like Asami, the reality is that much more difficult cases abound. Just recently, I had a conversation with two students, where I did all of the speaking. The two students in question both had fairly high knowledge of English grammar and vocabulary, but could not apparently participate actively in conversation. They came to ask me a question about the project they were working on, but when I asked them to explain their problem, there was a painfully long silence. One of the students had sent me an email a few days before, so I assumed that the question was related to that content. To make a long story short, the entire conversation progressed with me guessing what they wanted to ask based on my knowledge of that previous mail, and the students merely nodding or shaking their heads in reply. Thus, the conversation became kind of a game of twenty questions, with me doing all of the talking. This conversation felt like a struggle for both me and, presumably, the students.

Fluency has been a notoriously difficult notion to define. Even distinguishing it from *proficiency* is something that never been clearly done in the field of Applied Linguistics (Fujimoto, 2012). There is a sense that *fluency* describes the apparent continuity, smoothness, and speed of spoken language, but this impression of smoothness and continuity is an illusion (Goldman-Eisler, 1968), as most spoken language is full of silences, hesitations, restarts, and repetition. As Rühlemann (2007) notes, there is no perfect fluency. The spoken language of native speakers also shows signs of planning, re-planning, searching for words, and adjusting and re-adjusting language to take the current context into account. This problem is somewhat avoided with learners of English by comparing them with proficient speakers. As a result, too often a key problem with studies of fluency is that the fluency of learners tends to be described in terms of what they *can't* do, rather than what they *can* do. This leads to an unrealistic view of learners' language, with hesitations viewed as problems rather than the natural product of producing spoken language in real time.

1.2 Fluency as co-constructed

Conversation is one of the most basic forms of language use, and may have evolved with two people as its most basic context (Stivers, 2015). In the mainstream cognitive view of second language acquisition, conversation tends to be viewed as a ping-pong of alternating monologues, bouncing back and forth between the two people. This view has been reinforced by the fact that conversation is often studied in written transcription, which McCarthy (in the preface of Rühlemann, 2007) notes makes conversation appear to be created sequentially in space, while in fact it is created by a range of simultaneous factors beyond the content of the speech itself, including prosody, gaze, rhythm, and gestures. The skillful orchestration of all of these features, combined with the social preference for one person to speak at a time, create conversations that have a fluency or smoothness resulting from the combined speech of the two participants. McCarthy refers to this co-constructed fluency of the conversation as *confluence* (McCarthy, 2005), describing how speakers complement each other in conversation, for example, by finishing each other's sentences or adding to each other's statements, and by using discourse markers and "heads" and "tails" (Carter & McCarthy, 1995) to connect their turns with the turns of their interlocutor. Through the use of such conversational devices, two proficient speakers are able to create a conversation that feels normal and fluent to the participants in spite of these measurable "dysfluencies".

Conversation, in its natural form, is something that is done socially. Even in the field of language testing, the need for a social dimension has come to be recognized (McNamara & Roever, 2006). For example, Brown (2003, 2005) found that oral interviewees received higher scores when examiners were more supportive. Fujimoto (2012) has argued for oral testing to be done in groups or pairs, with the same grade given to the entire group. Some psycholinguists have noted that conversation is easier, and therefore more basic, than monologic speech, due to the fact that speakers affect and support each other in multiple, sometimes unconscious, ways (Garrod & Pickering, 2004). The present study, therefore, aimed to look more closely at the concept of *confluence*, or what I term *dialogic fluency*, in order to explore the details of the complex process by which fluency is co-constructed by participants in a conversation. Although previous research has described these dialogic features of fluency (for example, Fiksdal, 1990; Götz, 2013; McCarthy, 2005; Rühlemann, 2007), the details of these processes in their complexity have not been well-examined.

1.3 Research methods

The present study focuses on spoken fluency through parallel case studies of two Japanese learners of English, using multimodal data, and mixed methods of analysis. Much previous research on spoken fluency (for example, the work of Goldman-Eisler and later pausologists) and aspects of the co-construction of conversation (such as McCarthy's work on *confluence*) has looked at data involving native speakers of the language. This is natural in that it is assumed that native speakers of the language are fluent. In the present study, I focused on two learners of English, and particularly on one learner who was not at an advanced level of proficiency. Studying L2 data is of course useful from the point of view of possible applications to language teaching, but it also can highlight features that might go unnoticed in conversational data involving only native speakers. Aspects of dialogic fluency that are easily maintained by proficient speakers in a conversation may not be salient to the researcher until problems are observed with these aspects in conversations involving lower-proficiency learners.

Previous research on the fluency of second language speakers has generally employed small sets of data, analyzed using a microanalytic approach with temporal features timed to the millisecond, and quantitative methods comparing higher and lower fluency learners, as judged holistically by native speaker raters (for example, Riggenbach, 1989; Towell, Hawkins, & Bazergui, 1996). The time-intensive nature of transcribing spoken data has generally made corpus-based studies of spoken fluency difficult, and when these have been done (such as Götz, 2013; Rühlemann, 2007), they lack many of the detailed features and accurate timings of the microanalytic approach. Additionally, very little of this previous research has been able to incorporate gestures and gaze with video data. Research that has been done with this type of data has generally been focused on gestures and gaze (for example Knight & Adolphs, 2008), rather than fluency more generally. Recent technological advances with software and recording devices have made it possible to explore new means of recording and analyzing multimodal spoken data with transcriptions and video time-aligned to the audio data (Adolphs & Carter, 2013; Adolphs, Knight, & Carter, 2015). In the present study, I used both audio and video data in order to include as much of the relevant context as possible in the analysis—particularly gestures, head nods, and gaze.

Generally, research in second language spoken fluency has sought to measure a variety of (mostly temporal) variables with the goal of determining which variable correlates most highly with ratings of fluency made by native speaker evaluators—a search for *one factor to rule them all*. The results of this approach to researching fluency not been conclusive, nor very theoretically satisfying. Riggenbach (1989), for example, concluded that fluency was a collection of related features functioning together, with a problem in one or more of these features contributing to an impression of weak fluency. She also suggests that fluency is a nonlinear dynamic system, and that fluency can be achieved (or not achieved) in several different ways, with several profiles of fluent or nonfluent speakers. More recently, Götz (2013), in a corpus-based study, has come to a similar conclusion. With such a large number of interrelated factors all being claimed to influence judgments of the fluency of learners, in the present study, a detailed case study methodology was employed. Case studies allow the researcher to examine one case moment-by-moment to see how various factors interact, and how fluency is created or lost as the speech unfolds. Unlike quantitative methods, case studies allow researchers to notice details in the data that go against the trends, which can often be a key to a deeper understanding of the subject. In the present research, two parallel case studies were conducted with a lower-intermediate learner and an advanced learner. This allowed further comparison between the two cases, with the advanced learner serving as kind of a baseline for comparison.

The present study uses mixed methods in several ways. In terms of the collection of data, the study was naturalistic in that the speech and conversations were spontaneous retellings of the story of a movie, but it was also semi-experimental in that the topic was controlled, and the participants were asked to repeat the retelling in monologic and dialogic contexts. Furthermore, the analysis of the data employed a variety of methods to triangulate the data, and facilitate comparisons with previous research. For example, statistical methods were used to compare across the variable of monologue vs. dialogue, and between the two learners. Corpus linguistic methods were used to analyze the use of discourse markers and filled pauses, albeit with a very small corpus. Discourse analytic methods were used to examine the construction and co-construction of fluency as evidenced in detailed transcripts of the recordings.

1.4 The structure of the thesis

In the following two chapters, this introduction will be followed by an overview of the literature upon which this study is based. Fluency is a complex aspect of language, which has been approached from many angles and points of view over the past several decades. Therefore, the review of the literature was divided between two chapters. In Chapter 2, I first review the multiple and sometimes contradictory definitions of *fluency* that have been proposed in the literature, comparing these with the folk definitions and contrasting them with the concept of *proficiency*, and relating them to the common sense perceptions of speech as being “smooth” or “flowing”. This is followed with a review of temporal phenomena related to fluency, including rate of speech, silent pauses, and other hesitation phenomena. I then look at work on the psycholinguistic bases of fluency in terms of automaticity of language production, and in particular the relationship of *formulaic sequences* to fluent speech production, particularly focusing on the issues related to second language spoken English. Chapter 2 closes with a look at models of the factors affecting spoken fluency (Segalowitz, 2010) and of second language speech production (Kormos & Dénes, 2004). This chapter also reviews the important distinction drawn by Segalowitz (2010) between *cognitive fluency*, reflecting the automaticity and speed of the underlying cognitive processes involved in speech production, *utterance fluency*, the resulting observable output of those cognitive processes measured with the various temporal variables of speech, and *perceived fluency*, the impression on the listener of the smoothness of a speaker’s speech production processes.

The literature presented in Chapter 2 looked at fluency in terms of cognitive processes and as a property of the speaker in isolation—what I term *monologic fluency*. In Chapter 3, I review literature dealing with interactive elements of fluency. In the first half of the chapter, I look at factors that have been shown to influence the *perceived fluency* of a speaker, other than the temporal factors discussed in the previous chapter. These include prosody and the use of non-verbal aspects of communication, such as facial displays and hand gestures. Then, I examine the role of the listener in conversation, reviewing research that shows how the listener and the speaker co-construct the fluency of the conversation, through the use of discourse markers and listener responses. Turn boundaries are crucial points in the co-constructed fluency of a conversation, and I will review literature on what contributes to make these smooth, and to minimize gaps and avoid overlaps, which has been argued to be a universal aspect of natural conversation (Stivers et al., 2009). Finally, I will examine the interactive alignment that has been shown to occur between speakers in a conversation.

In the second half of Chapter 3, I present the results of a pilot study, the purpose of which was to bring together the various factors associated with both monologic and dialogic fluency and examine them in a detailed study of a single case in order to examine the influence of dialogic aspects of fluency. The case in question involved a lower-intermediate Japanese learner of English in free conversation with a native speaker. This particular learner was not chosen for any particular reason other than that she volunteered for the study and was not particularly stronger or weaker than most of her peers. Unexpectedly, in the conversation with the native speaker, she was able to create a very favorable impression of fluency on the listener in spite of weaker utterance fluency, as measured by the traditional temporal variables. The results of this pilot study led to the formulation of the research questions for the main study, which were aimed at looking at the interaction of the features of monologic and dialogic fluency in a more controlled context.

Chapter 4 presents the methodology of the main study. This is a parallel case study of two Japanese learners of English retelling the story of a short movie, the "Pear Story" (Chafe, 1980b). Each subject was asked to retell this story four times—twice in monologue (alone with an IC recorder), and twice in dialogue (with a native speaker interlocutor). Each genre (monologue and dialogue) was done twice to take into account practice effects. The dialogues were also video-recorded in order to create a multi-modal transcription including head nods, gestures, and gaze. This study is mixed-methods in that it combined features and methodologies of qualitative and quantitative research methods. It is qualitative in that it involved an in-depth look at the particulars of two specific cases—specifically, the lower-intermediate learner from the pilot study of Chapter 3 and a second advanced learner. On the other hand, it also had features of experimental research in that the content of the story retellings was controlled, and there were clear (but not necessarily independent) variables involved—i.e., the subject (implying proficiency) and the genre (monologue or dialogue). The advanced learner represents a baseline for comparison, as a non-native speaker of the same L1 at a higher level of proficiency. The lower-proficiency learner was interesting in that she showed weak fluency in terms of temporal variables, but was able to project a strong impression of fluency on her interlocutors. Therefore, one of the aims of this study was to determine how this lower-proficiency learner could compensate for her weaknesses in monologic fluency, and to what extent this behavior could be compared with that of the higher-proficiency learner. This study is also mixed-methods in terms of the procedures used to analyze the data. Statistical methods were used to compare features of temporal fluency, and corpus linguistic methods were used to examine the use of lexis, such as filled pauses and discourse markers. Additionally, discourse analysis techniques were used within a grounded theory approach to examine dialogic features of fluency, such as non-verbal features.

Chapters 5, 6, and 7 comprise the main analyses of the study. Chapter 5 is the quantitative part of this study and examines features associated with *utterance fluency*, especially those that have been employed in traditional fluency research since the 1950's. These include temporal variables such as speech rate, articulation rate, the length and location of silent pauses, and the length of uninterrupted runs of speech. They also include the use of those aspects of speech that have been referred to as *speech management*, *hesitation phenomena*, and *dysfluencies*, such as filled pauses, repetition, restarts, and false starts. These factors have typically been seen as a sign of the limitations of *cognitive fluency*, with the caveat that a speaker may not necessarily always speak at the limit of their cognitive fluency ability. Additionally, recurrent sequences in the multiple retellings of the same story was examined to look for signs of the use of formulaic sequences, which are claimed to be very important to spoken fluency (Pawley & Syder, 1983; Segalowitz, 2010).

Chapter 6 begins the analysis of dialogic aspects of fluency using segments taken from the eight Pear Story retellings from both subjects. The segments chosen comprised the *complication* segment of the story (Labov & Waletzky, 1967). This segment was chosen because it occurred in all eight retellings—and because it does not occur at the beginning or ending of the retellings. Once the story had been introduced and set-up in the first part of the retelling, the complication scene could be told. Each of the eight segments was studied line by line, examining the use of prosody, gestures, gaze, discourse markers, and speech management devices. The listener was also included in this analysis, looking at head nods and verbal responses.

Additionally, the establish of rhythm created by the placement of stress throughout utterances was analyzed, using the methods of Auer et al. (1999) and Fiksdal (1990). As features were identified throughout this analysis, they were compared to previous noted examples to confirm or update hypotheses generated by the data, using an approach modeled on the method for doing Conversation Analysis outlined by Have (2007), which is similar to a *grounded theory* approach.

Chapter 7 brings together the conclusions drawn from the line by line analysis of Chapter 6, organized by themes. First, I examine rhythmic alignment and its role in the co-construction of fluency in conversation. In the second part of the chapter, I discuss the use of discourse markers and speech management strategies. Following that is an examination of how listener responses, both verbal and non-verbal, are aligned with the speaker—both in terms of the co-constructed rhythm of the conversation, and in terms of the creation of a shared attention space between the speakers. The next section of the chapter deals with turn boundaries and how the subjects and their interlocutors work together to successfully negotiate these. Finally, the effect of listener feedback on the story itself is also examined, comparing how each subject expresses the same concepts in the monologues and dialogues.

Finally, Chapter 8 is the conclusion of the thesis. In this chapter, I discuss the models of the factors influencing fluency presented by Segalowitz (2010) and Rühlemann (2007), and how these can be updated to include a focus on dialogic fluency. I also discuss how this study may fit within the sociocognitive approach to second language acquisition outlined by Atkinson (2011). The chapter continues with an initial consideration of implications for teaching and testing of L2 spoken language, as well as a discussion of the limitations of the current study and the methodological implications for future research in second language spoken fluency.

Finally, the chapter concludes by examining the implications of this dissertation for our understanding of the nature of fluency and the need to redefine the traditional view of this concept to better capture the complexity of fluency in the case of L2 language learners.

2 Fluency

2.1 Defining fluency

Fluency is a concept that is central to the fields of English language teaching, second language acquisition, and linguistics in general. However, it is still not well understood and notoriously hard to define. Its importance in these fields can be seen in the frequent use of the term in second language teaching methodology, where it is often contrasted with *accuracy*, and its frequent appearance in rubrics for the testing of speaking. In fact, it is one of the main goals of language learning, if not *the* main goal.

Although *fluency* is a central concept, it has proven very hard to define due to ambiguity in the term. First, there exists a lay term *fluency* with its own uses and connotations. Second, there are differences in researchers' definitions, with there being two broad categories of definitions. One meaning is close to general proficiency in the language, and the other is a more technical sense, related to particular features in the spoken language, such as rate of speech. This is further complicated by the use of the term to refer to native speakers as well as nonnative speakers. For example, are all native speakers by definition fluent, or do native speakers vary in levels of fluency? Should nonnative speaker fluency be described in terms of resemblance to native speakers or should nonnative language be analyzed on its own terms? Guillot (1999, p. vii) notes that fluency "crosses over boundaries in a way which has made it resistant to analysis", and then adds

> Yet it is peculiarly available to all, language specialists and non-specialists, as a measure of oral performance, and is used with a confidence which hardly seems justified in view of the scarcity of accounts governed by anything other than intuition. (1999, p. vii)

Clearly, *fluency* is an important concept that needs to be unpacked. In the following sections of this chapter, I will look more closely at previous definitions of fluency in order to clarify this issue and arrive at a more clear and useful concept for second language acquisition. Although it is possible to speak of fluency in writing as well as speaking, and in receptive skills such as reading and listening, in this chapter I will only focus on spoken language, and only touch on listening as it relates to fluency in speaking.

2.1.1 Common sense definitions of fluency

Definitions of fluency in English language teaching and applied linguistics all share at least one thing in common; they are essentially based on the common sense non-technical meaning of the word. Therefore, I would like to start by looking at dictionary definitions, which can create a picture of the common uses of the term *fluency*, as used by non-linguists. Looking at the adjective *fluent*[1], the Merriam-Webster Online Dictionary gives two relevant definitions ("fluent," 2008).

> 1 (a) capable of flowing**:** fluid (b) capable of moving with ease and grace <the *fluent* body of a dancer>
>
> 2 (a) capable of using a language easily and accurately <*fluent* in Spanish> <a *fluent* writer> (b) effortlessly smooth and flowing**:** polished <a *fluent* performance> <spoke in *fluent* English> (c) having or showing mastery of a subject or skill <*fluent* in mathematics>

---

[1] In this dictionary, *fluency* was simply defined in terms of *fluent*, which is taken to be the central part of speech.

The first sense, "capable of using a language easily and accurately", implies second language use in the example, but the second, "effortlessly smooth and flowing", is also relevant to first language use, as well as other non-linguistic skills. There is also a more literal use of the word, defined as "capable of flowing" and "capable of moving with ease and grace". These uses are evidence of a conceptual metaphor of smooth motion and flowing as a liquid. This metaphor is applied to language in the definitions above, where we get the idea of *flow* in language, that is, language used in a way that is smooth and flowing. It is interesting that definition 2a, regarding second language speakers, uses the word *easily* to describe fluency, and definition 2b, regarding first language use, uses the word *effortlessly*. Both of these imply that fluency is the ability to speak a language in a way that involves little or no effort on the part of the speaker, or another interpretation would be that the language is spoken in such a way that there is no *perceived* effort on the part of listeners.

The second language use of *fluency,* definition 2a, also includes the idea that the language is used accurately. This is the basis of the use of *fluency* as overall language proficiency. This is interesting in the lay definition, because in English language teaching, these terms have been put in contrast to one another. Although the dictionary definitions do not provide a comprehensive definition of the concept of fluency, they do provide a starting point for examining it. The common sense meaning of *fluency* implies that normal language is smooth and flowing. Koponen and Riggenbach (2000) note that many languages have similar metaphors of motion and flowing currents. In Japanese, the relevant word is *ryucho* (流暢) which has as one of its two characters, the same character that is used for the word that has the literal meaning of "flow".

Although the lay definition of fluency equates it with "flow" in language, the picture that emerges from looking at natural spoken language is quite different. Goldman-Eisler (1968, p. 16; cited in Wingate, 1987) observed that:

> Somehow the phenomenon of speech has become associated with images which suggest continuity of speech production. We speak of even flow [...] and many words relating to speech derive from descriptions of water in motion. [...] The facts however show these images to be illusory. If we measure vocal continuity by the number of words uttered between two pauses and call "phrase" the sequence uttered without a break we obtain a picture of fragmentation rather than continuity.

Goldman-Eisler (1968) in her studies of spoken language also noted that "spontaneous speech is a highly fragmented and discontinuous activity" and "When, even at its most fluent, two thirds of spoken language comes in chunks of less than six words, the attributes of flow and fluency must be judged an illusion." (quoted in Guillot, 1999, p. 3). Furthermore, Wingate (1987), working in the field of speech pathology, describes normal native speaker speech as dysfluent, saying that it is regularly misperceived as fluent. Koponen and Riggenbach also note that "oral proficiency and its perception are so complex in linguistic description that considering them unidimensionally as flows or currents constitutes an overly simplified view of language and speech" (2000, p. 13). The question that needs to be answered therefore, is what creates the impression or perception of fluidity or flow in spoken language in spite of the observed realities of spoken language.

Clearly the common sense notions of *fluency* are not precise enough for applications either in psycholinguistics, second language acquisition research, or language teaching and testing. In this chapter, I would like to move toward a clearer understanding of what fluency is from a theoretical perspective, in order to provide a definite, realistic, and useful concept for linguistics and second language acquisition research, as well as pedagogical applications.

2.1.2 Definitions in linguistics and SLA research

One of the first important papers to attempt to define fluency was Fillmore's (1979, 2000). His discussion of fluency includes four senses of the term, which are variable properties of first language users, i.e., they are characteristics of the "maximally gifted wielder of language" (Fillmore, 2000, p. 52), and not necessarily qualities inherent in being a native speaker. The first sense is the ability to speak at length, at speed, and with few pauses. In other words, "they don't have to stop many times to think of what to say or how to phrase it" (Fillmore, 2000, p. 51). The second sense of fluency is the ability to speak in coherent, semantically dense sentences. He implies that this is the ability to plan speech well, and gives the example of Noam Chomsky and William F. Buckley[2]. Fillmore's third sense is the ability to have appropriate things to say in a wide variety of situations, i.e., never being at a loss for words. This sense implies more than just "flow" in speech; it also implies lexical and pragmatic knowledge. The fourth sense is the ability to be creative, witty, and imaginative, adding that people with this ability appear to be doing "rapid pre-editing" (Fillmore, 2000, p. 51) of their language before they speak. This relates to a higher level of conversational ability, a kind of cultural fluency, where the speaker should try to be funny and interesting. In a second language, this is one of the most difficult aspects of language to master, but it is certainly an important part of casual conversation. However, regarding nonnative speaker

[2] William F. Buckley was a conservative television commentator, famous for his large vocabulary and for speaking in a very dense manner that resembled writing. His speaking was not, however, particularly rapid, in contrast to Fillmore's first sense of fluency, which is sometimes referred to as *DJ fluency.*

speech, “rapid pre-editing” resembles monitoring, which is generally considered to hinder fluency. Brumfit (1984) notes that this last and the second of Fillmore’s senses of fluency could also be attributed to people who are not particularly fluent, suggesting that these might be better associated with personality. Kormos and Dénes (2004) also comment that Fillmore’s definition of fluency is hard to separate from overall oral proficiency. A clearer concept of fluency is particularly necessary for second language acquisition, where fluency is typical construed as something that learners may not have and want to achieve.

Although in Fillmore’s essay fluency is a collection of variable qualities of native speaker (NS) speech, in general, the term is used much more with nonnative speakers (NNSs) (Koponen & Riggenbach, 2000). For NNSs, fluency is an evaluative criterion, as testing rubrics, where NNSs are described as being nonfluent, or more or less fluent, typically on a scale that ends with “nativelike” abilities.

Second language testing is also an area where fluency figures prominently as the name of a criterion upon which second language learners are evaluated. Koponen and Riggenbach (2000), looking at the 4-level fluency bands of the Test of English for Educational Purposes (TEEP, used at University of Reading Centre for Applied Language Studies), note that the terms used to describe aspects of fluency within each band, such as *continuity*, *evenness*, *halting*, *groping for words*, and *fragmentary*, are very unspecific and depend heavily on the rater’s interpretation. Furthermore, the band descriptors collapse different features of language such as hesitations, coherence, speed, and reformulation into the category of fluency, and they note that it could be argued that the definition of fluency changes from one band to the next. Similar observations were made by North (1993) regarding the ACTFL oral proficiency interview fluency bands.

Freed (1995) also comments that fluency is often not defined at all for raters, which leaves raters to apply their own holistic conception of fluency. Esser (1996; cited in Hasselgreen, 2005) also found that judges had very little inter-rater reliability with regard to fluency, and that each judge had their own personal conception, which they were often unable to define clearly, making fluency an intuitive holistic judgment.

In the area of research, different scholars have used quite different definitions of fluency, falling into several categories (Koponen & Riggenbach, 2000). One view is to equate fluency with a range of temporal phenomena, such as rate of speech or lack of unfilled pauses. Another is to equate fluency with proficiency in the second language. A third is a psycholinguistically-oriented definition, where fluency is the automatization of the underlying cognitive processes of language production. A fourth is a listener-oriented definition where fluency implies the ease of the listener's ability to comprehend the message. This sense is psycholinguistic in that it is related to the cognitive load placed on the listener to process the message, but some researchers also relate it to textual coherence and the ability to participate smoothly in conversation (Olynyk, Sankoff, & d'Anglejan, 1983). Finally, there is a pedagogical notion of fluency, which is contrasted with accuracy with regard to classroom activities and tasks (Brumfit, 1984).

With regard to fluency as a range of temporal factors, Riggenbach (1991) and Lennon (1990) have both noted that rater's judgements of fluency rarely depend only on temporal factors, but also seem to be influenced by other aspects of language such as lexical and syntactic complexity. On the other hand, equating fluency with proficiency leaves us with a redundant concept in second language acquisition research, and is a problem for learner English, when a learner is described with words such as "Talking to him is like talking to a native speaker, but if you listen really carefully to what he says, his grammar is all over the place[3].” One reason that it is so difficult to separate fluency from proficiency is that impressions of fluency on listeners or raters depend on many other aspects of language, and is it nearly impossible to control other variables in experiments. For example, if fluency depends on vocabulary knowledge (more vocabulary knowledge means less pausing to search for words or to rephrase), then an increase in vocabulary knowledge would likely, but not necessarily, result in both an increase in fluency and overall proficiency. In any case, the definition of fluency is crucial in that it must include all factors or components that are important to encompass the concept of smoothness and flow in language, without being so general as to be equivalent to overall proficiency.

Lennon (2000) addresses this issue by proposing two senses of fluency: a broad sense and a narrow sense. The broad sense is a higher-order fluency that is roughly equivalent to overall oral proficiency. The narrow sense is a lower-order fluency that essentially refers to speed and smoothness of speech. His claim is that the higher-order fluency is dependent on the lower-order fluency, but the lower-order fluency corresponds to aspects of spoken language that can be measured objectively through temporal factors. In second language acquisition, Lennon’s higher-order fluency would be better replaced with *proficiency*, as it is redundant as discussed above.

Lennon’s (2000) definition of the narrow sense of fluency is one of the most often cited, that is, “the rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing” (p. 26). This definition of fluency is applicable equally to native

[3] Paraphrased from a comment by a colleague while discussing a student we both knew.

speakers or nonnative speakers, and attempts to capture both fluency as automaticity of psycholinguistic processing and as a range of temporal factors. The temporal factors, such as rate of speech and mean length of run, can be seen then as outward manifestations or observable epiphenomena resulting from the automaticity of the underlying processing of language. Schmidt (1992) similarly associates fluency with processing in real time, saying that it is an automatic procedural skill. Lennon (1990) further notes that “fluency is an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently” (p. 391). Here Lennon is referring to *perceived fluency,* i.e. fluency from the point of view of the listener, while in the definition above, fluency is assumed to be internal to the speaker.

Segalowitz (2010) clarifies these definitions by proposing three senses of fluency. The first is *utterance fluency*, which includes the measurable, temporal aspects of speech, such as speech rate, length of run, and lengths of filled and unfilled pauses. The second sense is *cognitive fluency*, which is the efficient and automatic operation of the language production process. Utterance fluency clearly depends on and is limited by cognitive fluency, but many other factors may affect the temporal aspects of speech. For example, a speaker may speak more slowly to accommodate a weaker speaker, or a speaker may be distracted by something around them. The third sense is *perceived fluency*, which is, following Lennon, the perception by the listener that the speaker’s speech processes are smooth and efficient. Utterance fluency must influence perceived fluency, but there may be other factors that affect it as well.

An important aspect of defining fluency is to avoid equivocation between different senses of the word, and to explain what the relationship is between these senses. Segalowitz highlights three of these senses: the observable temporal aspects of speech, the rapid and efficient cognitive processing of speech, and the impression on the listener that speech processing is rapid and efficient. Dysfluencies make the listener aware of problems or slowdowns in the speaker's language processing (Lennon, 1990). This leaves the possibility of a mismatch between fluency from the speaker's point of view and the listener's point of view. For example, in the case of the strategy of *avoidance*, the speaker may have difficulty expressing something (i.e., inability to translate communicative intention into language), but by simply not saying it or saying something else, they may appear more fluent to the listener, while feeling a lack of their own ability. On the other hand, speakers could feel comfortable with their speech without realizing that their speech has many dysfluencies apparent to the listener (Lennon, 2000). Segalowitz (2007) also includes a wider range of language skills, explaining that fluency includes speaking or reading at an acceptable rate, comprehending spoken or written language at speed, and being able to perform in a variety of social and physical situations.

Hedge (1993), who defines *fluency* as "the ability to link units of speech together with facility and without strain or inappropriate slowness or undue hesitation" (p. 275), claims that fluency increases when learners use fillers and paraphrases rather than silent pauses. Koponen and Riggenbach (2000) note that this might be considered a kind of compensatory fluency, that is, having techniques to compensate for problems in the learner's knowledge of language. This brings up an interesting point which will be discussed in more detail below, that of whether it is possible for a learner to appear fluent to listeners, but to lack essential features of fluency from the speaker's point of view.

Ejzenberg (2000) notes that fluency should be seen from both a speaker-based, psycholinguistic perspective, and a listener-based, sociolinguistic perspective. She claims that fluency is a "display" or "manifestation", defined as "the perception of the ease with which a speaker delivers the message, making it appear to be smooth and naturally paced to the listener" (p. 287), which is variable and dependent on the context or situation. Fluency is ultimately the speaker's ability to project an "image of fluency" (p. 288) to the listener. Ejzenberg (2000), noting the relationship of fluency to overall proficiency, states that it is one component of oral proficiency, dependent on language knowledge, i.e., grammatical competence and discourse competence, but accomplished mainly by strategic competence, which is defined as ability to deal with breakdowns in communication (Canale, 1983) .

Sajavaara (1987) takes a broader view of fluency including both the linguistic acceptability and smooth continuity of speech, stating that fluency "equals the communicative acceptability of the speech act, or 'communicative fit', and expectations concerning this fit vary according to the situation" (p. 62). Just as native speaker speech contains pauses, hesitations, false starts, and rephrases, it is expected that nonnative speaker speech should also contain these "dysfluencies" (Lehtonen, Sajavaara, & May, 1977), and therefore, fluency must be evaluated in context (Lehtonen, 1978).

> Learning to speak fluently does not always imply an uninterrupted flow of speech that is sequentially and grammatically irreproachable. The "good" speaker "knows" how to hesitate, how to be silent, how to self-correct, how to interrupt, and how to complete expressions or leave them unfinished. Speech must meet the expectations of the speech community and represent normal, acceptable and relaxed behavior (Sajavaara, 1987, p. 62).

This implies that fluency is not only dependent on the speaker, but also on (1) the listener, (2) the context or situation, and (3) the speaker's ability to predict how the listener will interpret the speaker's utterances in context. Sajavaara describes participants in conversation as problem solvers, where the problem is to communicate using previous experience and whatever practical means are available, including facial expressions and gestures. Listeners and speakers both construct interpretations by taking the other person and their own discourse history into account.

Hasselgreen (2005) also takes a listener-oriented and interactive perspective on fluency when she defines it as "the ability to contribute to what a listener, proficient in the language, would normally perceive as coherent speech, which can be understood without undue strain, and is carried out at a comfortable pace, not being disjointed or disrupted by excessive hesitation" (Hasselgreen, 2005, p. 134). This definition includes temporal aspects (“comfortable pace” and “excessive hesitation”) as well as the overall sense of perceived fluency as being ease of comprehension. The psycholinguistic side of fluency is hinted at as the “ability” to produce speech that is perceived as fluent. The problem is in the details. What exactly is “normal”? To answer this question, it is necessary to look more closely at the observable components of fluency. This will be dealt with in the next section.

Another influential view of fluency is that of Brumfit (1984, 2000). Brumfit contrasts fluency with accuracy, but, unlike the previous definitions, this is intended to be a pedagogical distinction rather than a theoretical one, initially designed to guide teaching methodology. Brumfit therefore applies this to the distinction between classroom activities that promote fluency, and activities that promote accuracy. In this context, fluency is defined as "natural language use, whether or not it results in native-speaker-like comprehension or production" (Brumfit, 1984, p. 56) and "the maximally effective operation of the language system so far acquired by the student" (Brumfit, 1984, p. 57). Under this view, all learners have fluency. It is hindered by attention to accuracy, that is, the speaker excessively monitoring their own speech. Fluency activities, in Brumfit’s view, result in the learner’s maximally effective operation of the acquired language system by minimizing dysfluencies caused by the learner monitoring their own output. Brumfit acknowledges that even native speakers monitor their speech, therefore, for nonnative speakers, the issue is with hesitations caused by conscious application of memorized rules or other mnemonic-like devices while speaking. Ejzenberg (2000) also notes that conscious monitoring in nonnative speech results in lower perceived fluency, and it is clear that thinking about declarative grammar knowledge can be a reason for a hesitation.

Guillot (1999), looking at language learners' own informal definitions of fluency, notes that many people see fluency as including accuracy. All of the 24 accounts in her study mention productive ability, but about 30% also refer to listening skills. Ten out of 24 (42%) mention absence of hesitation, and 14 of 24 (58%) include ease or facility with the language. Approximately one-third of the learners included accuracy in the definition. Although Guillot makes no mention of it, two of the six examples that she gives refer to "thinking in the language" and not having to translate from the L1.

As can be seen from the range of definitions surveyed above, there is little agreement as to exactly what fluency is, although the common sense notions of smoothness and ease of speaking and/or listening are found throughout the definitions. It appears that most, if not all, of these definitions are intuitive definitions of the researchers or educators. What is needed is a research-based definition, looking at the basis for intuitive impressions of smoothness and ease, and then deriving a definition from those results. The next several sections of this chapter will move towards this goal.

2.2 Fluency as a temporal phenomenon

Fluency research in the 1980's and 1990's particularly focused on identifying quantifiable temporal variables which could distinguish fluent and nonfluent speech.

The type of factors considered include length and duration of unfilled (silent) and filled non-lexical (*um* and *uh*) pauses or hesitations, length of uninterrupted runs, rate of speech, rate of articulation, amount of speech, phonation-time ratio, and presence of other dysfluencies such as uncompleted utterances, lengthened sounds (drawls), repetitions, and restarts.

One of the difficulties with comparing fluency studies is that they often use very different methodologies. Some compare learners classified as fluent with learners classified as nonfluent (Ejzenberg, 2000; Riggenbach, 1991; van Gelderen, 1994). Others are longitudinal studies, usually comparing learners before and after a period of study abroad (Freed, 1995; Lennon, 1990; Rossiter, 2009; Towell et al., 1996). Although some studies look at fluency in interaction (Ejzenberg, 2000; Morales-Lopez, 2000; Riggenbach, 1991), most look at fluency in monologue production through speaking tasks (for example, Kormos & Dénes, 2004; Lennon, 1990; Rossiter, 2009). Also, different topics and tasks can cause differences in measurable aspects of fluency (Goldman-Eisler, 1968; Skehan, 1998).

Quantitative studies of fluency have generally found that more fluent L2 speakers have comparatively fewer pauses than less fluent speakers, a higher rate of speech, and longer uninterrupted runs of speech (Ejzenberg, 2000; Freed, 1995; Hasselgreen, 2005; Lennon, 1990; Riggenbach, 1991). Riggenbach (1991) found that students judged as fluent (rather than nonfluent) by their teachers tended to have fewer unfilled pauses and higher rates of speech. Deschamps (1980), comparing speakers in both their L1 and L2, found that in the L2, speakers had a lower rate of speech, and more frequent pauses, but not longer pauses. Freed (1995) and Lennon (1990) found that more fluent speakers had higher rates of speech, fewer unfilled and non-lexical filled pauses, fewer repetitions, and increased mean length of run. Cucchiarini, Strik, and Boves (2002) also found that speech rate, phonation-time ratio, duration of unfilled pauses, and mean length of run were significant factors in fluency judgments. Towell, Hawkins, and Bazergui (1996) found that improvements in learners' fluency after a period of study abroad were due to longer runs of speech, but not shorter pauses. Higher rates of speech were also observed, but this was claimed to be due to the increased lengths of run. Ejzenberg (2000) also found higher rates of speech to distinguish fluent and nonfluent L2 speakers as rated by trained raters listening to recordings of the speakers performing several tasks. However, she also notes that the subjects with the highest fluency ratings were not always the ones with the highest speech rates. Hasselgreen (2005) also confirmed that more fluent speakers had higher rates of speech, longer runs, and fewer pauses within clauses. Kormos and Dénes (2004) found that fluency correlates with speech rate, mean length of run, phonation time ratio, and *pace,* which is defined as the number of stressed words per minute (Vanderplank, 1993). They claim that pauses and other hesitation phenomena are not important factors in fluency judgments. However, Rossiter (2009) found that fluency judgments correlated with pause duration and pruned speech rate[4]. Riggenbach (1991) found that most fluent speakers resemble each other, but that nonfluent speakers can differ in many ways, implying that fluency may be a bundle of related features, where a deficiency in any one of those features could result in problems with fluency. While some temporal features tend to be correlated in many studies, there is much variability, showing that it is difficult to pin down particular temporal features as defining fluent performance.

[4] Defined as the number of syllables after removing self-corrections, repetitions, false starts, and filled non-lexical pauses (Rossiter, 2009).

2.2.1 Rate of speech

Although fluency judgments of learners generally correlate with higher speech rate, Munro and Derwing (2001), based on research by Anderson-Hsieh and Koehler (1988), found that speech rate had a curvilinear relationship with fluency judgments, i.e., there is a best range for speech that is easy to comprehend, with speech that is too fast or too slow being perceived as less fluent. Furthermore, native speakers' preferred speech rate for nonnative speakers was found to be lower than for other native speakers. This could be due to native speakers having more difficulty processing nonnative pronunciation or non-nativelike phraseology. Also, speech rate (as measured in syllables, rather than words, per minute) has been shown to vary based on genre and context, for example, with speech rates in conversation being higher than those in lectures (Tauroza & Allison, 1990).

2.2.2 Unfilled pauses

Chafe (1980a) notes that there are many reasons for hesitations in speech, with some of them being dysfluencies but others being a normal part of fluent speech. Pauses are first of all necessary, because speakers must sometimes stop in order to breathe. Furthermore, pauses between clauses or *intonation units* (Chafe, 1994) are necessary for planning the next utterance (Dechert, 1980; Goldman-Eisler, 1968; Pawley & Syder, 2000), and therefore are not necessarily dysfluencies (Adolphs & Dahlmann, 2007). Pauses can be further divided into planning *what* to say and planning *how* to say it.

It is also claimed that the location of the pauses is critical as well, with pauses that are within clauses being more dysfluent than those between clauses. Deese (1980) found that pauses that were within clauses are actually perceived as being longer than pauses between clauses. Pawley and Syder (2000) found that pauses before first clauses of multi-clause constructions were 0.5 to 2.0 seconds in duration, pauses before clauses that were not first were 0.2 to 1.0 seconds, and pauses for word search were about 0.2 seconds. Chafe (1980a) claims that pauses within clauses are for grammatical or lexical choice rather than planning content, which would explain why they are more frequent with L2 users. Furthermore, less fluent L2 speakers tend to have more clusters of dysfluencies occurring together, which can result in the impression of "choppy" sounding speech (Riggenbach, 1991). In researching L2 speech, it is difficult to determine when a pause is a normal pause for planning purposes, and when it is a sign of the language system "breaking down" (Lennon, 1990, p. 394). Furthermore, some hesitations are a normal part of the pragmatics of certain language functions, such as in refusing invitations (Sacks, Schegloff, & Jefferson, 1974).

Raupach (1980), comparing speakers in L1 and L2, found that a speaker's L1 pause profile transferred to the L2, but that in the L2, there were more pauses internal to syntactic clauses. Also, he claims that hesitation devices operate together and do not need to be treated separately, although others (Chambers, 1997; Riggenbach, 1991) imply that hesitation devices need to be categorized and studied separately to determine which are perceived as fluent and dysfluent.

Searching for words is one of the major reasons for hesitations (Chafe, 1980a), and Fillmore (1979, 2000) suggests that vocabulary size can be one measure of fluency, but that it is situation-dependent in that it depends on whether the other speaker can understand the vocabulary used. Since adult L2 speakers usually have a large mismatch between what they can say in the L1 and what they can say in the L2, they are likely to find themselves in situations where they simply lack knowledge of the needed lexical item, and have to resort to circumlocution, avoidance, reliance on help from the interlocutor, or silences.

There is also evidence that hesitation phenomena may not just be necessary from the point of view of the speaker's ability to process language, but may also be helpful in comprehension. Corley, MacGregor, and Donaldson (2007) show that hesitations (*uh*, in the case of their study) which preceded words that had low predictability as completions of the given sentences, were more easily integrated into context, and resulted in greater retention of the following word. Their study implies that hesitations before unpredictable words reduce the cognitive processing required for comprehension by the listener. This implies that hesitations can sometimes be fluent rather that dysfluent. The listener can benefit from the temporal delay in speech regardless of whether the hesitation is a filled pause (*um*) or a silence (Corley & Hartsuiker, 2011).

2.2.3 Other hesitation phenomena

Restarts or false starts are generally considered to be dysfluencies, although they are often found in native speaker speech as well. Carroll (2004), studying conversations of groups of three novice learners of English, observed that some apparent dysfluencies could be argued to contribute to the overall fluency of the conversation rather than detract from it. By viewing video of the conversations, he found that some restarts were the result of the speaker holding their turn while waiting for full eye contact from the person the utterance was directed to. Once eye contact was made, signaling to the speaker that the recipient was paying attention, the utterance was completed. Carroll argues that this behavior is not dysfluent, but rather shows interactional competence. This is supported by Goodwin's (1979) finding that speakers can use a pause or restart in the middle of a phrase to request the gaze of a participant in a conversation.

The "fluent" speech of native speakers is full of these apparent dysfluencies, which native speakers are usually themselves unaware of (Corley & Stewart, 2008;

Wingate, 1987). The difficulty in determining whether temporal phenomena are dysfluent or not lies in the fact that they are the symptoms of interactive aspects of the conversation and the underlying psycholinguistic processes of speech production. On one hand, the “dysfluencies” of spoken language are a window into the limitations of human language production for all speakers and into the difficulties of second language production for learners, but they can also be viewed simply as the normal techniques used to facilitate the production of spoken language in real time. In this thesis, I will use the term *speech management phenomena* (Allwood, Nivre, & Ahlsén, 1990), rather than *hesitation phenomena* or *dysfluencies* to refer to phenomena such as repetition, restarts, and filled pauses, since, as noted by Rühlemann (2007), there is no “eufluency”, i.e., perfect fluent speech with no hesitation phenomena.

There is a danger in attributing anything that fluent speakers do to fluency because this equivocates fluency with overall proficiency, and skips over the most important issue of what is behind the observable features. Any explanation of fluency must be grounded in a theoretical framework of how fluency works, so that the correlations are testing the theory rather than simply correlations. For example, an increase in speech rate may correlate well with levels of fluency, but this does not explain what enables the increase in speech rate. Furthermore, language teachers cannot simply teach learners to increase their speech rate to be more fluent.

In summary, temporal factors such as length of run, speech rate, and location and length of pauses, are evidence of fluency in speech, but lists of these factors do not always hold up to experimental testing, and they only address one kind of fluency—utterance fluency. The surface features of language are always ambiguous (Lennon, 2000). What is needed is a clearer theoretical concept of fluency. The next section will begin to address that by looking at the psycholinguistic basis of fluency.

2.3 Fluency as automaticity

2.3.1 Psycholinguistic basis of fluency

The temporal factors of fluency discussed above can be seen as epiphenomena of the efficient processing of language in the brain. Fluency can also be associated with procedural knowledge as opposed to declarative knowledge (Færch & Kasper, 1984; Schmidt, 1992). "Whereas such elements as idiomaticness, appropriateness, lexical range, and syntactic complexity can all be assigned to linguistic knowledge, fluency is purely a performance phenomenon; there is (presumably) no fluency store" (Lennon, 1990, p. 391). Linguistic knowledge is what is stored in memory, but cognitive fluency is the speaker's ability to access those memories and process them into speech within the constraints of real time conversation. For learners, the problem is not simply to increase the speed of their speech, but rather to improve the efficiency of the organization and production processes of language in the mind, which then results in more fluent speech.

It has long been recognized that planning speech requires time away from articulation (Goldman-Eisler, 1968), and that many apparent dysfluencies are a necessary part of natural spoken language. Chafe (1980b) and Pawley and Syder (2000), both looking at native speaker conversation, argue that there is an upper limit on what speakers are able to plan for a single uninterrupted utterance. Chafe (1992) proposes that this limit is based on the capacity of working memory (Baddeley, 1992; Miller, 1956; Schweickert & Boruff, 1986) and that speakers can at most plan the lexical content of one new idea, or one focus of consciousness, at a time, which often coincides with a single clause. Pawley and Syder (2000) also show that chaining single clauses—that is, producing one clause at a time connected serially—results in more fluent speech in native speakers than more complex grammar involving embedded clauses. That working memory capacity limits what can be planned in a single fluent utterance is also supported by research correlating temporal measures of L2 fluency with working memory (Mota, 1997, 2003) and with phonological memory (O'Brien, Segalowitz, Freed, & Collentine, 2007).

2.3.2 Automaticity

Within the limits of working memory, fluency is achieved through automaticity of language production (Schmidt, 1992), although the term *automaticity* has been used in four different ways in the literature: (1) speed of processing, (2) availability of relevant procedural memory, (3) absence of attentional control, and (4) ballistic (unstoppable) processing (Dörnyei, 2009, p. 287). Dörnyei notes that speed may not be an important component of a psycholinguistic definition of automaticity, because it does not entail a restructuring of the underlying language knowledge. It is generally accepted that fluent language production is not based on the rapid application of explicit rules. Rather, fluent language production is based on procedural knowledge within a theory such as Anderson's ACT-R (Anderson et al., 2004), or based on strengthening of connections to relevant memories of linguistic items within connectionist theories such as Logan's Instance Theory (Dörnyei, 2009; Logan, 1988).

Automatic processing frees working memory capacity for other aspects of speaking, such as planning content. Studying the neurobiological correlates of fluency, Dewaele (2002) states that fluency in both L1 and L2 depends on the existence, accessibility, and retrieval of procedural knowledge from long-term memory. Fluent speech results when the appropriate or required language is present and accessible in long-term memory, it can be combined efficiently, and there is sufficient storage in working memory. Problems in any of these areas, such as lack of language knowledge, an individual's lower capacity of working memory, or anxiety in the communicative situation, could lead to problems in fluency. Although it may be possible for fast operation of declarative knowledge to improve fluency to some extent, in general L2 speakers' reliance on declarative knowledge to speak results in reduced fluency due to overloading working memory (Dewaele, 2002; Paradis, 1994).

Along with automaticity, the efficient storage of language in memory is another factor that contributes to cognitive fluency (Segalowitz, 2010). Exactly how language is organized in the brain is still unknown, but it is known that formulaic sequences, explained in the following section, play an important role.

2.3.3 Formulaic sequences

The link between fluency and the use of formulaic language has been recognized for quite some time (Pawley & Syder, 1983; Raupach, 1984). In a study of learners' fluency development after a period abroad, Towell et al. (1996) concluded that the observed temporal aspects of fluent speech, i.e., higher rate of speech and longer uninterrupted runs, were a result of more efficient proceduralized knowledge. This could be due to automatization of the processes of speech production or the use of ready-made chunks of language, i.e., formulaic sequences. Larson-Freeman and Cameron (2008a) note that, although native speakers can theoretically produce infinitely many utterances that conform to the syntactic rules of the language, they rarely do, and rather tend to mix and match frequently used chunks. This is also one of the most important findings of corpus linguistics—that a large part of native speaker speech consists of recurrent chunks (Nattinger & DeCarrico, 1992; O'Keefe, McCarthy, & Carter, 2007; Schmitt & Carter, 2004). Formulaic sequences also contribute to longer runs of speech (Rossiter, 2009; Wood, 2006), assuming they are processed and uttered as single units.

Wray (2002, p. 9) defines a *formulaic sequence* as “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored or retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar”. By directly accessing the formulaic sequence, the speaker is able to circumvent application of syntactic rules in the construction of utterances in real time. Connectionist neural network theories propose that as particular sequences of language are repeatedly accessed in memory, their neural connections are strengthened, which leads to faster access later (Gasser, 1990; MacKay, 1982; Strayer & Kramer, 1990).

2.3.4 Formulaic sequences in learner language

It has been well established in corpus-based research that native speakers use formulaic sequences (Nattinger & DeCarrico, 1992; Schmitt & Carter, 2004; Wray, 2002) as a large part of their produced language. Their use in nonnative speech has been investigated less so, probably due to the lack of corpora, especially spoken corpora, of learner language. Furthermore, although analysis of native speaker corpora can show sociolinguistic patterns across a large group of speakers, determining how to place nonnative speakers into groups that are similar enough to be fairly compared is difficult due to differences in first language, proficiency, and individual differences in language development. However, there have been smaller-scale studies in the past. Raupach (1984) found that learners who had returned from study abroad and were perceived as more fluent, tended to use more formulaic language, particularly the kinds of word and phrases that have been called *smallwords* (Hasselgreen, 2005). Dechert (1983) proposed that formulaic sequences act as *islands of reliability* for learners in language production. That is, they ease the processing load between stretches of language that have to be constructed creatively or translated from L1.

Another study that looks at the relationship between the use of formulaic sequences and fluency is Oppenheim's (2000) study of advanced nonnative speakers of English performing short speeches. She found in her study that fluent nonnative speakers relied heavily on the use of recurrent sequences[5], and conversely, that the use of recurrent sequences is the basis of fluent speech. In her study, nonnative speakers were recorded performing short speeches. They performed speeches on two topics, with each speech performed twice (with the second speech within 30 minutes of the first). She found that all the speakers used recurrent sequences, calculated at between 48 and 80 percent of the speech. She noted that the percentage of sequences varied more between topics than retellings.

It is also possible for an L2 speaker to have idiosyncratic formulaic sequences which are stored and accessed whole in memory, but which are not "nativelike", and would not be a recurrent chunk found in a corpus of native speaker language (Oppenheim, 2000; Wood, 2006). Their use would promote "nativelike fluency", but would not be "nativelike selection" (Pawley & Syder, 1983). However, the use of idiosyncratic rather than nativelike formulaic sequences could result in greater processing required by native speaker listeners (Millar, 2011), which could result in the perception of weaker fluency from the listener's point of view.

---

[5] I use Oppenheim's term here, rather than the term *formulaic sequences*, which is preferred by many writers after Wray (2002), because her term *recurrent sequences* reflects her methodology, which was to look for repeated occurrences of phrases between two repetitions of the same prepared speech. The term *formulaic sequence* as used by Wray (2002) implies a cognitive process where the sequence is accessed whole in memory.

It could be that the use of formulaic language, either non-productive (fixed) sequences (Nattinger & DeCarrico, 1992) or sequences with slots that can be easily manipulated, results in fewer pauses within the sequences, and therefore results in longer runs of speech and a faster rate of speech (Chambers, 1997). Studies of pauses within recurrent sequences, such as Erman (2007) and Adolphs and Dahlmann (2007), give further evidence of the direct access and storage of formulaic sequences, implying that formulaic sequences are usually spoken as uninterrupted runs, which as explained above, contributes to temporal aspects of fluency.

2.4 Conclusion

In this chapter, I have discussed the problems with expressing fluency only in terms of temporal variables, such as speech rate or mean length of run. Even though many studies have shown that these variables correlate with raters' judgments of fluency, the results are not very consistent. Furthermore, many studies simply compare "fluent" speakers and "nonfluent" speakers, without distinguishing between different levels (Hasselgreen, 2005). Also, for learners, there is not a linear development in temporal features from, for example, extremely long pauses to no pauses (Fulcher, 1996). Also, there are many reasons for the presence of apparent dysfluencies in speech, with some being evidence of lack of proficiency in learners' speech, but others being normal or unavoidable speech management phenomena.

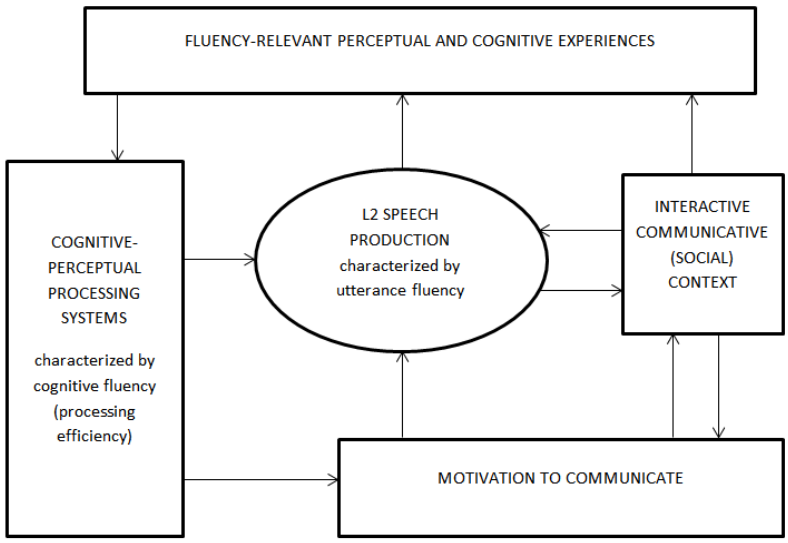

Figure 2.1 Framework for thinking about the dynamic relationships between sources influencing L2 fluency (Segalowitz, 2010, p. 164).

One recent way of looking at fluency, particularly utterance fluency and cognitive fluency, is Segalowitz's (2010) framework. This framework (Figure 2.1) is an attempt to show the dynamic relationship among factors that influence L2 fluency. In this framework, L2 speech production, which is characterized by utterance fluency, is influenced by the speaker's motivation to communicate and the social context of the communicative interaction, which also influence each other. Cognitive fluency is based on the language processing system, and influences speech production. The box labelled "Cognitive-Perceptual Processing Systems" (Figure 2.1) contains the language production system, for which one of the most current models is the Levelt model, adapted by DeBot, and later again by Levelt, and finally by Kormos (Figure 2.2). Last, there are fluency-relevant experiences, which, through exposure to language and chances for repetition, can build automaticity in the language production system.

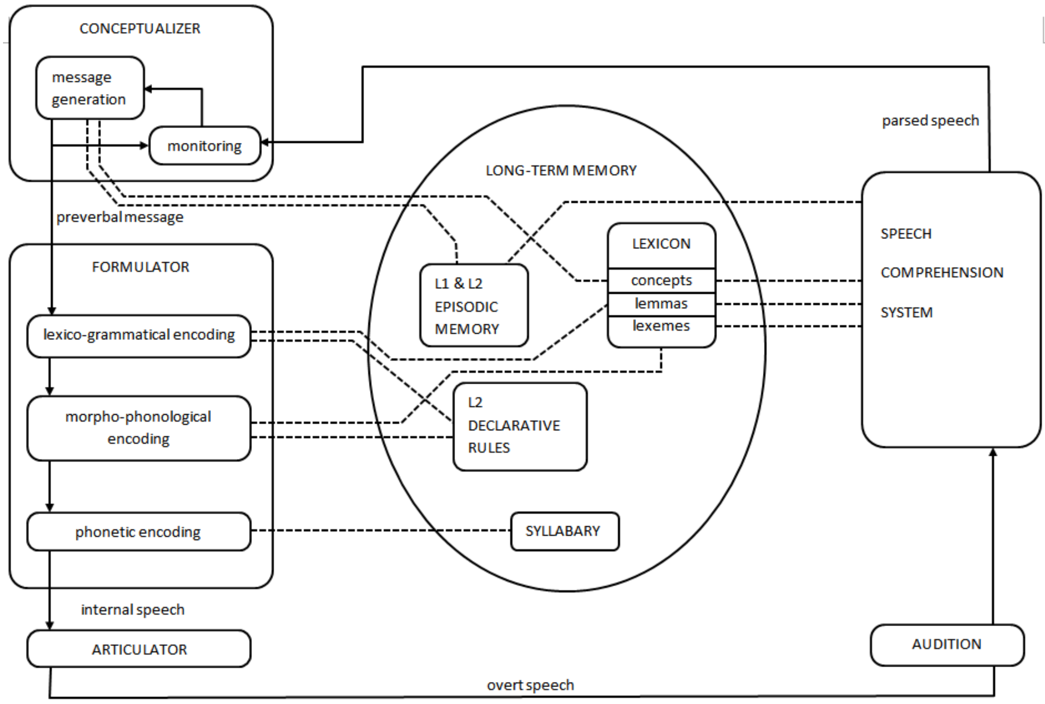

Figure 2.2 The model of bilingual speech production of Kormos, based on DeBot's adaption of Levelt's modular language production model (Kormos & Dénes, 2004, p. 168).

This chapter has looked at what I term *monologic* aspects of fluency. This includes features of utterance fluency, i.e., temporal variables of speech such as speech rate, articulation rate, length of uninterrupted run, and the distribution of silences. It also includes what have traditionally been called *hesitation phenomena*, such as the distribution of filled pauses (as well as unfilled pauses), restarts, repetition, and false starts, although an increasing amount of research reviewed in this chapter challenges their status as "dysfluencies", and points toward some of these features as contributing to fluency rather than necessarily detracting from it. An alternative term for these phenomena is *speech management*, which highlights how these phenomena are natural means for speakers to deal with planning and producing language under the time constraints imposed by conversation in real time.

This chapter also looked at features of cognitive fluency, that is, automaticity in language production and the use of formulaic sequences. Cognitive fluency is enabled by efficient processes of the language production system, and underlies utterance fluency. The language production processes are made more efficient through the use of formulaic sequences, which are pre-fabricated chunks of language, and through the automaticity in the retrieval and use of language elements.

The next chapter will complement these features of monologic fluency by looking at features of what I term *dialogic fluency*. In particular, I will look at interactive aspects of fluency, including *perceived fluency* and *confluence*, with the goal of determining how these factors can fit into a framework for defining and researching fluency.

3 Interactive elements of fluency

3.1 Other components of fluency

In Chapter 2, I reviewed the previous literature on the definitions of fluency, and on the components of utterance and cognitive fluency, that is, fluency from the point of view of automatic and efficient speech production on the part of the speaker. In this chapter, I would like to look at other components, particularly those factors that are related to more interactive aspects of fluency, including the perception of the fluency of a speaker by the listener, and the construction of a fluent conversation by both participants.

3.1.1 Prosody

Wennerstrom (2000) proposes that intonation is another component of fluency. In her study, native speakers' ratings of non-native speakers' speech were compared with an analysis of whether or not the non-native speakers used pitch to differentiate such things as new and old information in the structure of the discourse, and intonation at the end of turns. She notes that the lack of pitch accents, where all words have equal pitch, results in a choppy, one-word-at-a-time effect, which results in a perception of poor fluency by native speaker judges. Furthermore, the lack of nativelike use of boundary tones at pauses and end of turns can result in the nonnative speaker losing control of their turn. The speakers who were judged to be less fluent used inappropriate boundary tones – either too frequently placed, cut short, or with an inappropriate contour, which often resulted in interruptions by the native speaker interlocutor. However, appropriate use of boundary tones allowed even relatively long pauses to be tolerated without interruption by native speakers. She concludes by noting, “Thus it is not longer utterances or shorter pauses per se that lead to a perception of fluent speech; instead, it is the ability to speak phrasally rather than word-by-word, focusing the main idea of each utterance in a coherent manner and collaborating in the turn-taking process” (Wennerstrom, 2000, p. 125).

3.1.2 Nonverbals

Bavelas (2000) studied the relationship of nonverbal aspects of communication to fluency. Bavelas considers certain nonverbal acts, in particular, “those visible acts that convey meaning and are tightly synchronized with speech” (Bavelas, 2000, p. 91) to be genuine linguistic units. This includes hand gestures and facial displays (but not facial expressions, which she notes are traditionally considered as inadvertent expressions of emotion). Bavelas found that these facial displays and hand gestures were non-redundant, in that they added meaning to the speech or to the discourse in ways that were not present in the spoken words. In fact, in some cases, listeners responded verbally to nonverbal signals given by the interlocutor. This implies that nonverbals are an important element of communication, even though they have often been associated with lack of fluency, that is, as a crutch rather than an essential part of communication. Most fluency research has not used multimodal data, which misses this important aspect of communication.

Research has shown that the learners who use gestures are evaluated more highly in terms of proficiency (Gullberg, 2008), and as learners become more proficient, their gestures become more meaning-oriented (Gregersen, Olivares-Cuhat, & Storm, 2009). However, it has also been argued that gestures and speech form a two-way system of cognition, where gesture informs and alters verbal thinking (A. Clark, 2011). That is, learners can use gestures to off-load some of the cognitive load of words, grammar, and other parts of language planning that must be kept in mind while speaking (Gullberg, 2008). Thus, gestures can help both the speaker and the listener in the creation of fluent speech.

3.2 Fluency as interaction

Temporal variables are often ambiguous, and psycholinguistic aspects of fluency are generally inaccessible or difficult to measure. For these reasons, the fluency of L2 speakers is usually measured holistically by native speaker raters. Fulcher (1996) created bands of descriptors derived from studying the language of learners grouped into five categories of fluency, and found that speakers put into different categories of fluency varied in more ways than just temporal variables. Differences in the five bands included listening comprehension, length of utterances, presence or lack of backchanneling, expansion or lack of expansion of utterances, and the ability to repair problems in communication. This suggests that fully capturing what constitutes perceived fluent performance in conversation involves more than just temporal factors on the speaker’s side.

In conversation, all participants are both speakers and listeners, and comprehension problems of the nonnative speaker as a listener can lead to breakdowns in the overall fluency of the conversation. Morales-Lopez (2000) compared eight conversations between nonnative speakers and native speakers, where the nonnative speakers were divided into three fluency levels, and concluded that fluency is the ability to (1) "maintain a certain degree of parity in turn exchanges - balancing the amount of information by each participant”, (2) "repair and renegotiate topic after conversation 'flow' has been disturbed”, and (3) "convey and interpret the multiple communicative inferences that underlie and contribute to the development of the interaction" (Morales-Lopez, 2000, p. 283). This includes social and cultural knowledge. In her study, she also found that the listening comprehension of the nonnative speakers was very important. The intermediate-level learners had good comprehension skills with weak production skills, but were still able to maintain conversational flow with some help from the native speaker. The lower-level learners in her study, with weak comprehension skills, sometimes interrupted the native speaker and seemed to lack awareness of where the conversation was going.

3.2.1 The role of the listener

Bavelas, Coates, and Johnson (2000) found that listeners' verbal and nonverbal backchannels function to show the speaker that communication has been successful, and more specific lexical backchannels can even contribute to the construction of the speaker's narrative. In their study, when listeners were given a concurrent task which prevented them from backchanneling, the speakers' narratives became much less fluent and often fell apart. Furthermore, speakers tend to use more words when there is no feedback (Krauss & Weinheimer, 1966), and listeners are able to understand more when they can provide feedback (Kraut, Lewis, & Swezey, 1982). Wolf (2008) also found that Japanese non-advanced learners performed better in terms of several temporal measures of fluency (such as speech rate) when both verbal and nonverbal backchannels were used by the listener, and worst when listeners used no backchannels.

This suggests that fluency in conversation may be co-constructed by both (or all) participants. McCarthy (2005) calls this aspect of fluency *confluence*, and notes that speakers scaffold each other's performance, for example, by finishing each other's sentences or suggesting words that the other speaker might be searching for, which results in a fluent conversation in spite of the apparent dysfluencies of the individual participants.

There is also evidence that the rhythm of speech is a factor in dialogic fluency. Fiksdal (2000), studying conversations between native speakers and L2 speakers, found that the periods when the participants felt that the conversation was most fluent was when they were able to match tempo (explained in more detail below) and used the same rapport system. She argued that the relationship to the interlocutor, the level of formality, the topic, and other more social factors may affect fluency, through the ability or lack of ability to build rapport. This includes such things as using phrases like *you know* to establish common ground in the conversation. However, mixing rapport systems, such as in her example of an Asian student using *deference* face strategies to repair a misunderstanding with an American who is using a *rapport* face system, results in what she calls an "uncomfortable moment", which is felt as dysfluency by the participants in the conversation.

3.2.2 Smallwords

Relevance Theory (Sperber & Wilson, 1995) can help to explain how the interaction between participants in a conversation can lead to successful communication and fluency. Relevance Theory assumes that utterances are underspecified for meaning, and successful communication is when the listener is able to make the correct inferences intended by the speaker. To do this, participants must be able to identify the communicative intention of the speaker, the context for interpretation of the utterance, the effect of previous utterances, the degree of vagueness or commitment to the utterance, and the state of success of communication so far. Hasselgreen (2005, p. 135) claims that *smallwords,* which are "small words and phrases, occurring with high frequency in the spoken language, that help to keep our speech flowing, yet do not contribute essentially to the message itself", accomplish many of these functions. These overlap with some of what others refer to as *backchannels*, *discourse markers*, and *vague language*. Smallwords, such as *well, right, you know, like,* and *kind of*, help the speaker by guiding the listener to the appropriate interpretation of the utterance. Hasselgreen (2005) found that in addition to temporal variables such as rate of speech, nativelike use of smallwords is a mark of higher L2 fluency.

3.2.3 Turn boundaries

McCarthy (2008) noted that fluent conversation is a joint enterprise where both participants attend to one another. By attending to one another, every turn is a response to the previous turn. Schegloff (1996) states that turn boundaries are critical points in conversation. Turn beginnings help the recipient anticipate the shape and the type of the ensuing turn (Schegloff, 1987), which then helps the listener predict the ending of the turn. Tao's (2003) corpus-based study of turn initial items found that turn beginnings were usually syntactically independent smallwords such as *yeah*, *mm hm*, *oh*, *and*, *well*, and *right*, that served to link that turn with the preceding turn, or comment on the preceding turn.

Certain lexical items may also facilitate turn endings. Evison and McCarthy (2013) identified lexical items and multi-word chunks that typically occur at the end of a speaker's turn and invite speaker change, especially by eliciting some kind of agreement token from the interlocutor. Examples are vague expressions such as *and things like that*, and evaluative adjectives such as *lovely* or *awful*, which invite a convergent response.

Smooth turn boundaries are crucial to creating the impression of flow in dialogic fluency, and are dependent on each participant's ability to predict the endings of the other's turns. Stivers et al. (2009), in a study comparing turn-taking across 10 languages, found that in all languages, speakers try to maintain the shortest pauses between turns and avoid overlap. Different cultures were found to have different norms for what is considered a delay, but these differences were relatively small, and when taken into consideration, support the universality of minimal pauses and avoidance of overlap. Pauses that deviate from the local norm generally have significance, for example, signaling a problem with the previous utterance (Stivers et al., 2009), and could be problematic for an L2 speaker who applies L1 turn-taking practices to the L2. They also found that questions accompanied by eye contact resulted in faster response times, showing again the importance of nonverbal cues in dialogue. Furthermore, what they call *visual responses*, i.e. head nods and other gestures, were also faster than purely verbal responses.

In order for speakers to participate in conversations with minimal gaps and minimal overlap, they need to be able to anticipate the end of turns, i.e., they need to be able to anticipate *turn completion points (TCP)*, which occur at *transition relevance places (TRP)* (Sacks et al., 1974). A TRP is a point in the conversation where the turn could go to another speaker, or be continued by the same speaker. De Ruiter et al. (2006) investigated the causes of the ability to anticipate turn endings, by artificially removing words, intonation, or both from utterances. They found that listeners relied more on the words, that is, the syntactic and lexical cues, and lack of intonation did not prevent listeners from being able to anticipate turn endings. However, their results suggest that intonation can be used to keep a speaker's turn during a pause where the speaker intends to continue, which supports the findings of Wennerstrom (2000), noted above. Furthermore, lexical cues, particularly the filled pause *uh* can also override other lexical and syntactic cues which would indicate a TCP.

3.2.4 Rhythm and tempo

There is also evidence that the rhythm of speech is a factor in dialogic fluency. Rhythm refers to the regular beat of stressed syllables in time, and tempo or pace refers to the rate of stressed syllables per minute (Vanderplank, 1993). When conversation feels fluent, speakers tend to converge on a single isochronous rhythm, which is maintained over turn boundaries between speakers, and in spite of false starts and hesitations (Auer et al., 1999; Fiksdal, 1990). It has been suggested that the short pauses that occur between turns occur in order to maintain this rhythm, and that anything other than one beat of silence is marked and has some kind of meaning (Schegloff, Koshik, Jacoby, & Olsher, 2002). For example, it could be a sign of a dispreferred response (Pomerantz, 1984). Synchrony in rhythm is one factor that enables smooth turn boundaries (Couper-Kuhlen, 2001), for example, by helping listeners to anticipate the next word (Martin, 1972). When the rhythm of the speakers is not synchronous, it is recognized as a problem in the conversation (Gumperz, 1982). Therefore, within the framework of dialogic fluency, pauses between turns could be a necessary component of speech to maintain synchrony in rhythm between speakers.

Synchrony in rhythm is not, however, a necessary or automatic feature of conversation, according to Auer et al. (1999). Rather, they argue that it is something that is achieved by the speakers in the conversation, perhaps when there is rapport between the speakers, as suggested by Fiksdal (1990). In the case of learners, there may be delays that break the rhythm of the conversation because of language problems. In particular, Carroll (2000) found that Japanese learners had delays in the starts of turns that occurred after dysfluent turns by the previous speaker, i.e., turns that contained lengthy and/or numerous pauses, and these kinds of pauses could result in breaking the rhythmic pattern of the conversation.

3.2.5 Interactive alignment

Traditionally, accounts of language production have focused on monologue, and treated dialogue as if it is merely alternating bits of monologue. This can also be seen in the dichotomy above between psycholinguistic theories of fluency as automaticity, which tend toward a monologic perspective of speakers, and sociolinguistic theories of fluency, which focus on interaction and turn exchanges. A recent psycholinguistic theory that may help to bridge the gap is Pickering and Garrod's (2004) theory of *interactive alignment*. They consider dialogue to be more the more basic genre of speech than monologue, and argue that language production in dialogue is easier than in monologue because the utterances of each speaker act to prime the same linguistic representations in the listener. These automatic priming mechanisms operate at all levels including lexical items, syntax, prosody, pronunciation, and semantics. This results in the linguistic representations being available for the listener in speech production, as well as alignment of *situation models*, which provide for a shared context of interpretation for the utterances used in the conversation. This theory accounts for why speakers tend to converge in many aspects of language, including speech rate, pause length between turns, syntactic patterns, and lexical choice, as has been reported in many studies (Aijmer, 1996; Kousidis & Dorran, 2009; Tannen, 1989). Pickering and Garrod also claim that this process can result in the use of *routines*, which are temporary formulaic utterances used in the same way and repeatedly by all participants in the conversation.

In their theory, monologue is simply a special case of dialogue, where there is less or no interactive alignment between speaker and listener, so that the speaker and listeners must rely on more conscious assumptions about shared knowledge and therefore require more processing to produce language, resulting in less fluency. As above, considering the role of the listener, monologic speech production occurs independently of a listener, and requires the speaker to consciously try to guess what the listeners (whether present or not) will understand. In dialogue, interactive alignment automatically creates a shared situational model, and primes language that eases speakers' production.

Interactive alignment has the potential to account for many phenomena in dialogic speech, such as (1) speakers converging on lexical use, syntactic structures, pause length, prosodic features, and speech rate (Kousidis & Dorran, 2009; Street, 1984), and (2) the ability of interlocutors to sometimes complete the speaker's sentences, and otherwise construct structures that connect across turn boundaries, as has been observed with *which*-clauses (Tao & McCarthy, 2001).

3.2.6 Interactive competence

The view of fluency as co-constructed aligns closely with Sociocultural Theory, where mental functioning and cognitive development are described as socially shared (van Compernolle, 2014). Sociocultural Theory focuses on the social function of language, and conversation is seen as one kind of joint, culturally organized activity. The participants in a conversation support each other by aligning to each other and scaffolding each other's performance (Lantolf & Thorne, 2006). A second language learner could be able to participate in a conversation at a level beyond their abilty to speak alone with help from a more proficient interlocutor, assuming that the learner has enough competence to reach their *zone of proximal development,* where they are not overwhelmed by the difficulty of the task.

The interactive skills needed to co-construct fluent conversation could also be seen as components of *interactional competence*, which is the knowledge that people use to engage in interaction with others, interpret the actions of others, and connect their own actions to others (Hall & Pekarek Doehler, 2011). This was proposed as a reaction to earlier views of communicative competence (composed of grammatical, sociolinguistic, discourse, and strategic competencies), which were seen as static cognitive properties of individuals. Interactional competence recognizes that one person's competence varies with the interlocutor, so that "interactional competence is not the knowledge or the possession of an individual person, but it is co-constructed by all participants in a discursive practice" (Young, 2011, p. 101). Dialogic aspects of fluency fit well within the construct of interactive competence.

3.3 Summary

In summary, it seems that there are two main ways of looking at fluency. One is a psycholinguistic construct (cognitive fluency), based on the efficient organization of the language in long-term memory, and rapid access to those memories, for example, with the use of formulaic sequences, which is manifested in longer uninterrupted runs of speech and higher rate of speech, and fewer pauses internal to clauses (utterance fluency). The other sense of fluency is dependent to some degree on the psycholinguistic construct, but co-constructed by all participants in the conversation, and manifested through the co-construction of a conversation with smooth turn exchanges (confluence).

The monologic construct of fluency is the traditional way of studying fluency, and is seen mainly as function of automaticity in language production, which depends on how language knowledge is stored in long-term memory, and the ability of the speaker to rapidly access these memories when necessary. Of course, even native speakers have times when these memories are not easily accessible, for example, when someone forgets a word for something when speaking, but in general, the more that the words, constructions, or chunks of language are used, the more readily they will be accessed. Second language users can have additional difficulties due to inefficient storage, lack of formulaic sequences, lack of vocabulary knowledge, or weak connections to memories of previously encountered language. Monologic fluency or cognitive fluency corresponds more closely to psycholinguistic definitions of fluency. Monologic fluency is independent from accuracy in the sense that interlanguage knowledge can be inaccurate but rapidly and efficiently accessed, in the same way that proficient speakers of different dialects of English can have different ideas of what is accurate, but both would still be fluent.

On the other hand, there is a dialogic construct of fluency, or confluence, which is co-constructed by the interaction between participants in a conversation. Dialogic fluency is influenced by many factors, such as the use of nonverbals and alignment of rhythm in speech, as well as alignment in grammatical and lexical choice. Deviations in grammatical constructions and lexical choice that is unfamiliar to the listener can produce difficulties in the anticipation of turn completion points, as well as difficulties with the construction of relevant interpretations of the speaker's utterances, which could lead to problems in the confluence of the conversation.

The common sense notion of fluency as smoothness of speech, and ease of speaking and understanding, could mainly be a function of dialogic fluency, where the two (or more) speakers work together to co-construct a conversation that as a whole is fairly smooth and continuous, with minimal pauses between turns facilitated by rhythm through turn exchanges. However, dialogic fluency is also dependent on the cognitive fluency of both speakers. The more closely that the language organization of one speaker resembles the other, the more interactive alignment (through priming) should take place, so that each speaker helps the other to become more fluent, and therefore more confluent.

The distinction drawn between monologic and dialogic fluency mainly reflects the point of view of the researcher. Monologic fluency is the traditional view of fluency from a cognitive point of view, where fluency is a property of an individual speaker. Dialogic fluency is from more of a sociocultural or sociocognitive point of view, where fluency is seen as co-constructed. This distinction, however, is somewhat artificial in that there are modes of speaking that are more interactive, such as a typical conversation, and modes that are more monologic, such as a lecture, such that the difference between monologue and dialogue may be a continuum. Even self-directed speech could be in some sense dialogic if the speaker imagines an interlocutor or potential interlocutor.

The next section will investigate the interaction of these various factors through the close examination of a conversation between a nonnative speaker and native speaker of English. A conversation with a non-advanced learner of English has the potential to offer more insight into the mechanisms that are involved in the co-construction of a conversation, because it is likely that there will be more opportunities for problems or breakdowns than might be found in a conversation between proficient users of the language.

3.4 Pilot study: Analysis of a single conversation

3.4.1 Purpose of the study

In this and the last chapter, the concept of fluency was questioned and the relevant literature was examined, leading to two separate but interrelated concepts: *monologic fluency,* or *cognitive fluency,* as automaticity of language processing, and *dialogic fluency* or *confluence*, co-constructed by the participants. In order to investigate more closely both of these tracks of fluency in second language spoken English, it is necessary to look closely at examples of learner language for evidence of these two tracks of fluency, and their various components as discussed in the previous chapter. Because dialogic fluency is a construct that has not been researched well, particularly in second language speech, the pilot study was a case study employing discourse analytic techniques informed by Conversation Analysis. As a type of qualitative inquiry, these methods are classified as hypothesis-generating rather than hypothesis-testing (Markee, 2000; Richards, 2003; Yin, 2003), which was appropriate for looking for evidence of dialogic fluency in spoken second language data and the relationships between the various factors involved. However, as will be explained in more detail below, many quantitative measures of the traditional temporal variables used in fluency research were used to make comparisons between the speakers in the study and with previous research. In looking at a single conversation in detail, it is possible to look at the wide variety of features that have been related to fluency at the same time, and their relationships within the single conversation.

The study reported in this chapter is a case study of one learner in a recorded conversation with a native speaker. One learner was chosen, rather than several, in order to more closely examine the features of this learner's English and what they reveal about both monologic and dialogic fluency, and the impression of fluency on the native speaker interlocutor, without having to make generalizations across a group of learners that might hide the complex relationships that occur in any actual single case. The learner in this study was a female Japanese university student at a major university in Japan, majoring in English Communication. Overall, her English proficiency as demonstrated in classes was not outstanding, and she would be classified as intermediate or lower intermediate, or a B1 in the Common European Framework. At this point, she had not taken any English proficiency exams that could be used as a reference, such as TOEIC or TOEFL.

3.4.2 Methodology

This study involved recording a single conversation between the learner and a native speaker. The native speaker was a male American teacher at the same Japanese university, but the two of them had never met prior to the conversation. The fact that they did not know each other could have meant that the learner would have been very nervous, but she reported afterwards that although she was nervous at the beginning, within about one minute into the conversation, she relaxed and felt comfortable.

After completing the conversation, I interviewed each participant individually while listening together to the recording of the conversation. The purpose of interviewing the native speaker was to get more detailed information on the native speaker's impressions of the learner's fluency, and to be able to identify places in the conversation where the native speaker felt that the conversation was particularly fluent or not fluent. Furthermore, in cases where the reason for interruptions or long pauses was unclear from the conversation, the native speaker could be asked directly about this. In the case of the learner, I asked the learner to explain the reasons for hesitations that seemed noticeably long, to identify parts of the conversation that she felt were particularly fluent (either from her perspective as a second language speaker, or looking at the conversation as a whole) and parts that were particularly dysfluent. In the case of hesitations, it is impossible to determine from recorded conversations alone whether hesitations are due to lack of knowledge, or slow processing, or simply thinking about content. It has been noted, however, that the speaker's introspection may not be reliable in all cases (Hopper & Drummond, 1990; Lennon, 2000). There are several limitations with this methodology. One is that the conversation must be fairly short to avoid the participant getting tired while listening to it. In the case of a five or six-minute conversation, the following interview took about 30 minutes. A second problem is that if the interview is done immediately after the conversation, the researcher does not have time to identify places that need clarification, which usually become apparent during transcription. On the other hand, if the interview comes after the transcript is prepared, the participant's memory of the conversation will have likely faded, making the introspection even more unreliable.

The duration of the conversation was approximately six minutes, and was recorded in wav format using a Edirol R-09 IC recorder. The quality of the recording was high enough to identify the words clearly and to hear breaths before speaking. The recording was transcribed using ELAN linguistic annotation software (*ELAN — language archiving technology*, 2012), which allows annotations to be visualized in time along with the original waveform of the audio recording, and which allows the two speakers' transcriptions to be placed on separate tiers. Pauses and overlap in speech can be identified visually very easily and pause length can be measured accurately to the msec (see Appendix A for transcript of this pilot study).

ELAN allows multiple tiers of annotations to be created, so the first stage of transcription was to create two tiers—one per speaker—defined as *utterances*, which were defined as sequences of speech between silent pauses. At this point, following Riggenbach (1991), I included filled pauses such as *um* and *uh*, in the utterances, as well as all other vocalizations, such as laughter (see the transcription conventions inTable 3.1). Still, many questions arise regarding annotations of this sort. Should pauses that occur between turns be attributed to the previous or following speaker? In general, the methodology of previous research and theoretical considerations may provide a method for annotating. However, in many cases, these questions can only be answered by looking at each individual example and the context around it. Initially, pauses between turns were annotated on a separate third tier, and then analyzed in more detail later. Furthermore, pauses that were shorter than 200 msec are generally considered to be articulatory rather than hesitations (Kormos & Dénes, 2004; Riggenbach, 1991), and were therefore not transcribed. Transcription conventions followed those typically used in Conversation Analysis (Richards, 2003) and by Chafe (1994) (Table 3.1).

Table 3.1 Transcription conventions.

-	Cut-off intonation, either mid-word or mid-utterance
.	Low pitch boundary tone
(none)	Plateau boundary tone, or partially falling boundary tone (at the end of an annotation preceding a pause, no symbol implies a plateau boundary tone)
,	Low rising boundary tone
?	High rising boundary tone
[u]	Brackets around IPA symbols are used to mark vowels added to the end of some words. In general, "non-standard" pronunciation is not annotated; this annotation is only used when the non-standard final vowel is elongated as a hesitation device.
hn hn	Laughter is usually spelled as it sounds with hn, huh, or heh.

The second stage of transcription was to separate filled pauses from continuous utterances. Child tiers dependent on the associated utterance tier of each speaker were created, and filled pauses were marked as separate annotations from the other words of the continuous utterances. A third tier combining single speaker turns into single annotations was then created, and an additional tier annotating the pauses between turns was made. Finally, a tier of commentary from the post-conversation interview with each participant was added to allow the commentary to be lined up with the relevant moment of conversation.

The next two sections will report the results of the analysis of the conversation. First, I looked at fluency from the traditional monologic perspective—looking at the learner's individual utterances and comparing them with the native speaker in the conversation. In particular, this section will deal with temporal variables and formulaic sequences used. Second, I looked at fluency as interaction, or confluence, by looking at the fluency of the conversation as a whole, especially as evident through turn exchanges and alignment of rhythm.

3.5 Monologic fluency

3.5.1 Temporal variables

In order to evaluate fluency as a temporal and psycholinguistic phenomenon of the speaker (utterance fluency and cognitive fluency), it is necessary to measure the durations of utterances and silences themselves. The following temporal variables were calculated.

1. Mean length of run (MLR): A run is a continuous utterance between pauses of more than 200 msec. This calculation followed Riggenbach (1991) by including filled pauses as words.
2. Speech rate: This is the number of words or syllables divided by the time spent speaking during a turn. Pauses that occur between turns (that is, between speaker changes) were not figured into the calculation, following previous research (Riggenbach, 1991). It could be argued that some turns between pauses were really turn-internal pauses, where the other speaker jumped in to help. However, this is sometimes difficult to determine with certainty, and would likely not be numerous enough to affect the overall result significantly. Furthermore, utterances containing substantial amounts of laughter were not included into the calculation, because although laughter is relevant to conversational interaction, it is not necessarily intentional and cannot be clearly classified as speech for calculations of words or syllables.
3. Articulation rate: This is calculated the same way as speech rate, except that unfilled pause duration is not included.

4. Phonation time ratio: This is the total time speaking divided by the total time of the speaker's turns including unfilled pauses.
5. Pruned speech rate: This was calculated following previous research (Derwing, Rossiter, Munro, & Thomson, 2004; Kormos & Dénes, 2004; Rossiter, 2009), which involves removing filled non-lexical pauses, self-corrections, repetitions, and false starts. However, the fact that filled pauses may add to comprehensibility (Corley et al., 2007) and that repetition and apparent false starts can serve communicative functions (Carroll, 2004) makes this temporal variable somewhat suspect. Furthermore, in dialogue (and in this particular conversation), some false starts are due to reacting to the interlocutor, and should not be considered dysfluencies. When it was clear from the conversation analysis that this was the case, the relevant syllables were included in the calculation.
6. Number of pauses within turns: A frequently used temporal variable is pauses per minute, and average pause length. However, looking at the conversation closely, it became apparent that these measures are too simplistic to be useful. There are many reasons for hesitations, with some of them being signs of fluency rather than dysfluency, as noted above. Therefore, the number of pauses and types of pauses, combined with the speech rate were looked at first, and then particular cases of long pauses were examined individually.

The results of these measurements for both the nonnative learner and the native speaker in the conversation were calculated and compared (Table 3.2).

Table 3.2 Temporal variable results.

	learner	*native speaker*
mean length of run (words)	2.14	4.57
mean length of run (syllables)	2.60	5.63
no. of pauses within turns	53	21
no. of turns	113	111
no. of utterances	182	120
speech rate (words/min.)	122	189
speech rate (syllables/min.)	149	233
average word length (syllables/word)	1.2	1.2
articulation rate (syllables/min.)	171	250
phonation time ratio	88%	93%
pruned speech rate (syllables/min.)	140	218

The native speaker's speech rates (189 words per minute and 233 syllables per minute) fell near the border between "average[6]" and "moderately slow" speech in conversation (Tauroza & Allison, 1990), while the learner was very much "slower than normal" (149 syllables per minute). The pruned speech rate for the learner was 140 syllables per minute (94% of the unpruned rate), and for the native speaker was 218 syllables per minute (also 94% of the unpruned rate), which implies that there was not a great difference between the learner and the native speaker in the use of repetition or restarts.

Mean length of run is considered to be an important measure of fluency because it directly reflects the automaticity of the underlying psychological processes (Towell et al., 1996; Wood, 2001). In this conversation, the mean length of run for the native speaker was 4.57 words, which is normal for English (Chafe, 1994). The learner's mean length of run was much shorter, at 2.14 words. One-word utterances accounted for 50% of the learner's uninterrupted runs, while this was

[6] According to Tauroza and Allison (1990) "average" is 190-230 words or 230-280 syllables per minute. "Moderately slow" is 160-190 words or 190-230 syllables per minute. Anything slower than 160 words or 190 syllables per minute is "slower than normal". They claim that the syllable rate is absolute, but the word rate varies based on the context. The speech rate quote here is based on casual conversation.

17% for the NS. For the learner, many of these were backchannels, but the others involved apparent difficulties constructing phrases and selecting lexical items and grammatical structures.

Looking at the hesitation devices used, I first labelled all unfilled pauses and non-lexical filled pauses for both speakers, dividing them into five categories based on the contexts they occurred in:

1. pauses directly after a turn initiator (for example, “but um” or “and uh” used at the beginning of a turn)
2. pauses used to initiate a turn
3. pauses before a content word (usually a noun or verb)
4. pauses between clauses
5. other—mainly pauses internal to clauses and pauses after false starts, which would be classified as dysfluent.

The learner used more pauses overall (see Table 3.3), but particularly, the learner relied much more on silent pauses than the native speaker, and, as would be expected, had many more clause-internal pauses (21 for the learner vs. 6 for the native speaker). The native speaker often used filled pauses at the beginning of turns, particularly in expressions like *but uh* and *and um*. These were often followed by a silent pause. On the other hand, the learner primarily used silent pauses after turn initiators. Pauses before content words, which usually are a sign of lexical choice (as reported by the learner in an interview after the conversation), were almost always unfilled in the learner’s case, but always filled in the native speaker’s case.

Table 3.3 Number of unfilled and non-lexical filled pauses.

	learner		*NS*	
	unfilled	*filled*	*unfilled*	*filled*
after turn initiator	10	3	9	14
as turn initiator	N/A	6	N/A	7
before content	7	1	0	6
between clauses	26	1	10	1
other (dysfluent)	19	2	1	5
total	62	13	20	33

The mean length of pauses (both filled and unfilled) was the same for both speakers, at 0.47 seconds. In one case, the learner used a cluster of three hesitations (a silent, filled, and then silent pause) for a total of 1.40 seconds, due apparently to lexico-grammatical choice, when she says *and then we talked about (.26) diary (.42) uh (.41) students each other*. In another case, she had a 1.51 second combination of unfilled and filled pause, while trying to remember the content of what she wanted to say. However, she began the filled pause after 0.85 seconds of silence, which she reported as feeling uncomfortably long. One of her more dysfluent turns was *yeah I know but (.57) I: (.21) I have to: (.76) I have- (.25) to:: work more because (.33) I (1.01)*, after which her sentence was completed by the native speaker. Another example that illustrates the different ways the native speaker and learner use pauses is a point in the conversation where the learner said *same (.21) to (1.16) Kirk* but the native speaker then corrects her with *same as uh Steve*. These two phrases, with essentially the same content, involve a pause before a content word, but the learner used an unfilled pause while the native speaker used a filled pause.

3.5.2 Formulaic language

Looking more closely at the learner's longer turns, many of them were constructed one or two words at a time, and have many grammatical errors. This single conversation, with 394 tokens and 138 word types, is too small to be subject to corpus linguistic techniques; however, looking at the sequences of words the learner is able to utter as uninterrupted runs, the data suggests that some sequences may be formulaic for the learner—where the term *formulaic* implies holistic storage—and might be idiosyncratic to the learner's interlanguage. *I want to go* was spoken twice and both times as an uninterrupted sequence. *So where are you from?* was one of her longest spoken runs. Another long run was *yes do you know Decks Tokyo Beach?* After the turn initiator *yes*, the next three words were uttered one other time in the conversation, also as a continuous phrase, and are not infrequent in learner input. The last three words make up the name of a shopping center, which would almost certainly be stored whole. So, this could very well be three sequences combined together. The learner's longest run in this conversation was *yes: so: last month, I: use a lot of money, to buy.*, which is 12 words long. Unfortunately, there was no evidence in this conversation for the formulaicity of any of the parts of this utterance because of the small size of this conversation as a corpus (390 words). One of the learner's 7-word utterances was *my seven days is almost all days.* The phrase *my seven days*, although not particularly nativelike, was used one other time a few turns before this occurrence. Also, the phrase *almost all days* I believe to be formulaic because I have observed this learner using this same (idiosyncratic and ungrammatical) phrase on at least two other occasions in other conversations. In general, most of the turns were made up of simple grammatical constructions, and sentences that involve more complex grammar generally seemed to result in more silent pauses. For example, *so where do you (0.6) often (0.8) go (.) to buy sh- (0.4) clothes* (see also Extract 1) involves an infinitival complement after the main verb.

3.6 Dialogic fluency

3.6.1 Strategic competence

As is to be expected, the conversation was generally led by the native speaker, who chose most of the topics and asked most of the questions. However, there were times, increasingly so as the conversation progressed, where the learner initiated topics and asked questions which moved the conversation forward. The native speaker commented that this was one of the aspects of the learner's conversational ability that impressed him the most, and made the conversation feel more balanced. To attempt to quantify this, the native speaker asked 23 direct questions during the roughly six minutes of conversation, and the learner asked 8. The numbers may not be equal, but in addition to being a native speaker, he is a teacher and he is older than the learner, all of which gives him natural power and authority in the conversation.

The learner generally showed good listening comprehension throughout the conversation, and was able to initiate repair when she did not know words. In one case, she did not understand the word *debt*, and immediately repeated it to ask for clarification. There were also two cases where she was able to clarify her meaning when the native speaker signaled a lack of understanding. In Extract 3.1, the learner (A in the transcription[7]) used an idiosyncratic phrase *dress style*, which was followed by a hesitation and repetition (line 148) by the native speaker (B in the transcription). She recognised the problem and rephrased the expression as *one-piece* (this is a Japanese loan word from English that means *dress*), which the native speaker shows he understands.

[7] See Appendix A for a full transcription of the pilot study.

Extract 3.1

```
144 A   yeah-
145     (744) [yeah.]
146 B         [so   ] uh what kind of fashion do you like.
147 A   (336) yeah dress style.
148 B   (522) dress style like u:[:h-    ]
149 A                            [one pie]ce.
150 B   (258) oh o[kay. ]
151 A             [yeah.] yeah.
```

Later (Extract 3.2), when asked where she is from, she used the Japanese place name *Shitamachi* in her answer, and then asked the native speaker if he understood to pre-empt repair.

Extract 3.2

```
227 B   how is what is uh monzennakacho like.
228 A   (492) u:[:h-        ]
229 B           [is it a ni]ce place?
230 A   (132) yes.
231     (204) shitamachi.
232 B   (174) ah [ok.       ]
233 A            [can you un]derstand? yeah.
234 B   mm hm.
```

Another time, when she was embarrassed at her lack of knowledge on the topic, she immediately initiated a topic change. Only once was there a problem that was not immediately resolved. At the end of the conversation, she misunderstood part of the question asked to her, and the conversation continued for a few turns until the misunderstanding was brought to her attention by the native speaker. Her strategic competence, shown by her ability to repair and control the topic, is quite good compared to her peers and appears to be a major contribution to the impression of high fluency.

3.6.2 Turn boundaries

Turn boundaries are important places to find signs of confluence in conversation. Out of a total of 164 turn boundaries, 22% were latched, 13% were early (overlapped), and 64% had silent pauses. The silent pauses between turns were a mean length of 0.42 seconds (SD = 0.26), with the longest pause at 1.36 seconds. The total pauses between turns accounted for 11% of the total time of the conversation. Of the five pauses that were longer than 1.0 second, two were cases of the learner being asked a question that involved a topic change, which could have been more taxing on her comprehension and required more time to think of a response. One of these was prior to the learner giving an answer that she felt was embarrassing, suggesting that the pause was intentional. The other one was where the learner faltered while searching for how to say what she wanted to say, until the native speaker completed her sentence for her. One of the longest pause (see line 199 in Extract 3.3) can actually be attributed to the native speaker, and comes when the native speaker was thinking of something to say on a topic started by the learner.

Extract 3.3

```
A   (120) so: where do you-
    (618) often
    (768) go to buy sh-
    (366) clothes?
B   (972) nn: i almost never go to buy clothes.
A   wow.
    (hoh hoh hoh ~h huh huh ~h huh huh ~h)
B   (heh heh ~h huh heh heh)
A   wow.
B   but nn::
    (1380) yeah [sometimes-]
A               [your      ] wife?
B   (264) sometimes my wife will buy me a shirt.
A   oh good.
```

This 1.4 second pause in line 205 was perceived as long by both speakers, as they both began speaking simultaneously afterwards, until the native speaker dropped out and the learner finished her question. This example shows that the learner was aware of when silences became uncomfortable, and understood that she was also responsible for attempting to keep the conversation going (which she also reported in the interview). The fact that both speakers began speaking at the same time after the uncomfortable pause, could show that they had converged on what was considered to be an unacceptably long pause. The native speaker commented that he was at a loss for how to continue the conversation because the topic was not of interest to him.

Generally, she was able to use appropriate boundary tones to signal the end of turns, or to hold turns (with a low-rise or flat boundary tone). However, there were two cases where she did not do this correctly, which resulted in being interrupted. In one case, shown in Extract 3.4, overlap occurred when the native speaker mistakenly interpreted the turn in lines 024-025 to have ended. In lines 020 to 023, she effectively held her turn even through long pauses with low-rise and plateau boundary tones and with the phrase *oh yeah,* but in line 025 she used a falling tone after *worked*. This tone, as well as the fact that grammatically the native speaker would not have expected a following direct object, could signal to the native speaker that the turn was complete (De Ruiter et al., 2006), and resulted in overlap with a question from the native speaker when the learner tried to complete her sentence.

Extract 3.4

```
017 B   so what did you do during your seven days.
018 A   (1032) my seven days?
019 B   mm hmm in your [seven day diary]
020 A                  [o:h           ] yea:h
021     (852) uh:: yest-
022     (390) yesterday,
023     (246) oh yeah.
024     (258) my seven days is almost all days-
025     (414) i worked.
026 B   (990) a:[h you have a part] time job?
027 A           [a part time j-   ]
028     ye[s. ]
029 B     [wha]t is your part time job.
```

Another case of interruption occurred earlier in the conversation when the learner again used a low (falling) boundary tone, but then continued the clause in the following utterance. The falling boundary tone occured on the word *diary* in line

010 of Extract 3.5. As above, in this case, the syntax of the phrase was likely a major factor in this point being interpreted as a turn completion point by the native speaker. Although the phrase in lines 090-010 was not error-free, this could be a point of syntactic completion, and without a plateau or low-rise boundary tone to hold the turn, it is likely to be interpreted as complete.

Extract 3.5

```
005 A   (815) yea:::h
006     (536) first,  we write do::wn
007     (260) seven days diary
008 B   (714) mm hm
009 A   a:nd the:n we talked about
010     (264) diary.
011 B   (420) a[:h ]
012 A          [u:h]
013     (411) students each other
014 B   (216) oh that sou[nds ] fun
015 A                    [yeah]
```

From the point of view of the learner as listener, in general, this learner showed a strong ability to respond quickly when the interlocutor’s turns ended, and as above, she showed an awareness of when a silence had become long. In Extract 3.6, the learner answered the native speaker’s question (lines 132-134) and the conversation stalled (lines 135-136), but both speakers began speaking simultaneously after a longer than average pause of 0.85 seconds (lines 137-138).

Extract 3.6

```
130 B   (246) but uh sounds like uh so you're working hard,
131 A   (540) yeah.
132 B   and buying clothes, what kind of uh- where do you buy your
clothes.
133 A   (558) ah u::::m
134     (648) ikebukuro.
135 B   (582) ikebukuro.
136 A   yeah.
137     (846) [i:-]
138 B         [ u:]:h-
139     (960) like at uh parco [o::r-]
140 A                          [yes. ] ye[s.    ]
141 B                                    [oh rea]lly?
```

Looking at turn initiators used by the learner, out of a total of 119 turns—following Tao (2003), backchannels and overlapping speech were also counted as turns—*yeah,* used 41 times, was by far the most frequent turn initiator. *Yeah* is also the most frequent turn initiator in native speaker corpora, but only accounts for 19% of turns (Tao, 2003), while in this learner's speech it accounted for 35%. Other frequent turn initiators were *yes* (18), laughter (7), *oh* (6), *uh* (5), and *so* (4). Although she lacked the variety of turn initiators used in native speaker speech, she did generally use appropriate turn initiators, which is a characteristic of more fluent non-native speakers (Hasselgreen, 2005). This learner also used backchannels appropriately, either with *yeah*, *mm*, or a response token like *oh good.* She also used two of the 19 smallwords in Hasselgreen's (2005, p. 165) list. She used *oh* 10 times to indicate a cognitive change of state, and *I know* once to indicate agreement. She also used *ah* as a smallword, but in the way it is commonly used in Japanese, to express a negative surprise, rather than in the nativelike use of a positive surprise (Hasselgreen, 2005).

In terms of the balance between the speakers in the conversation, the learner had 113 turns, and the native speaker had 111 turns. The learner's speech accounted for approximately 56% of duration of the conversation, and the native speaker's speech accounted for approximately 48%[8]. The learner took up more time with turns, but of course spoke more slowly and paused more frequently during the turns, which is to be expected and is therefore not a sign of higher fluency. The number of turns was almost equivalent. Although this is difficult to interpret by the raw numbers alone, looking at the conversation itself, the learner generally responded to the native speaker's turns without any problems. Furthermore, as the

[8] The total is greater than 100% due to overlapping speech.

conversation progressed, the learner sometimes elaborated her responses without prompting or further questioning from the native speaker.

3.6.3 Alignment of rhythm and tempo

The final analysis for this pilot study involved looking at the rhythm and tempo of the speakers for signs of alignment. The rhythmic pattern and the pace of the conversation is a perceptual gestalt which must be annotated by listening to the conversation for the regular beat of stressed words. It can vary as much as 20% in tempo and still be perceived as a continuous rhythm (Auer et al., 1999; Couper-Kuhlen, 1993). In order to verify my own reliability in perceiving the rhythm, I compared my own annotation of the thirty seconds of the conversation with that of an experienced researcher (Fiksdal, personal communication), and found only slight differences.

In this pilot study, I looked at several cases where the speakers' turn boundaries appeared to be smooth. In particular, the case shown as Extract 3.6 above, where the two speakers broke the silence at the same time, was analyzed in order to determine whether there was alignment of rhythm between the speakers that might have enabled the simultaneous breaking of the silence (Extract 3.7). The tempo of Extract 3.7 is approximately 100 bpm, which results in a single beat, annotated with a grave accent, being 0.6 seconds. Beats are defined by the presence of stressed syllables (the syllable immediately following each grave accent mark), but silent beats do also occur. Line 131 is an example of a single beat composed of a silence, followed by a response by the learner after the beat. This was followed in line 132 by a continuation of the prior turn by the native speaker with a single word anacrusis, leading to the stressed word *buying*. In line 133 the learner began her hesitation on the next beat after the native speaker's turn ended. Lines 134-135 both have single silent beats before the response, in one case by the learner and in the other by the native speaker. Finally, in line 137, there is a pause that encompasses two silent beats (0.85 seconds), which was perceived as abnormally long by both speakers. Both speakers break that silence on the following beat.

Extract 3.7

```
130 B   `but uh sounds like `uh so you're `working hard,
131 A   `(295) yeah.
132 B   and `buying clothes, so `what kind of uh- `where do you buy
your `clothes.
133 A   (558) `ah `u::::m
134     `(648) ike`bukuro.
135 B   `(582) ike`bukuro.
136 A   `yeah.
137     ``(846) [`i:-]
138 B           [`u: ]:h-
```

Extract 3.7 shows that there are cases where the learner and the native speaker converge on a single rhythm that continues through changes of speaker, and which facilitates smooth turn exchanges and simultaneous breaking of an abnormally long silence. However, contrary to previous research showing convergence of speech rates (Street, 1984) and convergence in articulation rate (Schweitzer & Lewandowski, 2013), in the present conversation, the two speakers have widely differing speech rates and articulation rates (see Table 3.2 above). It seems that the synchronization of rhythm is facilitated by the learner stressing a greater number of words, as compared to the native speaker. In lines 130-132 of Extract 7 above, the native speaker uttered a total of 23 words with 7 stressed syllables and one intervening silent beat, i.e., approximately one word in three is stressed. In contrast, in lines 195-198 of Extract 3.8, the learner uttered 10 words (including the cut-off word in line 197) with 8 stressed syllables and two intervening silent beats, i.e., approximately every word is stressed. Therefore, by stressing more words, the learner is able to maintain a certain rhythm with a lower speech rate.

Extract 3.8

```
195 A   (120) so: `where do `you-
196     `(618) `often
197     `(768) `go `to `buy `sh-
198     (366) `clothes?
```

3.7 Conclusions and research questions

To summarize, the speech of the learner in this study shows rather low fluency from a monologic point of view. Her speech rate was slower than normal, and she tended to produce short runs of speech with 84% of her utterances being three words or less. This implies that her fluency is weaker in terms of cognitive fluency (i.e., automaticity and efficiency of language processing), which could be due to a lack of vocabulary in general, and especially a lack of formulaic sequences, which would help to minimize the strain on cognitive processing and working memory.

On the other hand, the learner in this study may be somewhat atypical, in that she seems to be able to create an impression of higher fluency in conversation. At the level of strategic competence, she was able to show confidence, and was able to take control of the conversation, even if only to a small extent, and the ability to repair breakdowns in communication. She also shows the ability to use some nativelike turn initiators. She was also able to keep her turn while encountering language production difficulties by employing low-rise boundary tones, filled pauses, and repetition, and by avoiding long silent pauses. She also showed the ability to recognize when a silence had reached an uncomfortable length, and acted to fill it, which is part of "keeping up her end of the conversation."

One of the questions I asked of the participants in this study was to rate the fluency of the learner holistically on a scale from 1 to 10. Interestingly, the native speaker rated the learner as a 7, and the learner rated herself also as a 7. This confirms the results of previous research where learners who can "keep up their end of the conversation" are perceived to be fluent (Morales-Lopez, 2000). Both participants reported the impression that the conversation was comfortable, easy, and relatively smooth. It seems that the smoothness of this conversation is at the level of dialogic fluency, which is not the same as proficiency or Lennon's higher-order fluency (Lennon, 2000). Dialogic fluency is influenced by (1) each speaker's ability to predict the other person's turn endings, which in turn is influenced by each speaker's comprehension; (2) knowledge of vocabulary, collocations, formulaic chunks, and grammatical constructions; and (3) the use of prosody to maintain turns across silent pauses. Of course, it is also influenced by the converse: the ability of each speaker to construct turns that have predictable endings. In this sense, accuracy is a part of dialogic fluency because syntactic information is the main factor for predicting turn endings (De Ruiter et al., 2006).

An important part of dialogic fluency is the ability of speakers to construct relevant interpretations of the other speaker's utterances within the time constraints of real-time conversation. Listening comprehension, and the ability to predict what will the interlocutor will say is also important here. Using predictable formulaic chunks and collocations eases processing and is what allows listeners to sometimes finish the speaker's sentences for them (Garrod & Pickering, 2004; Millar, 2011; Pickering & Garrod, 2004). Again, this will be dependent on knowledge of vocabulary, grammatical constructions, and formulaic language, but it is also dependent on the proper use of turn initiators and smallwords. These words promote dialogic fluency by making the context for interpretation of the speaker's utterances clearer (Hasselgreen, 2005).

Finally, this learner showed some evidence of being able to converge on a synchronous rhythm with the interlocutor, which facilitates smooth turn boundaries and adds to the overall perception of a fluent conversation by participants (Fiksdal, 1990). For this particular learner, these dialogic aspects of fluency seem to be able to compensate for her lack of ability in monologic fluency. This suggests that there is more to the creation of a smooth conversation than the monologic aspects of fluency that have traditionally been the focus of fluency research. As I have argued above in this chapter, the perception of fluent conversation may be due in a large part to dialogic fluency, which are those aspects of fluent conversation that are co-constructed by both participants. Although the most current framework of fluency—that of Segalowitz (2010)—includes *interactive communication context* as a factor, his framework has the L2 speech production of a single speaker as the center of the model (see Figure 2.1 in Chapter 2), and does not account for the co-constructed, dialogic fluency of the conversation as a whole. Therefore, the present study seeks to answer the following major questions.

1. To what extent should co-constructed features of fluency be taken into account in definitions and frameworks of fluency?
2. What factors interact in a complex system that account for the dialogic (co-constructed) fluency of a conversation?

Within the above overarching questions, the following more specific questions will also be addressed.

A. Are co-constructed aspects of dialogic fluency able to create confluence and therefore, an impression of fluency, in spite of weaknesses in monologic fluency?
B. In particular, can alignment of rhythm between speakers lead to an overall higher impression of fluency in a conversation in spite of lower fluency in terms of temporal variables?

Additionally, in light of current debates about research methodology, including issues of qualitative versus quantitative methods (Dörnyei, 2007; Richards, Ross, & Seedhouse, 2012), and methods for investigating complex systems (Dörnyei, 2011; Larsen-Freeman & Cameron, 2008a, 2008b)—particularly for a complex construct such as fluency—the following question will also be addressed.

C. What are the most effective research methods for a study on fluency?

This will be done through a parallel case study comparing the learner from the pilot study in this chapter to a more typical high-proficiency learner, in order to be able to examine in detail the interactions between the various factors affecting monologic and dialogic fluency, as they develop in each individual case. The overall goal of this study is to develop a framework for looking at fluency as a dialogic, co-constructed phenomenon.

4 Methodology

4.1 Introduction

Fluency is a multi-faceted and complex phenomenon. Segalowitz (2010) defines three types of fluency: cognitive fluency, utterance fluency, and perceived fluency. Although *perceived fluency* begins to address the relationship between the listener and the speaker, it is still grounded in a view of fluency as something that is created by the speaker. In other words, he defines perceived fluency as the listener's perception (or evaluation) of the speaker's cognitive fluency based on the speaker's utterance fluency. *Utterance fluency* is based on cognitive fluency, but also affected by other factors[9]. However, in Segalowitz's model, *cognitive fluency*—the efficiency of the underlying cognitive processes of language production—is central and limiting. Utterance fluency is limited by cognitive fluency. It may seem to the casual observer that perceived fluency should be limited by utterance fluency, but as detailed in Chapter 3, this relationship may not be so simple.

In traditional models, fluency is monologic. It is fully dependent on the cognitive language processing of the individual, and conversation is viewed as alternating monologues between two or more speakers. However, as discussed in

[9] As discussed in Chapters 2 and 3, a speaker's actual speech output is not always as fast as the speaker's cognitive fluency allows. For example, speakers may intentionally pause for pragmatic or rhetorical reasons, or may intentionally speak more slowly for social reasons. Therefore, for many reasons, utterance fluency may not necessarily directly reflect cognitive fluency, although it may be limited by it.

Chapter 3, some researchers suggest that fluency is a dialogic construct, co-constructed by participants in the conversation. McCarthy (2005, 2009) introduces the term *confluence* to describe the overall fluency of a conversation, where the performance of individual speakers may not be as fluent as the conversation as a whole.

A comprehensive definition and framework of fluency needs to account for both the monologic and dialogic aspects. The present study aims to investigate the relationship between monologic and dialogic aspects of fluency in order to work toward a framework of fluency that explains the co-constructed fluency of a conversation. It is likely that fluency is a complex, dynamic system (Segalowitz, 2010), and therefore it may not be possible to isolate the various factors that have been argued to influence fluency. In other words, there are no independent variables that can be used as the basis for controlled experiments (Larsen-Freeman & Cameron, 2008b). Therefore, a methodology for approaching a complex system must focus more on individual cases, in all their complexity, without making aggregating generalizations across large groups (Atkinson, 2011; Larsen-Freeman & Cameron, 2008b). Case studies are the most appropriate methodology for these kinds of situations, where the researcher cannot control the events (Yin, 2003). The case study allows the researcher to focus on how the myriad of factors interact in the single case. Although the results of case studies are not generalizable to the entire population, they can inform theory. This approach can shed light on the interactions of the various factors contributing to fluency, which can lead to the design of a framework to guide future studies—both further case studies and controlled experiments.

In an effort to address the need for a combined account of monologic and dialogic fluency, the present study examined the contributions to the co-constructed fluency of a conversation made by both monologic and dialogic aspects, through a mixed-methods case study approach. Since individuals can vary widely in language ability, the case study approach was chosen to enable the details of the interactions between various factors related to fluency to be observed, as they unfold in the progress of the conversation. In the present study, this was complemented by doing two parallel case studies with two learners that have different profiles in terms of proficiency. Each case study is of a single Japanese learner of English in monologic and dialogic contexts. I examined what features exist in the dialogues that help to contribute to the construction of fluency of the conversation as a whole. Monologic features across both the monologues and dialogues were compared to act somewhat like a control group in an experimental design, i.e., in order to identify the co-construction of fluency beyond the monologic fluency of the participants.

Quantitative research focuses on trying to isolate variables, while qualitative research, such as case studies, look at the data in all its complexity (Holliday, 2002; Richards, 2003; Richards et al., 2012). As such, the goal of this type of research is to generate hypotheses rather than test them (Markee, 2000; Richards, 2003). In particular, in the present study, the goal of the case studies and the comparisons between them was to identify the inter-relationships between the various factors that contribute to the co-construction of a fluent conversation, in order to lead toward an improved definition and framework for fluency.

This is also a mixed-methods study. Although it primarily relies on a discourse analysis, case-study-based approach, it was supported with quantitative data comparing each subject across two contexts, and between subjects. Statistical analyses and corpus-linguistic analyses were used to compare fluency-related features, but as a case study, it is not possible or intended to generalize the results beyond these cases. Rather, the intention of this study was to examine the relationships between the multitude of fluency-related variables in order to elucidate the interactions between them and their effects on the overall conceptualization of fluency. Within paradigms of mixed-methods research, the present study can be classified as a *concurrent triangulation design* (Creswell, Plano Clark, & Garrett, 2008; Hashemi, 2012), where qualitative analyses are conducted concurrently with qualitative analyses, and the various analyses inform each other in drawing conclusions.

4.2 Participants

The subjects of the two case studies were both female Japanese learners of English, majoring in English at a university in Japan. Both learners had had six years of English study in junior high school and high school, as is typical in the Japanese context. Although students in Japan come to university with six years of experience learning English, their proficiency is generally low. English in Japanese schools is taught using a grammar-translation method, which trains learners to analyze and translate English sentences into Japanese, and does not help and may even be detrimental to their spoken language skills (Mulligan, 2005). In spite of pressure from the Japanese government to adopt more communicative methods of teaching English, this method is preserved by washback from the university entrance exam system, where speaking skills are not tested at all (McVeigh, 1997). Therefore, the typical Japanese university student has weak spoken fluency skills and relies on translating from Japanese to English, resulting in long silences in conversation (King, 2011).

Two other important aspects of Japanese culture are relevant to spoken fluency. The system of *sempai* and *kohai* is based on Confucian values (King, 2011) and requires, for example, people older than the speaker to be classified as *sempai* and people younger than the speaker to be classified as *kohai*. *Kohai* must show respect for *sempai*, and are expected to defer to the *sempai*, resulting in their appearing reticent. The system of *uchi* and *soto* places others either within (*uchi*) or outside (*soto*) one's social group, resulting in outsiders being ignored and again apparent reticence (Nakane, 2007). These effects are visible in the face-saving system used in Japanese culture, where not speaking is a way to protect the speaker's own face, which can be interpreted negatively by non-Japanese (Nakane, 2007), and lead to negative impressions of the fluency of conversations with English speakers due to clashes with the face-saving systems used (Fiksdal, 2000). This is relevant to the current study in that in the dialogues the student subjects spoke with North American English teachers who were older than the them.

The subject of the first case study was a female Japanese learner of English, who was a second-year undergraduate student. For purposes of anonymity, she will be referred to as Asami. Although she was chosen at random from approximately 60 students enrolled in a first-year course I taught, her performance in the pilot study (see Chapter 3) was remarkable in that she gave an impression of fluency in conversation that appeared to be higher than what was expected based on my own impressions of her vocabulary level and speaking ability from her performance in my class. Therefore, this subject was chosen for this case study because she seemed to be an atypical example of a lower-intermediate learner—particularly in terms of fluency features. It was thought that a case study of a more atypical learner might lead to insights into fluency that would be missed with a more typical lower-proficiency learner.

As a baseline measurement of the subject's English speaking proficiency, she took the *Versant English Test* (Pearson, 2011). This test was chosen since it focuses on the assessment of spoken production[10] and has been linked to cognitive aspects[11] of language production (Van Moere, 2012). The subject achieved an overall score of 41 (Figure 4.1), which corresponds to the *A2-Basic User* level of the *Common European Framework of Reference (CEFR).* According to the Versant test results documentation, the subject's capabilities can be described as:

> Candidate can handle short utterances using common words and simple structures, but has difficulty following a native-paced conversation. Pronunciation may sometimes not be intelligible; candidate speaks slowly and pauses, but can convey basic information to a cooperative listener.

The subject's fluency subscore was 40, which is described as:

> Candidate speaks with uneven or staccato pacing, although speech may contain some smooth runs containing several words. Frequent obvious pauses result in an irregular speech rate and some disconnected phrases.

Lastly, the CEFR description of an A2 learner is:

> Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g., very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate

---

[10] Versant is taken over the phone, and is automatically graded by computer. It has been shown that scores on this automated test correlate strongly with oral proficiency interviews (Bernstein, Van Moere, & Cheng, 2010).

[11] For example, the sentence repetition task, although seemingly simple and not communicative, test the ability of the speaker to use formulaic sequences to overcome the limitations of working memory when the sentences to repeat are sufficiently long and/or complex (Van Moere, 2012).

> environment and matters in areas of immediate need (Council of Europe, 2001, p. 5).

The subject's score also corresponds to a 14-19 on the TOEFL iBT Speaking test, an overall score of 49-78 on the TOEFL iBT, or a 619-661 on the TOEIC.

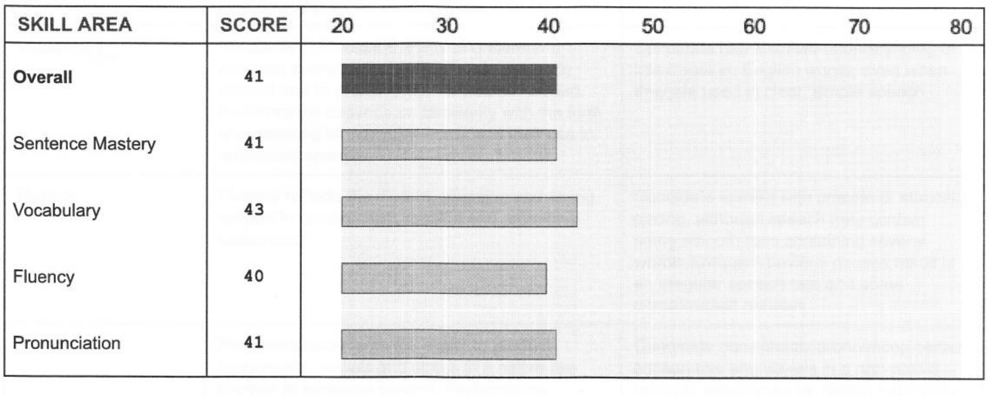

Figure 4.1 Versant English Test scores for Asami. Scores range from 20 to 80 points.

The second participant in this study was also a female Japanese learner of English in the same department at the same university in Japan, but was a fourth-year student. In this paper she will be referred to as Keiko. Comparing Asami with another learner of higher proficiency allowed the case studies to be more fairly compared than if she had been compared to a native speaker of English, as is typically used for the control in SLA research (Kachru, 1994; Sridhar, 1994). Both of the subjects were young Japanese women, both of them were majoring in English Communication, and both of them were outgoing, motivated students. Keiko had a similar background with regard to English language education, however, she was generally regarded as a higher-level student, and was consistently streamed into "advanced" sections of her required English courses. Her overall score on the Versant English Test was 50 (Figure 4.2), which is described in the results documentation as:

> Candidate can handle many utterances using a variety of words and structures, and can follow and sometimes participate in a native-paced conversation. Pronunciation

> is mostly intelligible; candidate can express some information on familiar topics to a cooperative listener.

Her fluency subscore was 54, which is described as:

> Candidate speaks with adequate rhythm but with some inappropriate phrasing and pausing. Hesitations and possible repetitions or false starts may sometimes interfere with smooth flow of speech.

This Versant score corresponds to a 17-21 on the TOEFL Speaking test, a 64-92 on the overall TOEFL, a 748-790 on the TOEIC, and a *B1* on the CEFR, which is globally described as:

> Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise while traveling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of a personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans (Council of Europe, 2001, p. 5).

It's interesting that with both participants, the CEFR rating based on my own experience as teachers of these students, and corroborated by the comments of other teachers, are higher than the level predicted by Versant. In the case of Asami, I described her as a B1 in the pilot study (Chapter 3), and I would have described Keiko as a B2. Keiko was chosen for the second case study because of her reputation as one of the highest proficiency students in that department, and that she seemed to exemplify a more typical higher proficiency learner, which could serve as a benchmark for comparison with Asami in the parallel case studies.

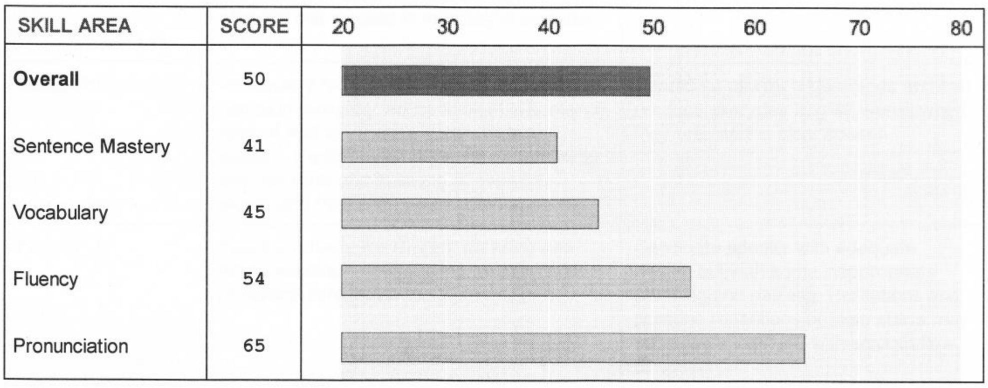

Figure 4.2 Versant English Test scores for Keiko. Scores range from 20 to 80 points.

4.3 Data collection

4.3.1 Speaking tasks and procedures

The main study in the present research used a multiple case study methodology, as described above, with two subjects. The case studies aimed to investigate the two learners in monologic and dialogic speech. The task was semi-controlled in that subjects were asked to retell the same movie to elicit the data. The movie used was the “Pear Story” (Chafe, 1980b), which is a five-minute color movie about a boy who steals some pears from a farmer. Although the movie has sound, there is no speech. The movie was designed with some retelling difficulties in mind. For example, there are several unnamed boys in the film, which prompts the use of different terms of reference or relative clauses to distinguish them. Furthermore, there is at least one item—a paddle-ball—which was designed to study how speakers deal with naming uncommon or unfamiliar items.

Both subjects were recorded retelling the Pear Story four times—two monologues and two dialogues. Each case—monologue and dialogue—was done twice in order to account for practice effects and to examine similarities between retellings in the same context. The actual recordings were spaced fairly widely, with the first monologue and dialogue being recorded on one day, and the second monologue and second dialogue being recorded on a different day, a few weeks apart.

In the monologue cases, the subject retold the story alone in a quiet office, i.e., with no interlocutor present. It should be noted that in some research, the term *monologue* is used to refer essentially to periods in a conversation where one speaker holds the floor for a substantial length of time. In the present study, *narrative* (following Chafe, 1994) is used to describe this type of conversation, and *monologue* is used to mean that there is no interlocutor. The monologic task is not realistic in terms of everyday speech, but it is not uncommon in language testing (such as the Versant test mentioned above). This was done to remove aspects of the speakers' language production that are not co-constructed, and therefore independent of other speakers.

In the dialogue cases, the subject retold the story to a native speaker of English, in the same room where the monologues were recorded. The story retellings followed approximately 15 minutes of free conversation, so that the learner could become comfortable speaking with the interlocutor. A different native speaker was involved for each retelling in order to ensure that each retelling was the same type of authentic communicative event. The four interlocutors were English teachers in their 30's, teaching at the same university that the subjects attended, but were unknown to the subjects. One was an American male, one was a Canadian female, and the other two were American females. For this study, native speakers of North American English were chosen as the interlocutors in order to minimize differences in the English ability of the participants. Also, these varieties of English were familiar to the subjects. However, it is impossible to truly control for individual differences between interlocutors. This was taken into consideration when drawing conclusions from the discourse analyses in Chapter 6 and 7.

The dialogues were audio-recorded as above, and also video-recorded using two video cameras—one showing the front view of each participant. Video recording was done to enable analysis of head nods and gestures, which have been shown to be important in conversation and in fluency (Adolphs & Carter, 2013; Bavelas, 2000; Schegloff, 1984), and which are often used instead of verbal backchannels (Bavelas et al., 2000), as well as gaze.

All the retellings were recorded as wav files with an Edirol/Roland R-09 recording device in a quiet office. This allowed for high quality recordings to be made, such that even breaths were audible on the recording. Video recordings were made with two Victor Everio hard disc camcorders. The researcher was not present during any of the recordings.

All recordings were transcribed with *ELAN* (*ELAN — language archiving technology*, 2012; Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006), which enables annotations to be time-aligned to the audio waveform (Figure 4.3), and allows video files to be synchronized with the audio. ELAN also allows multiple tiers of transcription for such things as different speakers, and separate tiers for utterances and nonverbals. This facilitates the analysis of events that overlap or occur simultaneously, and allows accurate time measurements to be made in milliseconds. *Praat* (Boersma & Weenink, 2012) was used in conjunction[12] with ELAN to accurately measure the length of utterances and pauses, and to analyze prosodic features.

The first stage in the calculation of temporal variables was to determine the durations of the segments of speech and silence (unfilled pauses). For the two dialogues, the portion of the story that encompassed the main narration was

[12] Praat can be accessed directly through ELAN using the Sendpraat script available from the Praat website, cited above.

extracted from the pre-narrative and post-narrative segments to make the dialogues more comparable with the monologues. For Asami's Dialogue 1 this was from time 00:31 to 04:36 (min:sec), and for Asami's Dialogue 2 this was from time 00:24 until time 02:48. For Keiko's Dialogue 1, this was from the beginning of the audio file until time 03:54. For her Dialogue 2, the entire length of the audio file was used[13].

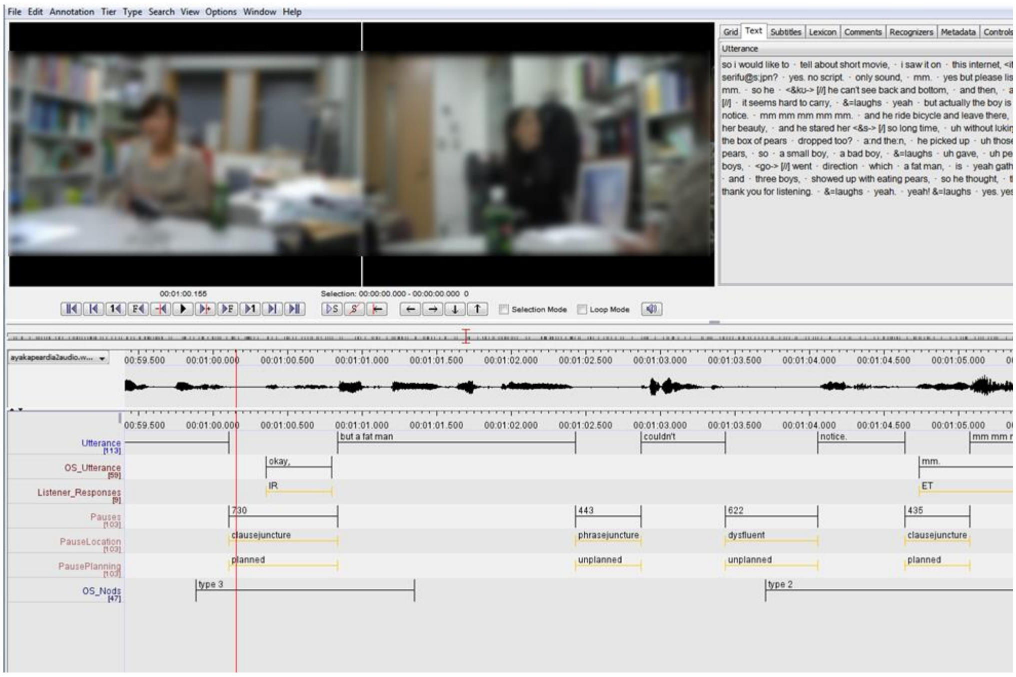

Figure 4.3 Screenshot of ELAN. In the top left corner are the two videos showing each speaker. In the middle is the audio waveform. Below the waveform are the tiers of annotation.

4.3.2 Units of transcription

The first step in transcription involved dividing the stream of speech into some kind of unit. There were several possibilities, each with their own advantages and disadvantages. In this study, the first level of transcription was done using *utterances*, defined as continuous runs of speech between silent pauses. Different

[13] The audio files (and accompanying video files) were cut from a single audio recording encompassing all tasks in a given session. That is, the session involved about 10-15 minutes of free conversation, followed by the retelling of the Pear Story. The individual segments were cut from this to facilitate transcription in ELAN.

researchers have used different minimum durations to define an unfilled pause, from lengths as short as 100 msec (Griffiths, 1991) to lengths up to 1 second (Götz, 2013), but values between 100 and 300 msec are the norm in most fluency research (Towell, 2002; Towell et al., 1996). Riggenbach (1991) classified pauses of less than 200 msec as *micropauses* rather than as *hesitation pauses,* suggesting that micropauses were a frequent and normal aspect of fluent speech, and therefore not hesitation phenomena.

Traditionally, the minimum pause duration has been set at 250 msec (Goldman-Eisler, 1967), but recently this threshold has been challenged (Campione & Véronis, 2002; Kirsner, Dunn, Hird, Parkin, & Clark, 2002) with the argument that current technology allows us to record speech in higher fidelity and to accurately measure pauses to hundredths of seconds. Kirsner et al. (2002) show that if no minimum threshold is used for pauses, so that all silences are measured, then the distribution of the log-duration of the pauses tends to be bimodal, with one mode at a very short duration (such as 100 msec) and another at a higher duration (such as 500 msec). The boundary between these (which might, for example, be 175 msec) should be used as the threshold for separating longer hesitation pauses from shorter pauses which are generally the articulatory silences that occur with phonemic voiceless stops.

The value of the threshold is important since many studies use automatic segmentation of pauses and speech with computer software, and setting a different value for this parameter will give different results (Kirsner et al., 2002). This problem was avoided in the current study by segmenting the pauses manually. In lieu of a minimum threshold, all silences were evaluated individually to determine whether they were due to articulation or could be classified as hesitation. Marginal cases (generally pauses less than 200 msec) were examined in Praat (Boersma & Weenink, 2012), and a decision was made based on the acoustic and linguistic environment. For example, if the pause was followed by a plosive consonant, I did not annotate it as a hesitation unless it was longer than approximately 100 msec, which seemed perceptually to work well in separating articulation from hesitation. There were hesitations as short as 100 msec that were accompanied by a short breath (audible from the recording), which is clearly a sign of an intended break in the speech. Oehmen, Kirsner, and Fay (2010), in a study of manual segmentation of pauses, found that intra-rater reliability was very high, while inter-rater reliability decreased with increased noise in the audio recording. Therefore, in the present study, since all the rating was done manually by the researcher, comparisons between the tasks and participants can be considered to be reliable. The distributions of pauses in the present study were also compared with the results of Krisner et al. (2002) to determine whether the log-transformed distributions of pause lengths were normally distributed. This would be expected from their results, if I successfully avoided articulatory pauses.

Other possibilities for units of transcription besides *utterances* include *turn construction units, intonation units, basic discourse units,* and *C-units*. *Turn construction units (TCP's)* are defined as segments of speech that end in a *turn relevance point (TRP)*, where a speaker's turn could potentially end (Sacks et al., 1974). TCP's are defined by a theory of conversation based on turns, where speakers generally speak one at a time, avoiding overlap and gaps between turns (Sacks et al., 1974). For story retelling tasks, turns are not a very meaningful unit because the floor is generally held by the speaker telling the story for the entire duration. However, the concept of TRP's was used in the discourse analysis of Chapter 6 and 7, particularly when looking at listener responses.

While turns are a higher level structure related to the organization of conversation between speakers, *intonation units* are implied by Chafe to be related to the capacity of working memory for speech, and are therefore a unit that should be more suitable for looking at cognitive fluency. The term *intonation unit* comes from Chafe (1994) and is described as a basic functional segment of discourse that can be identified by convergence of several features. These features include: (1) the presence of pauses before and after the intonation unit, (2) acceleration of syllables at the beginning and deceleration of syllables at the end, (3) an overall decline in the pitch of stressed syllables, and (4) a terminal pitch contour. There are six terminal pitch contours, usually termed *boundary tones*. They include high-rising, low-rising, partially falling, plateau, low, and no pitch (cut-off) boundary tones (Wennerstrom, 2001). Chafe also notes that intonation units contain *prominences*, which are stressed or accented words in the utterance, identified by pitch that stands out from the base pitch of the utterance (usually higher pitch), and increased volume. Chafe identifies three types of intonation units: *fragmentary* (incomplete), *substantive* (containing information about events, states, or referents), and *regulatory* (regulating the interaction or flow of the conversation). Intonation units (substantive) often have the form of a single clause, but more often contain one new idea, leading Chafe to propose the “one new idea constraint”, and arguing that a single intonation unit (particularly the substantive ones) corresponds to a single focus of consciousness (Chafe, 1994).

Another alternative unit is the *basic discourse unit* (Degand & Simon, 2009), which are defined by points were the prosodic boundary and the syntactic boundaries coincide. Degand and Simon argue that these points are perceived as complete by the listener, and therefore are points where the listener can more easily draw inferences and create coherence or relevance. The BDU is therefore a unit as perceived by the listener, rather than a cognitive unit, based on the speaker, like the intonation unit.

In research with spoken corpora, Biber et al. (1999) found that although the grammar of spoken and written language are basically the same, the sentence, which is the basic unit of written language, cannot be used as a basic unit of structure in spoken conversation. They use the *C-unit* (meaning “communication unit”), originally proposed by Loban (1963), which is an umbrella term encompassing both *clausal units* and *non-clausal units*. *Clausal units* consist of an independent clause and any related dependent clauses. *Non-clausal units* are the other utterances that are not complete clauses, including clause fragments that do not at least consist of a subject and finite verb. A clausal unit may consist of several clauses, and this tends to be achieved with the “add-on strategy” (Biber et al., 1999, p. 1068), which is what Pawley and Syder (2000, p. 177) refer to as “clause chaining”. This is where finite clauses are strung together linearly, rather than embedding. Another common feature is syntactic blends, or *anacoluthon*, where phrases are completed with grammatically inconsistent language, due probably to limitations of working memory, such as in the example, *About a hundred, two hundred years ago* ***we had*** *ninety-five per cent of people- i- in this country* ***were employed*** *in farming* (Biber et al., 1999, p. 1064).

In operationalizing C-units in the present study, independent clauses that begin with coordinating conjunctions (such as *and* or *but*) are annotated as new C-units, since it is very common in spoken English for turns to begin with these (Biber et al., 1999, p. 1070). Incomplete clauses that contain a finite verb in the main clause are classified as clausal C-units, following Biber et al. (see Note 10 in Biber et al., 1999, p. 1137). In the present study, C-units were mainly used in the analysis and categorization of pauses.

Initially, I intended to transcribe the data in this study with intonation units, however, identifying them in L2 speech proved to be problematic because learners do not necessarily use prosodic boundary tones in the same way that native speakers do. Therefore, I used the utterance, as defined above, as the basic unit of transcription because this unit was the most theory-free. This was deemed most appropriate for an investigative study using a grounded theory methodology, where I could examine the data without implying certain cognitive processes (in the case of intonation units) or certain perceptions of the speakers (as in the case of TCU's and BDU's).

In all cases of annotation and categorization, decisions must be made about difficult or borderline cases at the time, in order to facilitate the following analyses. As this was done, notes about the decisions made were kept in order to facilitate later interpretation and produce an "audit trail" of the annotation process (Lapadat, 2000, p. 216).

4.3.3 Annotation conventions

There are many conventions available for annotating spoken language, including the typical systems used in Conversation Analysis (CA), and variations on this such as that used by Riggenbach (1991). Since ELAN was used for transcription and data analysis in this study, it was not necessary to annotate features such as overlap, which are visible in the software and can be measured or observed directly through the ELAN. However, the transcripts printed in the Appendices and examples given in the text included annotations similar to the standards of CA, in order to make transcripts that were readable as text. The transcription system used in these case studies was based on the system for CHAT (MacWhinney, 2014), including the transcription conventions it allows for CA. This allowed text and annotations for fluency features that were easily searchable by computer, and allowed the possibility of using the analyses available in the CLAN software package (MacWhinney, 2013). One modification, however, was that I used a system for transcribing boundary tones that was based on the system used by Auer, Couper-Kuhlen, and Müller (1999), using punctuation marks rather than the arrows used in prosody research (Wennerstrom, 2001). The symbols used in the present study are outlined in Table 4.1. Upper case letters were only used for the first letters of proper names. Numbers were spelled out. Words and phrases spoken in Japanese were spelled out in the Latin alphabet and marked with the symbol @s:jpn, as in *nan@s:jpn darou@s:jpn.* Incorrect words were annotated with the correct substitute afterwards in square brackets, as in *the box of paired [: pears].* Incomplete words or reduced forms were transcribed with the missing sounds in parentheses, as in *could (h)ave* or *jus(t).* Normal contractions such as *I'm, it's,* or *don't,* were transcribed as normally written, i.e., as single contracted forms.

Table 4.1 Transcription symbols

?	high-rising boundary tone	:	lengthening of a vowel
,	low-rising boundary tone, or similar prosodic tone that implies continuation	xx	unintelligible words
none	plateau boundary tone (no punctuation at the end of an utterance)	@u	phonological transcriptions (particularly for vowels added to the end of words, when that ending is prolonged or noticeable)
;	partially falling	@s:jpn	Japanese words
.	falling	&=laughs	laughter
-	cut-off intonation contour	[=! laughing]	laughter within a word
(325)	silent pauses in msec	[=! text]	paralinguistic annotations (e.g. audible inhalations, but only when they serve a communicative purpose)
&	phonological fragment, such as &t, or phonological transcription of an unrecognizable word (these are not counted as words)		
[/]	repetition	[///]	reformulation
[//]	self-correction	[/-]	false start
*	timing of gesture annotations	2====+	head nods (number indicates type, lenth indicates duration)

In representations of the data in the text, line breaks were made after each utterance, with silent pauses having their own line (Extract 4.1). Since square brackets are used for many annotations in the CHAT system (see Table 4.1), overlap was shown in transcriptions using right-angle brackets as in the following extract from Keiko's Pear Story Dialogue 1 (lines 083-084).

Extract 4.1

```
074   KEIKO with rope?
075         (818)
076         an(d) pass past the tree.
077         (780)
078   NS3   okay,
079         (95)
080   KEIKO &=laughs
081   NS3   wait a different man or the same man.
082   KEIKO oh different man.
083   NS3   a different ┌man okay.  ┐
084   KEIKO             └yeah differ┘ent man.
```

4.3.4 Annotation tiers

Once the tier of utterances was completed (for each of the participants in the case of the dialogues), a separate tier of the resulting pauses was created. This allowed the annotations and their associated time stamps and durations to be exported for further analyses with spreadsheet software, text editors, and R (R Development Core Team, 2012).

4.4 Monologic features of fluency

For all four retellings by each of the two subjects, the traditional measurements of fluency-related features were made, in order to compare the monologues with the dialogues, and Asami with Keiko. This section outlines those variables and how they were defined and measured.

4.4.1 Temporal variables

4.4.1.1 Length of run

One of the most commonly used measurements of fluency is *mean length of run* (MLR), which is defined as the mean number of words per uninterrupted run of speech. An uninterrupted run of speech is what I have defined above as an *utterance*, which is a run of speech between silent pauses. Words that were not uttered completely were not included in the calculation, and laughter was also not counted. Filled pauses (*uh* and *um*) were counted as words, following Riggenbach (1991) and others (Rühlemann, Bagoutdinov, & O'Donnell, 2011).

4.4.1.2 Articulation rate

Articulation rate reflects the actual speed of the speaker's articulatory processes, rather than any reflection of the speed of information communicated. Articulation rate is generally measured in syllables per second, and is calculated as the number of syllables in an utterance divided by the duration of the utterance.

Generally, this is calculated over the entire range of the speech of the subject (i.e. not including silent pauses), giving an average articulation rate for the entire text, but this can also be calculated per utterance. In this study both methods were used, so that the variations in articulation rate could also be analyzed in addition to the overall average rate. Native speakers of Japanese have a tendency to add syllables to some words, particularly by inserting vowels at the end of words or to break up consonant clusters. In calculations involving counting syllables, the actual number of syllables pronounced was counted rather than the number of syllables that a dictionary entry for the given word would show. Partial words that contained an identifiable vowel were counted as syllables.

4.4.1.3 Speech rate

One of the most common temporal variables in fluency research is *speech rate*. This is generally calculated as the number of syllables or words spoken divided by the duration of the speech. In this case, unlike articulation rate, unfilled pauses are included in the calculation. Therefore, speech rate reflects the speaker's speed of transmission of information, with longer pauses causing speech rate to decrease. Tauroza and Allison (1990) argue that speech rate should be calculated in syllables per minute, rather than words per minute, because the average number of syllables per word varies with different genres. In the current research, the genre, conversation, is the same for all the tasks, so it is expected that the length of words would only vary based on the particular characteristics of the learners. Therefore, words per minute was used.

In order to facilitate comparison between the monologue and dialogue retellings, minimal listener responses were not counted as turns in analyses of monologic features of fluency. Minimal listener responses tend to fill the spaces that would otherwise be left silent in monologues, thus if these spaces were not counted as pauses for the speaker, then the result would be that dialogues would have an artificially high phonation-time ratio compared to monologues. Backchannels were sometimes accomplished with non-verbal signals such as head nods, which would further complicate the calculations if they were treated as turns. In cases in the retellings where there was a turn exchange, this data was excluded from the calculation to facilitate comparison with the monologue retellings.

4.4.1.4 Phonation-time ratio

Another common temporal variable is *phonation-time ratio*. This is simply the ratio of the amount of time spent speaking to the total amount of time the speaker has the floor. In the current research, this is calculated as the sum of the durations of all the speaker's utterances divided by the sum of the durations the speaker has the floor, similar to the calculations for speech rate above.

4.4.2 Analysis of performance phenomena

Streams of speech are not only broken up by silences, but also by words such as *uh* and *um*, repetitions of words, restarts, and words or phrases that are cut off and not completed. The greatest tendency in native speaker speech is for these phenomena to occur at the beginning of a clause or other unit of conversation (Biber et al., 1999), which agrees with Pawley and Syder's (2000) "one-clause-at-a-time" hypothesis, and similarly with Chafe's (1994) "one new idea at a time" proposal—both of which propose that speech can be planned ahead for only one syntactic clause or one "focus of consciousness".

Traditionally these have been classified as dysfluencies or hesitation phenomena (for example, Goldman-Eisler, 1968; Riggenbach, 1991); however, more recently some researchers have called attention to the ways in which these phenomena contribute to natural speech and to the naturalness of speech (for example, Carroll, 2004; Rühlemann, 2006). Rühlemann (2006), in fact, refers to them as *speech management phenomena*, in order to highlight the way that they are a necessary and natural part of speech, generally not noticed by native speakers (Wingate, 1987), and part of the way that speakers are able to produce bursts of speech that are perceived as fluent. In fact, speaking without hesitation phenomena would result in speakers sounding unnatural and stiff (Götz, 2013). In the current research, these phenomena were annotated and analyzed, looking at both the ways they contribute to fluent speech and detract from it.

4.4.2.1 Unfilled pauses

As explained above in the section on utterances, unfilled pauses (silent pauses) were pauses of any duration, manually distinguished from articulatory silences. Measurements were made locating the silences in the waveform displayed in ELAN, and creating a corresponding time-aligned annotation. When necessary, the duration was verified using a spectrograph created in Praat (Boersma & Weenink, 2012). As explained above, this was done to eliminate articulatory silences from being annotated as unfilled pauses.

Many studies calculate the mean pause length (for example, Kormos & Dénes, 2004; Towell, 2002), but because there are many reasons for pauses and types of pauses, the resulting number is affected by too many variables to be very meaningful. Calculations such as phonation/time ratio and speech rate already include all the pauses in the speech segment and therefore reflect the length of the pauses in relation to the speech. Also, some studies have shown that mean pause length has no correlation with fluency ratings (Cucchiarini et al., 2002). In the present study, pause lengths were analyzed in terms of distribution, following Kirsner et al. (2002).

Mean pause lengths vary considerably based on their location (Chafe, 1994; Pawley & Syder, 2000), and location even seems to influence the perception of pause length; it has been found that within-clause pauses were noticed by listeners when they were longer than 200 msec, while between-clause pauses were not noticed unless they were more than 500 to 1000 msec in duration (Boomer & Dittman, 1962). Furthermore, as the syntax of speech becomes more complex, the threshold for noticing pauses increases—that is, listeners do not notice pauses that they expect (Butcher, 1980). In general, pauses in spoken English occur between clauses, after discourse markers and conjunctions, and before accented content words (but after the preceding adjective or determiner) (Hansson, 1998).

In the current study, pauses were categorized according to location in terms of syntactic structure, and type in terms of speech planning, as evidenced by the preceding prosodic boundary tone. For location within syntactic structure, pauses were classified as being either (1*) between clauses*, (2) *between phrases*, or (3) *phrase internal*. If pauses were located at junctures between C-units (either clausal or non-clausal), they were classified as *between clauses.* If they were located within the C-unit, but between phrasal constituents, they were classified as *between phrases*. Other pauses were classified as *phrase internal*, which meant that they broke up phrases. Phrases were defined as constituents, so that in the clause

(1) [The opposition] [was demanding] [a more representative government],

the square brackets subdivide individual phrases (Biber et al., 1999, p. 94). In cases of pauses followed by repetition, if the phrase was completed prior to the repetition and restarted from the beginning of the phrase, it was classified as *between phrases.* Classifications of these kind often present problems for the researcher, as there are often borderline or unclear cases where “the language obstinately refuses to divide itself into the categories prepared in advance for it”

(Sinclair, 2002, p. 358). To illustrate the kind of decisions made in annotating the pauses, take the following example from Asami's Pear Story Monologue 2.

Extract 4.2 Asami Pear Story Monologue 2

```
044 ASAMI (792)
045      a:nd uh
046      (988)
047      a small boy,
048      (1363)
049      u:h &fain^(741)dʊ: [: find]
050      (314)
051      a:
052      (348)
053      some box of <pears,> [//]
054      (394)
055      gathered pears,
056      (569)
057      and he:
058      (185)
059      robbed
060      (649)
061      a: box
062      (285)
063      of
064      (991)
065      pears.
```

Extract 4.2 contains eleven pauses within two clauses. The first clause (lines 046 to 056) is *and uh a small boy find a some box of pears gathered pears,* and the second clause (lines 058 to 066) is *and he robbed a box of pears.* The pauses at the beginnings of each clause are classified as *clause juncture* pauses, the ones between grammatical phrases are *phrase juncture* pauses, and the ones within phrases are classified as *phrase-internal* (see Figure 4.4 for an illustration of the phrase structure). One pause in the first clause (line 050) actually occurs in the middle of a word. Although it is in practice quite difficult to determine what exactly constitutes a "phrase", for the purposes of this study, phrases were defined syntactically to facilitate the identification of pauses that would be more or less disruptive to the structure of speech. Therefore, phrase juncture pauses were identified as pauses that did not occur with a syntactic phrase that did not contain another phrase. The three-level classification was deemed necessary because C-units, as defined above, can become very long when subordinate clauses are used. This would have resulted in an overly large number of *phrase-internal* pauses which would not normally be considered dysfluent.

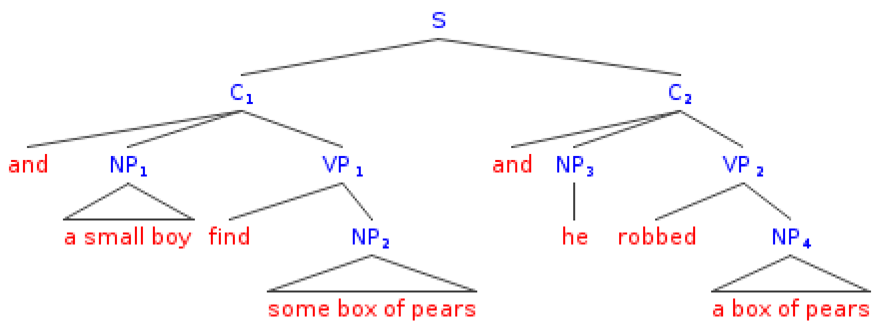

Figure 4.4 Phrase structure diagram of an example sentence from Asami Pear Story Monologue 2 consisting of several phrases and two clauses.

Furthermore, following Pawley and Syder (2000), pauses were classified as *planned* if they followed terminal or continuing prosodic boundary tones, and as *unplanned* if they did not. Some judgment was involved in classifying these. The intention of the classification, as reflected in the labels, was to separate pauses by whether they were intentional or not. The most obvious sign of an intentional pause is a prosodic boundary tone such as a falling tone or a low-rising tone. A flat intonation contour is usually a sign of an interruption in speech. However, there are some cases where it was unclear whether the pause was intentional or not. In these cases, other factors, such as the utterances surrounding the pause and the grammatical structure, were taken into consideration (see Figure 4.5 for an example).

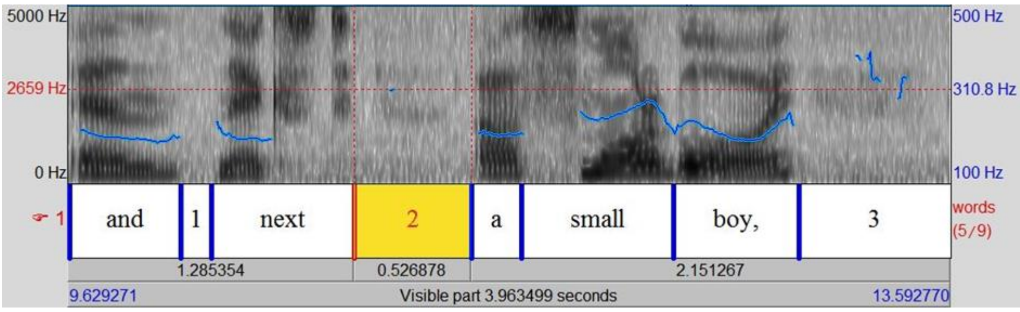

Figure 4.5 Example of using Praat to determine planned and unplanned pauses with boundary tones. The numbers (1, 2, 3) represent pauses. Pause 1 is a 150 msec silence following a plateau boundary tone, which is therefore classified as *unplanned.* Pause 3 follows a low-rise boundary tone, commonly used to imply continuation, and is therefore classified as *planned*. Pause 2 is a difficult case. It follows a plateau boundary tone, but the preceding co-text is a common discourse marking phrase, commonly followed by a silence, so this can be interpreted as a planned pause.

Ideally, the subjects could also be asked to explain the reasons for each pause or to classify them herself. However, intuition may not be reliable source of information, since the speaker does not have full access to their language production process (Lennon, 2000). Furthermore, speakers' intuitions in retrospect about their own speech may be inaccurate and reconstructed in light of the speaker's desires rather than being objective (Hopper & Drummond, 1990). In testing with my own second language retellings of the Pear Story, I found that in most cases I was not able to explain the reason for the pauses any better than the method above, particularly after the considerable time lapse between recording the conversation and having a segmented transcription. In the present study, the learners were interviewed after the conversations, and listened to the conversations with the researcher, however, in the end, this did not provide enough evidence to reliably draw inferences about the reasons for pauses.

4.4.2.2 Filled pauses

Filled pauses (realized as *uh, um,* and similar sounds[14]) are a very common device used for hesitations and speech planning in native speaker speech (Biber et

[14] In the actual learner speech data of this study, filled pauses are often realized as [ɛn] or [an], but were annotated as *uh* or *um* for ease of analysis.

al., 1999). Although they have traditionally been considered to be signs of a lack of fluency, many researchers now consider them to be normal ways of dealing with the speech planning process (Rühlemann, 2006). In this study, filled pauses were initially treated as words (see H. H. Clark & Fox Tree, 2002) and included in the calculations for speech rate and length of utterances, in line with previous research (Bosch, Oostdijk, & Boves, 2005; Campione & Véronis, 2002; Riggenbach, 1991 and others). The usage by the subjects was also examined using corpus linguistic techniques, e.g., KWIC (key word in context) lists.

4.4.2.3 Repeats

Repeats are defined as multiple iterations of a word due to hesitation, which should be distinguished from intentional repetition (Biber et al., 1999). Biber et al. (1999) found that repeats are a very common form of hesitation phenomena with a distribution very similar to filled pauses. In the current study, repeats were annotated using the CHAT system, where the speech that is repeated is surrounded with angled brackets, and marked with the symbol [/], as in *and a man, (310) is trying to (737) <pick> [/] (1070) pick up (215) some pears*. Afterwards, these were analyzed individually to determine whether they were hesitation phenomena or intentional repetitions for rhetorical effect. This was done by making an informed guess based on the video and the context of the two speakers engaged in the dialogues, as *yeah yeah yeah. it is so short movie.* In this example, *yeah* is repeated three times, but this is a commonly used device by this particular speaker and does not have any signs of being a dysfluency or even a rhetorical repetition. In the monologues, there is no video, and repetition to hold the floor or to wait for the listener’s attention (Carroll, 2004) is not appicable.

4.4.2.4 *Self-corrections and false starts*

Another common feature of spoken language are various types of incomplete or cut off words and phrases. These include *false starts*, where the speaker starts a phrase and then abandons it—often with a cut-off intonation boundary. These are annotated following the CHAT system as in the following example: *<he:> [/-] (878) yeah. (883) um (890) first, (334) &=laughs i don't know why but,* where the abandoned phrase is enclosed in angled brackets. *Retracing,* or a *retrace-and-repair sequence* (Biber et al., 1999, p. 1062) is similar, but in this case the speaker stops in the middle of saying something, and then repeats it with some kind of correction to the language, but without changing the basic message. This is annotated following CHAT as in *is coming <on the:> [//] on a bicycle.* Finally, *reformulation* is when a speaker stops mid-utterance and then rephrases it without changing the content as in *i don't know <the-> [///] where it is.* The false start is generally considered a sign of dysfluency, while the retrace can be viewed positively, as the speaker clearly recognizes a mistake and takes action to correct it. Mean length of utterance can be calculated with or without the retraced material. In this study, it was calculated initially with repeated material in temporal variable analyses, and later calculated without it for analyses of performance phenomena.

4.4.3 Formulaic sequences

The language used in the retellings was also examined to determine if more fluent runs of speech correspond to formulaic sequences for the speaker. In this case, the target of the analysis was the psycholinguistic construct of *formulaic sequence* as defined by Wray (2002), rather than recurrent sequences identified in, for example, native speaker corpora. Formulaic sequences in the psycholinguistic sense are assumed to have a direct relationship with fluency, as sequences stored whole in memory are assumed to reduce the cognitive load of language production to fluent moments of speech—the assumption being that longer runs of speech are facilitated by formulaic sequences and chaining structures sequentially (Pawley & Syder, 2000).

It is impossible to identify which sequences are formulaic for a given speaker with certainty in such a small corpus, however, the fact that the same story was retold four times enabled side-by-side comparison that could lend some evidence toward identifying language that is formulaic for the particular speaker. In this study, analysis of formulaic sequences was performed by comparing the four retellings of the stories in parallel across a single learner, and looking for sequences that recur across the multiple retellings.

4.4.4 Grammatical structure and complexity

In the current study, as described above, speech was initially segmented in terms of utterances, which are defined purely acoustically, by the presence of silent pauses. Then, the speech was segmented into C-units (classified as *clausal* or *non-clausal*) in order to examine the clausal structure and complexity of the subject's speech. Structure and complexity were not quantified in this study, but were considered when looking at particular cases that showed signs of higher or lower fluency within a given conversation.

4.4.5 Summary

In summary, the first part of the analysis of these conversations involves looking at features of fluency that are typically associated with monologic fluency—that is, cognitive fluency. These include the temporal variables of utterance fluency: length of uninterrupted run of speech, articulation rate and speech rate, the location and duration of unfilled and filled pauses, and the presence of repetitions, reformulations, and false starts. Additionally, other factors related to cognitive fluency are the use of formulaic sequences, the grammatical structure and complexity of the speech, and the use of compensatory strategies. The results of these analyses are presented in Chapter 5. The next section will detail the methodology for looking at aspects of co-constructed fluency.

4.5 Co-constructed and dialogic aspects of fluency

4.5.1 Turn boundaries

Turn boundaries are important points in conversation—points where speakers seem to try to minimize gaps and avoid overlap (Stivers et al., 2009). In order to examine the confluence of the dialogues, which is facilitated by smooth turn boundaries, temporal analyses were also carried out over each retelling as a whole. In these analyses, although I continue to the use the term *turn boundaries,* listener responses were included as turns, as these have important contributions to the overall smooth flow of conversation. Comparisons within the two retellings by a given speaker and comparisons between the two speakers were made in order to shed light on differences in the confluence attained, and what factors contribute to this. Furthermore, particular points of large gaps or overlap were looked at individually for signs of problems that could result in breakdowns of fluency or unpredictable turn boundaries.

4.5.2 Smallwords

Discourse markers have gone by a number of names, including *fillers, initiators, pragmatic particles, response signals, discourse particles, phatic connectives,* and *pragmatic markers,* just to name a few (Brinton, 1990; Fraser, 1999), and there is widespread disagreement about what exactly constitutes a discourse marker, and how they function in language (Fox Tree, 2010). However, most researchers agree on some general characteristics. Discourse markers most typically occur at the beginning of turns and utterances (Fung & Carter, 2007), although they do also occur in medial and final positions. They function to control discourse and refer to the language itself (reflexivity), and can point to other items in the discourse and situational dimensions (indexicality) (Aijmer & Simon-Vandenbergen, 2004). They tend to be short and occur outside of syntactic structures (Brinton, 1990). They also occur with particular prosody (Fung & Carter, 2007) and are optional (Del Saz Rubio, 2007). Discourse markers are argued to contribute to fluency in that they compose constraints on implicatures, helping listeners to construct relevant interpretations of the speaker's utterances (Blakemore, 1987, 1992). They also contribute to temporal fluency by being short and easy to produce—filling gaps and thus decreasing pauses, and contribute to perceived fluency by virtue of their high frequency and familiarity to listeners (Hasselgreen, 2005).

Similar to discourse markers are Hasselgreen's *smallwords* (Hasselgreen, 2005). These are defined as "small words and phrases, occurring with high frequency in the spoken language, that help to keep our speech flowing, yet do not contribute essentially to the message itself" (Hasselgreen, 2005, p. 162). In particular, Hasselgreen analyzed the use of 19 expressions and sets of expressions: *well, right, all right, okay, oh, ah, you know, I think, I mean, like, sort of/kind of, a bit, just, or something, not really, and everything/that/stuff/things, I know, you see,* and *I see* (Hasselgreen, 2005, p. 163), and found that more fluent learners used these expressions in more nativelike ways.

In the present research, I examined the use of *smallwords* because this does not necessarily entail that the speaker used the particular expression to function as a discourse marker. It is possible that learners and native speaker interlocutors may use these markers for different reasons (Aijmer, 2004). Therefore, in this study, I included as a smallword any short words or expressions that appear to be functioning as a smallword or discourse marker, by looking at the position, prosody, and context of each particular expression, keeping in mind that particular expressions might have been used differently from how a native speaker would likely use it. This was then compared across story retellings and conversations, and between the two subjects of the case studies.

4.5.3 Alignment of rhythm and tempo

Another stage of the analysis of the data is to look at the rhythm of the speech, created by the timing of stressed words, in each retelling. This will also contribute to an analysis of turn boundaries, which are considered to be key points in conversation (Sacks et al., 1974) and important elements for creating dialogic fluency (McCarthy, 2009). It has been argued that conversation feels more fluent for the speakers when the rhythm and tempo of the speakers is aligned (Fiksdal, 1990, 2000). *Tempo,* or *pacing,* is defined as the rate of stressed words per minute, and *spacing* is the ratio of stressed words to total words (Vanderplank, 1993). Pacing and spacing for an individual speaker have been shown to be effective measures of fluency (Kang, Rubin, & Pickering, 2010), and this information was also looked at in analyses of fluency in the current study[15]. However, for the purposes of looking for signs of interactive alignment, the analyses involved looking at how the pacing of the conversation was established and carried over successive utterances or turns.

In some previous studies, *pace* seems to be a simple measurement of the number of stressed syllables per unit (Kang et al., 2010; Vanderplank, 1993), however, in studies that look more closely at rhythm in speech, a particular tempo

---

[15] Note that Kang et al. (2010) defined *pacing* as the number of stressed syllable per run, rather than per unit of time, in their study.

or pace implies that the stressed syllables come at equal intervals (Auer et al., 1999; Couper-Kuhlen, 1993; Fiksdal, 1990, 2000). The approach used in the present study was based on the methodology of Auer et al. (1999), described below, who argue that rhythm in conversation is a perceptual gestalt by the listener (and speaker), and not entirely predictable from the acoustic signal alone. Rhythm is created by prosodic prominences which are created by lengthened vowels, pitch accents, and changes in intensity (volume). These acoustic phenomena contribute to the perception of rhythm, but (1) just because a syllable is stressed in some way, does not entail that it will be perceived as stressed within the rhythmic pattern, and (2) it is possible for the perceived rhythmic pattern to be maintained over silences where there is no stress—or no syllable, for that matter—but where the beat is still perceived. Initially, at least three beats at equal intervals are necessary to establish a rhythmic pattern, and the fourth or any subsequent beats can be silent, although generally more than three silent beats in a row will cause the rhythmic pattern to be lost. Double-timing, syncopations and other polyrhythmic characteristics can occur without breaking up the established rhythmic pattern, although these (silent beats and syncopated beats) need to be evaluated individually to determine whether the pattern has been broken or not. Therefore, in the methodology of Auer et al. (1999), the rhythmic patterns are determined by the perception of the researcher, i.e., by listening to the recording and determining when gestalt rhythmic patterns are set up.

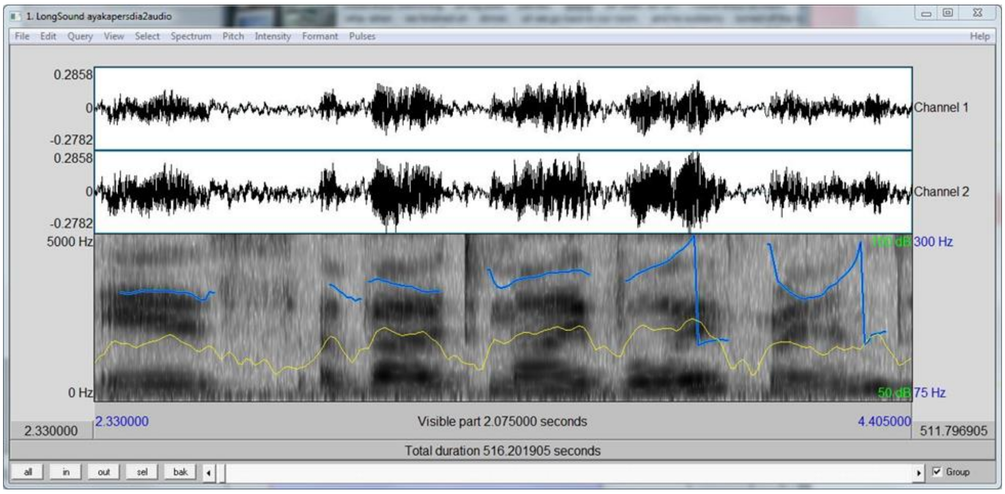

Figure 4.6 Screenshot of Praat software used for rhythmic analysis. The blue line superimposed over the spectrogram shows pitch, and the yellow line shows intensity.

In the present study, rhythm was evaluated as above, by listening to the audio recordings, as well as consulting the acoustic data as displayed in Praat (Boersma & Weenink, 2012), which shows the waveform, the intensity curve, the F_0 pitch curve, and the spectrogram (see Figure 4.6). This information was consulted to assist in measuring the tempo where rhythmic patterns were perceived, and determining whether rhythmic patterns existed or not. Although it may seem overly subjective to examine rhythmic patterns by perception of the researcher, there are several reasons that justify this. One is that rhythm is something that is perceived by listeners in the conversation as well, although probably not consciously. Second, it is not possible to accurately measure the interval between beats. The measurement could be taken from the beginning or mid-point of the vowel, or at the onset of the consonant preceding the vowel, or at the high point of the pitch or intensity curve for the stressed syllable, but none of these criteria results in accurate predictions of the perception of rhythm by listeners (Auer et al., 1999)[16]. However, there are some

[16] The discrepancy between the acoustic data and perception is not unique to rhythm. As stated elsewhere in this chapter, pauses can be perceived as longer or shorter

other facts about the acoustic data that can help to determine rhythmic patterns. Auer et al. (1999) found that rhythmic intervals less than 0.23 seconds and greater than 1.2 seconds were not perceived as rhythmic patterns. Also, Couper-Kuhlen (1993) found that the "just noticeable difference" between intervals—that is, whether an interval is perceived as the same length as the preceding interval or not—can be up to a 30-37% difference. In the present study, the tempo was determined using a metronome. Additionally, other native speakers were asked to listen for rhythmic patterns in segments of the data to corroborate the researcher's findings.

4.5.4 Interactive competence

In addition to the linguistic and temporal features of language discussed above, the participants' personalities, friendliness, and willingness to contribute to the conversations also may have an effect on the fluency of the conversation. For example, the rapport between the speakers and the ability of the speakers to scaffold each other to contribute to the confluence of the conversation are thought to be important factors in fluent conversation.

4.5.4.1 Personality and rapport

It has been claimed that conversation feels more fluent when there is rapport between speakers (Fiksdal, 2000), and that factors such as likeability affect convergence of temporal features (Schweitzer & Lewandowski, 2013). In the present study, measures such as *likeability* or *friendliness* were not quantified, however, evidence from the video and audio recordings of the conversations, as well as notes from the post-conversation interviews with participants were used to get insight

---

depending on their locations, and regardless of their actual length. Pauses can even be not perceived at all in some cases, or perceived where there are none (Butcher, 1980).

into the personalities of the speakers and the existence of rapport between the speakers.

4.5.4.2 Evidence of scaffolding and confluence

In addition to the features of interactive alignment and turn boundaries discussed above, there are other things that speakers do that contribute to the smooth flow of conversation, or confluence (McCarthy, 2005). For example, speakers may scaffold each other by completing one another's sentences, or jumping in to fill long gaps where the conversation is faltering. These aspects of the conversations were not quantified, but were looked at qualitatively throughout the data.

4.5.5 Nonverbal features

Nonverbal aspects of language have been shown to improve productive fluency (Bavelas, 2000; Finlayson, Forrest, Lickley, & MacKenzie Beck, 2003) as well as perceived fluency (McCafferty, 2006; Nambiar & Goon, 1993; Puccinelli, 2010). This includes features such as head nods or facial expressions which can act in lieu of verbal communication, and speech-preparatory gestures and gestures that complement speech. These actions may contribute to the co-constructed fluency of the conversation in several ways. Nonverbal backchannels may substitute for verbal backchannels (Bavelas et al., 2000). Nonverbal feedback from the listener may contribute to the construction of relevance, thereby making the storytelling task easier for the speaker. Furthermore, gestures and facial expressions by the speaker can assist communication by adding non-redundant information to the content of the story (Bavelas, 2000), and, because they are faster than speech, can help the interlocutor to predict upcoming speech (Kendon, 2004). On the other hand, overuse can contribute to a negative impression of fluency (Gullberg, 1998; Stam, 2007).

Head nods were annotated on a separate tier in ELAN for the listeners. They were time-aligned with the audio waveform and categorized using the framework of Knight and Adolphs (2008). In this framework, there are five categories based on the intensity and number of the nods. Small head nods can occur as a single nod (Type 1) or as multiple nods (Type 2). More intense head nods can also occur as a single nod (Type 3) or as multiple nods (Type 4). Type 5 nods are a combination of small and intense head nods.

For the scope of this study a detailed transcription system such as the Facial Action Coding System (Ekman & Friesen, 1978) was not deemed necessary. Therefore, in the current study, annotations for gestures, facial expressions, and gaze were comprised of a short description of the action placed in line with the transcription, and labeled with an asterisk to annotate the timing of the action. When necessary during the analysis, it was always possible to refer back to the video recording itself to look more closely at the particular action.

4.5.6 Summary

In summary, the features of dialogic or co-constructed fluency include smooth turn boundaries between speakers (avoiding overlap and minimizing silence), rhythmic alignment between the speakers, the use of backchanneling and nonverbal actions to facilitate communication, and interactive competence. Some other factors, such as accuracy, accent, and lexical diversity will not be quantified, but will be looked at within the context of particular examples in the analysis.

4.6 Analyses

The present study used a mixed-methods approach to multiple case studies. As a qualitative case study, the conversations were looked at individually and in detail for evidence of factors of fluency—both monologic and co-constructed—that contribute to more or less confluent sections of the conversation, using a methodology informed by Conversation Analysis. At the same time, quantitative methods were also used to compare conversations to look for differences (1) within recordings by the same learner (particularly in terms of monologue vs dialogue) and (2) between learners. When possible, these were analyzed graphically and/or statistically using the R software package (R Core Team, 2015) and the ggplot2 add-on package (Wickham, 2009). Transforming the data into various graphical forms enables the researcher to find patterns that might be difficult or impossible to find by piecemeal analysis of individual examples in the conversations. On the other hand, close analysis of individual examples is necessary to examine the detailed interactions of factors in the fluency of the speakers and the conversation.

In the next chapter, the Pear Story retellings and free conversations will be analyzed in terms of the features traditionally associated with monologic fluency. From a traditional point of view, the difference in these features between retellings of the Pear Stories should not vary considerably between the monologue and dialogue retellings. On the other hand, there may be variation between the earlier and later retellings, since a practice effect would be expected. Furthermore, it is expected that there will be significant differences in the monologic fluency features between the two subjects. The following chapter, Chapter 6, will analyze these same conversations in terms of the features associated with dialogic or co-constructed fluency.

5 Comparing monologic features of fluency across monologue and dialogue in two case studies

5.1 Introduction

This chapter and the next present the results of parallel case studies of two learners of English, Asami and Keiko, in semi-controlled speaking tasks. Fluency research over the years has looked at a large number of variables in conversation, and many factors have been proposed to account for the perception of fluent speech, from temporal variables such as speech rate, to factors such as rapport and the use of non-verbals. Each of the two case studies involved looking at one learner of English in monologic and dialogic storytelling tasks in order to investigate the interaction of a wide range of factors that contribute to the construction of fluency in conversation. In this study, these factors were divided into two broad groups: those that have been associated with cognitive fluency (monologic features), and those that have been associated with interactive, co-constructed aspects of fluency (dialogic features). This chapter will look specifically at those monologic features. The purpose of this chapter is to identify those features in the data which have traditionally been related to cognitive fluency, and which would therefore be expected to be independent of the interlocutor[17], in order to better isolate the effects of dialogic components of fluency (to be addressed in Chapter 6).

In Chapter 3, it was found that Asami was able to contribute to a conversation with relatively high confluence, while having measureable weaknesses in utterance fluency. Therefore, in this chapter, Asami will be compared through parallel case studies with the more typically fluent, higher-proficiency learner, Keiko, in order to better identify the effects of dialogic components of fluency, and to examine how Asami's ability to co-construct confluence differs from Keiko's. One theme of this chapter is therefore to verify that Keiko's performance in terms of utterance fluency and cognitive fluency are significantly higher than Asami's, in line with expectations based on impressions and test results (see Chapter 4).

Both subjects were asked to watch a short movie, "the Pear Story" (Chafe, 1980b), and were recorded retelling the story. Each subject retold the story four times—twice in a monologue setting, with no interlocutor present, and twice in a dialogue setting, with one of four interlocutors who was a native speaker of North American English. Although some previous research has shown that dialogue is more fluent than monologue (Ejzenberg, 2000; Riggenbach, 1989; Witton-Davies, 2012), these studies did not use the same task for the monologic and dialogic components. From the traditional standpoint of cognitive fluency, it would be expected that the temporal measures of the fluency of both subjects would not vary significantly between monologue and dialogue retellings. On the other hand, it was expected that there would be a practice effect, with improvement over the course of the

[17] In this chapter and others, the term *interlocutor* is used to refer to the conversational partner, as opposed to the *speaker*, who is the subject of the case study. Of course, all participants are speakers and listeners throughout conversations, albeit to differing degrees depending on the topic.

retellings. The first and second sessions were also compared for each subject to take this into account.

5.2 Verification of segmentation

In order to accurately measure the temporal variables outlined in Chapter 3, it was first necessary to accurately segment the speech into *utterances*, which were defined as continuous runs of speech between silent pauses. Silent pauses were measured and individually evaluated by the researcher to determine whether they were actual breaks in the speech, or the result of articulation. The accuracy of this procedure was verified with Asami's Pear Story retellings. Pauses in speech tend to be positively skewed (Figure 5.1) such that the majority of pauses tend toward some short duration (limited by zero), with higher durations being widely spread and less frequent. Therefore, it is necessary to log-transform the pause durations (Kirsner et al., 2002). For all four of Asami's retellings combined, the mean[18] pause length was 489 msec (*SD* = 1.89), with a range of 95 to 3721 msec. This mean duration is only slightly longer than the mean duration of the longer pauses (453 msec) found in the spontaneous speech of native speakers of English (Campione & Véronis, 2002). The log-transformed pause durations (Figure 5.2) show a fairly normal distribution, W = 0.998, p = .07, confirmed by the linear nature of a quantile-quantile plot (Figure 5.3). Kirsner et al. (2002) show that log-transformed pause durations in native speaker speech tend to have a bimodal distribution, and argue that this is the result of two

[18] Log-transformed data cannot be summarized with arithmetic means and typical standard deviations. Therefore, the geometric mean is used here which is back-transformed from the mean of the log-transformed data. Standard deviation is calculated similarly and is multiplicative rather than additive, following Limpert et al. (2001). Therefore, the range of one SD about the mean is a high bound of 489*1.89 = 924 msec, and a low bound of 489/1.89 = 258 msec.

kinds of pauses—the “short” pauses due to articulation and “long” pauses representing hesitations. The normal distribution of Asami’s log-transformed pause durations imply that manually segmented hesitation pauses was a reliable method of avoiding articulatory pauses.

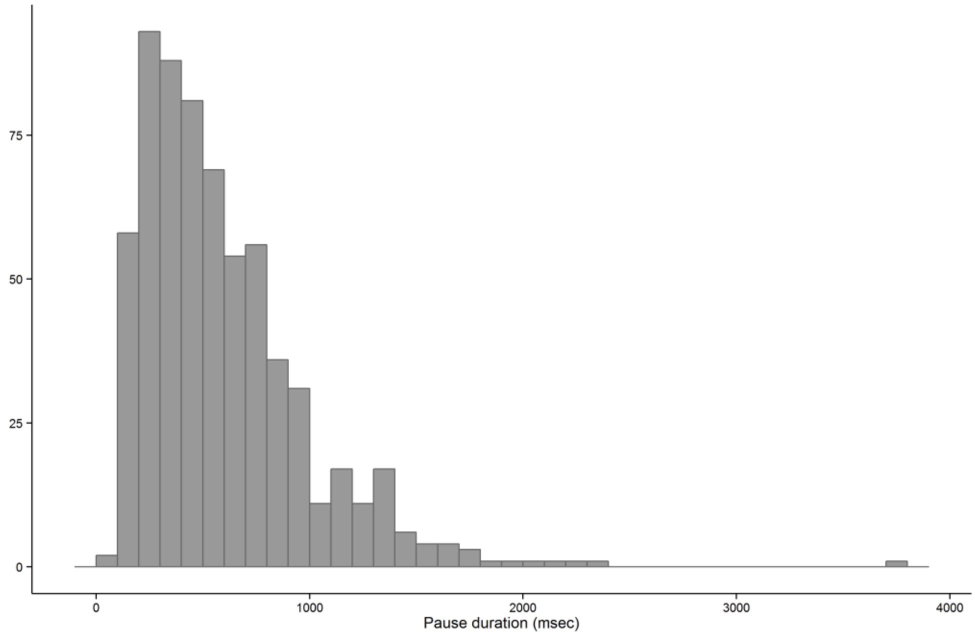

Figure 5.1 Frequency of pauses by duration from all Asami Pear Stories. Note the positively skewed distribution.

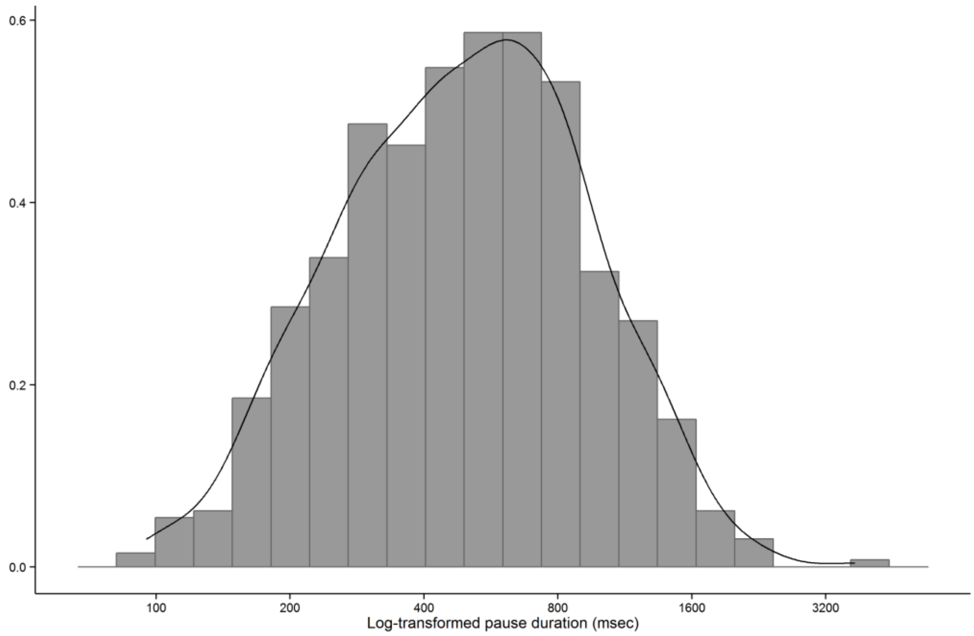

Figure 5.2 Frequency of log-transformed pause durations from all Asami Pear Stories.

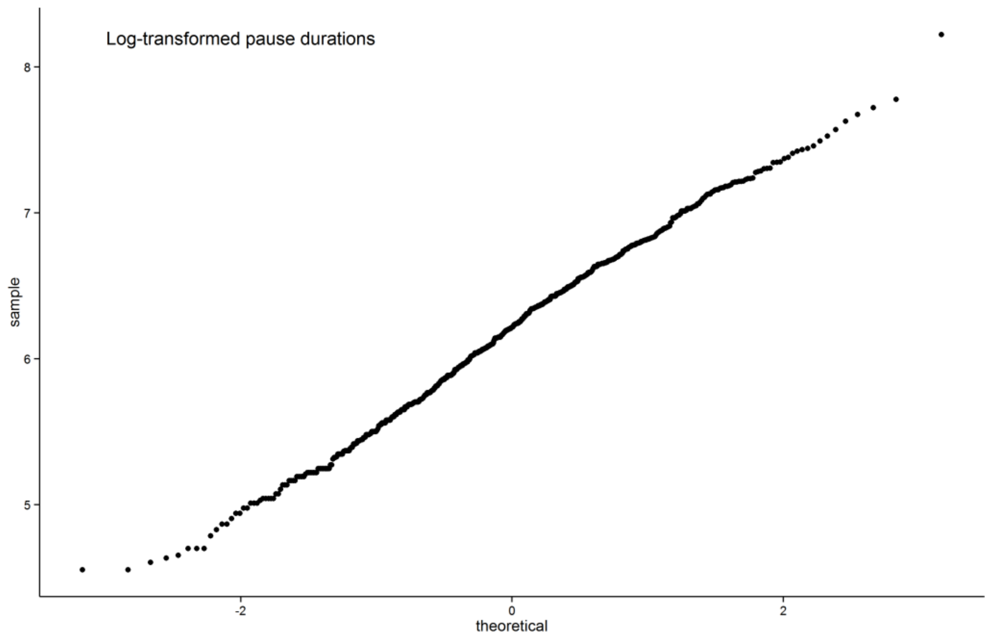

Figure 5.3 Quantile-quantile plot of log-transformed pauses durations from all Asami Pear Stories.

5.3 Length of run

The first temporal analysis was of utterance length, which is one of the most common measures of utterance fluency. In Asami's Pear Story retellings, the distributions of utterance length were clearly, and not unexpectedly, skewed toward shorter utterances (Figure 5.4). Mean length of uninterrupted run[19] was slightly lower for the monologues at 2.00 words (*SD* = 1.30) for Monologue 1 and 2.22 words (*SD* = 1.30) for Monologue 2, and higher for the dialogues at 2.23 words (*SD* = 1.44) for Dialogue 1 and 2.84 words (*SD* = 1.99) for Dialogue 2. This could suggest both a gain in mean length of run (MLR) for dialogue, as well as a practice effect (since the later retellings are higher than the earlier ones). However, since all four retellings were not normally distributed, $W > 0.767$, $p < .001$, the median is a more reliable measure for comparison. The median length of utterance was 2 words (mean absolute deviation = 1.48) for all retellings (Figure 5.5)—suggesting that length of run did not vary significantly between retellings. A Kruskal-Wallis rank sum test showed significance $H(3) = 15.7$, $p = .001$, but post hoc Wilcoxon rank sum tests, following the method of Siegel and Castellan (1988, pp. 213–214), were not significant with the exception of the comparison between Monologue 1 and Dialogue 2, suggesting an overall practice effect.

[19] *Mean length of run* is a standard temporal measurement in fluency research, and is therefore reported here. However, because the distributions are not normal, I chose to compare the data with the median and the mean absolute deviation, which are more appropriate for non-parametrically distributed data.

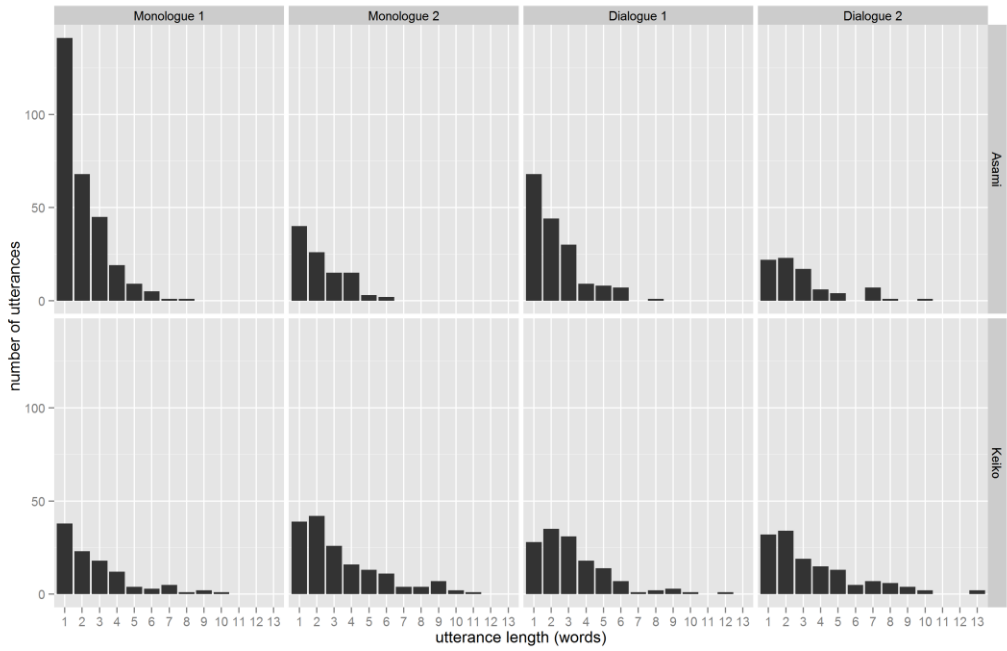

Figure 5.4 Frequency of utterances by length for both subjects and all Pear Story retellings.

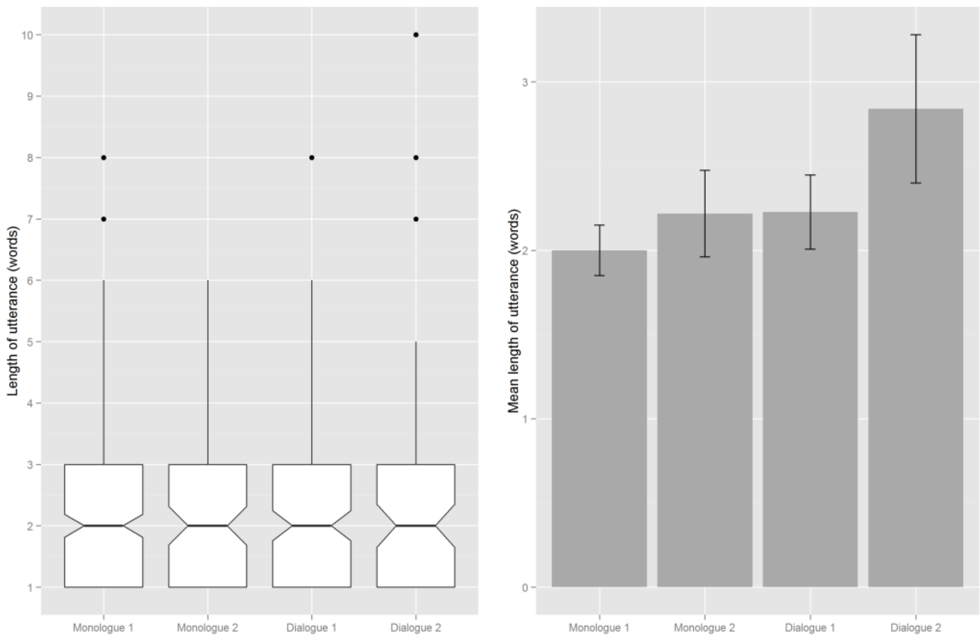

Figure 5.5 Asami's median utterance length (left) and mean length of run (right) for Pear Story retellings. Whiskers (left) extended to 1.5 times the interquartile range, with data outside this range displayed as outliers. Notches (left) and error bars (right) represent 95% confidence intervals.

There were 11 outliers for utterance length (Figure 5.5), with lengths greater than or equal to seven words. Eight of these eleven are in Dialogue 2, which again could suggest a practice effect. The longest run was of 10 words and was produced in Dialogue 2. The 10-word run was *are so kind, and they helped to pick up* *some*—two clauses chained together, each with internal pauses. The other longer runs also involved chaining two phrases (*the boy is okay but those boys are*) or repetition (*the three boy- one of the three boys*). One phrase was composed of what are possibly two formulaic sequences (*who are taking pears whole the movie*), since *taking pears* and *whole the movie* are both repeated at another point in the story retelling. Runs consisting of only one word were the most common overall, comprising 42.5% of the utterances. Out of 268 single-word utterances, most of the high frequency ones are filled pauses or smallwords, with the most common of these being *um* and *yeah* (see Table 5.1). This corresponds to previous research suggesting that lower-proficiency learners have a tendency to rely on a small range of discourse markers (Hasselgreen, 2005). Only 33 of the 268 single-word utterances involved lengthened vowels, which would be a marker of hesitation. The most commonly lengthened words were the filled pauses *um* (8 of 27) and *uh* (4 of 14), and the subject pronoun *he* (6 of 11). Again, this is expected considering that the filled pauses are hesitation markers which commonly occur in clusters with other hesitation phenomena (Riggenbach, 1991), and the subject pronoun *he* occurs at the beginning of clausal structures, where more planning would generally be required. For example, in the following extract (Extract 5.1) from Asami's Pear Story Monologue 1, *he* is drawn out as a hesitation three times while she considers how to word the following phrase (selecting between *watched* and *looked*).

Extract 5.1 Asami Pear Story Monologue 1

```
246 ASAMI he: watched
247        (445)
248        her,
249        (1140)
250        him yeah.
251        (320)
252        <he:> [/]
253        (1691)
254        u:m
```

```
255      (630)
256      yeah he:
257      (582)
258      looked her,
```

Table 5.1 Most frequent single word utterances (>3 occurrences) in Pear Story retellings for both subjects.

Asami (of 274 single-word utterances)				Keiko (of 137 single-word utterances)			
Rank	*Word*	*Freq*	*Percent*	*Rank*	*Word*	*Freq*	*Percent*
1	um	27	10.1	1	and	49	35.8
2	yeah	24	9.0	2	uh	19	13.9
3	and	21	7.8	3	he	8	5.8
4	uh	14	5.2	4	um	6	4.4
5	so	13	4.9	5	yeah	6	4.4
6	the	12	4.5	6	pears	4	2.9
7	he	11	4.1	7	then	4	2.9
8	then	9	3.4				
9	mm	8	3.0				
10	pears	8	3.0				
11	but	5	1.9				
12	is	5	1.9				

In the case of the higher-proficiency subject, Keiko, the mean length of run was higher for the later retellings at 3.33 words (SD = 2.37) for Monologue 2 and 3.53 words (SD = 2.61) for Dialogue 2, than the earlier ones 2.76 words (SD = 2.06) for Monologue 1 and 3.17 words (SD = 2.04) for Dialogue 1, which might suggest an effect due to practice, as well as a small gain in the dialogue contexts. However, there was no significant difference between the median utterance lengths[20] for the four retellings, $H(3) = 7.33$, $p = .06$ (Figure 5.6), with Monologue 1 having a median length of 2 words (*MAD* = 1.48), and the others having median lengths of 3 words (*MAD* = 1.48, except Dialogue 2 with *MAD* = 2.97).

[20] The utterance length distributions were all not normally distributed (Shapiro-Wilk tests for all four retellings $W > 0.81$, $p < .001$), as expected.

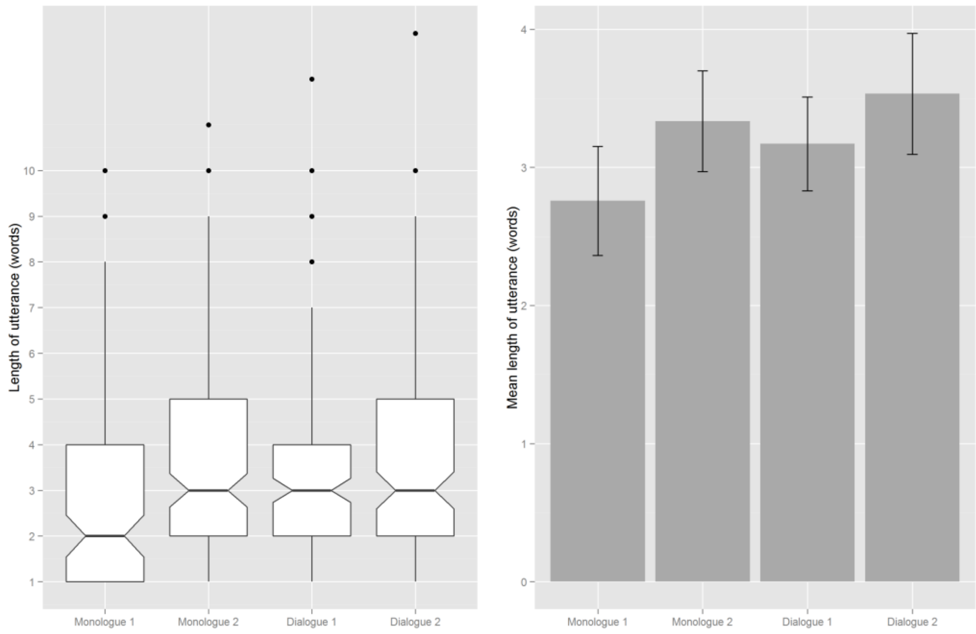

Figure 5.6 Keiko's utterance length (left) and mean length of run (right) by Pear Story retelling. Outliers (left) are greater than 1.5 times the interquartile range from the median. Notches (left) and error bars (right) represent 95% confidence intervals.

Comparing the two subjects, Keiko showed better performance in terms of length of runs (median = 3, mean absolute deviation = 1.48) than Asami (median = 2, mean absolute deviation = 1.48). This difference was significant, $W = 128466$, $p < .001$, but had a small to medium effect size, $r = -0.24$. Keiko also had a maximum length of 13 words as opposed to 10 words for Asami (Figure 5.7). As with Asami, the distribution of utterances by length tends toward shorter utterances (Figure 5.4) and does not vary significantly between retellings. The two longest runs for Keiko were *the man didn't talk to them or get angry or nothing so*, and *yeah the man uh picking up pears he didn't realize it,* which, unlike Asami, are single clauses rather than concatenations of smaller phrases. The phrase *didn't talk to them* is used three times over the four retellings, and the phrase *didn't realize it* is used four times, which suggests that they could be formulaic.

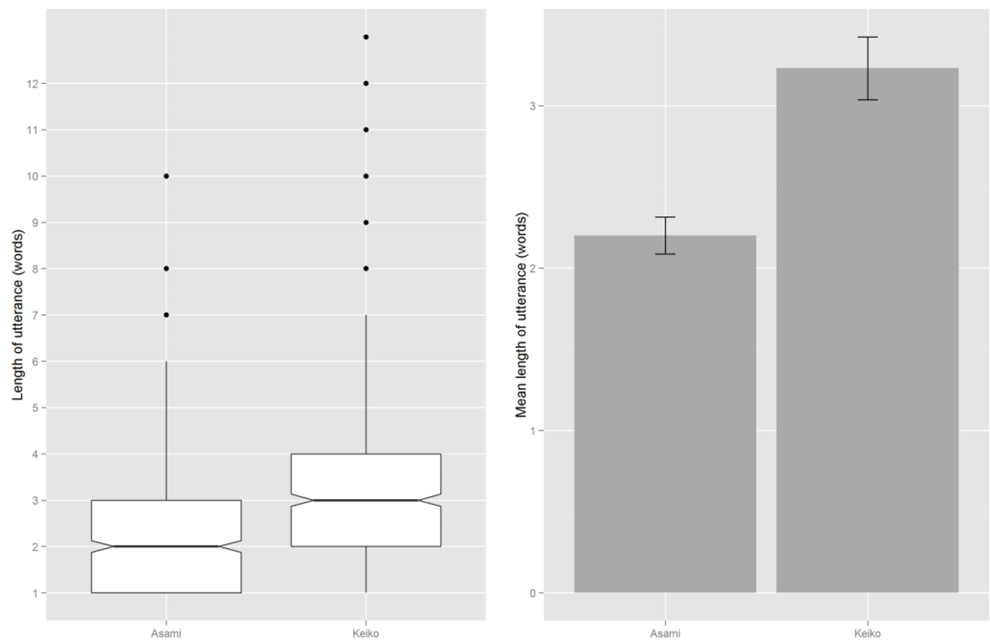

Figure 5.7 Comparison of Asami and Keiko with median utterance length (left) and mean length of run (right) for all Pear Story retellings combined. Notches (left) and error bars (right) represent 95% confidence intervals. The fact that the notches in the box plot do not overlap suggests a significant difference.

The one word utterances for Keiko are similar to those of Asami. Out of 137 single-word runs, 49 were *and,* 19 were *uh, and* 8 were *he* (Table 5.1). Similarly, many of these words were lengthened—11 of the 49 occurrences of *and*, nearly all of the occurrences of *uh* (17 of 19), *um* (5 of 6), and *he* (7 of 8). The most noticeable difference is Keiko's frequent use of *and* as a single-word utterance—making up 35.8% of the single-word utterances as compared with Asami's 7.8%, and much less frequent use of *yeah.* For both speakers, *yeah* is used as a response to the listener's reaction in the dialogues, however, Asami also frequently uses *yeah* in Monologue 1, where she seems to be reassuring herself of what she's said. The frequent use of *and* as a discourse marker is common usage in native speaker speech (Fraser, 1999) and could therefore be a mark of a higher-proficiency learner.

In summary, the length of utterances did not differ by genre (monologue vs. dialogue), but appeared to increase with practice—more so for Asami than Keiko. The higher-proficiency learner, Keiko, had significantly longer utterances than Asami.

There was some evidence that Keiko's longer runs were facilitated by the use of formulaic language.

5.4 Articulation rate

Articulation rate represents the rate of syllables per unit of time, calculated over continuous runs of speech. This can be calculated over each utterance or calculated over all the utterances combined. For all four of Asami's Pear Story retellings the distributions of articulation rate were log-normally distributed (for all retellings, $W > 0.987$, $p > .23$) and had equal variances, $F(3, 634) = 1.44$, $p = .231$. Surprisingly, the articulation rates were significantly higher in the dialogues than the monologues (Table 5.2), one-way ANOVA over the log-transformed articulation rates per utterance $F(3) = 17.073$, $p < .001$, $r = 0.27$. The significant differences were between the monologues and the dialogues, but there were no significant differences between Monologue 1 and Monologue 2, or between Dialogue 1 and Dialogue 2, with post-hoc t-tests using the Benjamini-Yekutieli correction method (Benjamini & Yekutieli, 2001). It is surprising that speakers would have higher articulation rates in dialogue settings than in monologue settings, since articulation rate is meant to measure the actual physical process of speech, as opposed to speech rate, which is more closely related to the rate of information transfer. It could be that Asami may feel pressured to speak more quickly in the dialogues since there is an interlocutor waiting for her to speak. Articulation rate does not include silent pauses, but it does include filled pauses and lengthened syllables. More frequent use of lengthened syllables in the monologues could account for these differences in articulation rate, and will be considered in more detail below in the analyses of performance phenomena.

In all four retellings, mean articulation rate was somewhat higher when single syllable utterances were not included (Table 5.2), due to the fact that single syllable utterances are often lengthened hesitations such as *u:m*. The minimum articulation rates per utterance were in almost all cases associated with single-syllable utterances. As above, one-third of the occurrences of *um* and *uh* were lengthened, and six of the ten occurrences of *he* as a single-word utterance were lengthened. Generally, in all the retellings, higher articulation rates were associated with longer utterances (Figure 5.8) up to lengths of eight or nine syllables, but the highest articulation rates, i.e., outliers, occurred mostly in utterances of between one and four syllables; the three most extreme outlier utterances were a single-syllable, cut-off *um* (123 msec in duration), and two phrases in Japanese spoken under her breath. The relationship between articulation rate and the length of the utterances appears consistent across the four retellings (Figure 5.8).

Table 5.2 Mean articulation rates for Asami Pear Story retellings.

	Including single syllable utterances			*Excluding single syllable utterances*		
	mean	*SD*	*overall*	*mean*	*SD*	*overall*
Monologue 1	2.32	1.45	2.39	2.59	1.40	2.57
Monologue 2	2.12	1.41	2.25	2.34	1.36	2.39
Dialogue 1	2.82	1.41	2.94	3.02	1.37	3.06
Dialogue 2	2.61	1.35	2.69	2.73	1.34	2.74

Overall AR is the entire number of syllables divided by the entire duration of speech, excluding silent pauses. Standard deviation (SD) is multiplicative, not additive, due to the log transformation.

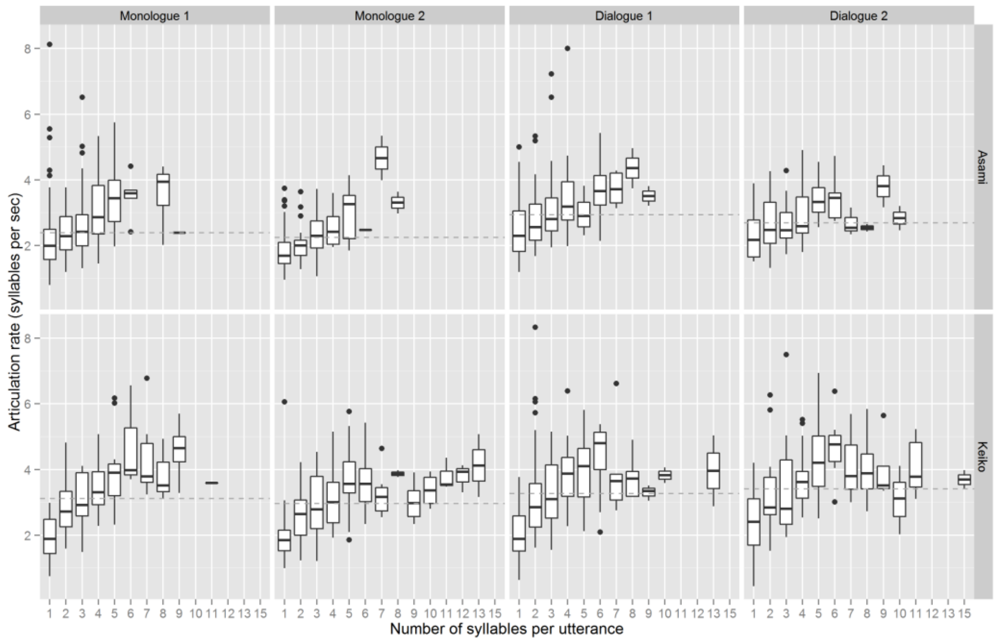

Figure 5.8 Median articulation rate by length of utterance for Pear Story retellings. The dashed horizontal lines are the overall articulation rates for each retelling.

Turning to the second case study, Keiko also showed a tendency toward a higher articulation rate in the dialogues (Figure 5.9). Three of the four retellings were not normally distributed (for Monologue 1, $W = 0.985$, $p = .31$, all others $W > 0.95$, $p < .05$), even when log-transformed, therefore, the median was used rather than the mean. Variances were not significantly different $F(3, 548) = 0.516$, $p = .67$ (Levene's test). Median articulation rate was significantly higher for the dialogues, $H(3) = 22.6$, $p < .001$, at 3.33 syllables per minute for Dialogue 1 and 3.29 syllables per minute for Dialogue 2, than the monologues, at 2.75 for Monologue 1 and 2.80 for Monologue 2 (Table 5.3). There were no significant differences between the two monologues, or between the two dialogues, which suggests that practice was not a major factor for articulation rate.

Table 5.3 Median articulation rates for Keiko Pear Story retellings.

	Including single syllable utterances			*Excluding single syllable utterances*		
	median	*MAD[1]*	*overall[2]*	*median*	*MAD[1]*	*overall[2]*
Monologue 1	2.75	1.49	3.12	3.33	1.33	3.49
Monologue 2	2.80	1.50	2.96	3.16	1.35	3.09
Dialogue 1	3.33	1.50	3.27	3.49	1.42	3.41
Dialogue 2	3.29	1.34	3.41	3.60	1.38	3.56

[1]MAD is the median absolute deviation, which is multiplicative, not additive, for log-transformed data.
[2]Overall AR is the entire number of syllables divided by the entire duration of speech, excluding silent pauses.

Keiko showed similar trends to Asami in that her articulation rate was also higher for longer utterances (Figure 5.8), and particularly lower for single-syllable utterances. Single-syllable utterances had a somewhat higher tendency to contain filled pauses and lengthened vowel hesitations (32.5%), compared to Asami (14%). Three of the six outliers in Figure 5.9 consisted of the word *and* with a lengthened vowel.

Overall, as expected for a higher proficiency learner, Keiko's performance is higher for articulation rates. In the monologues, Keiko's median articulation rate (2.78 syllables per second) was significantly higher than Asami's (2.24), $W = 38583$, $p < .001$, $r = -0.17$, implying a small effect. Also in the dialogues, Keiko's median articulation rate (3.28) was higher than Asami's (2.77), $W = 25979$, $p < .001$, $r = -0.15$, again implying a small effect. However, the trend that articulation rate was higher in dialogues was true for both participants regardless of proficiency (see Figure 5.9).

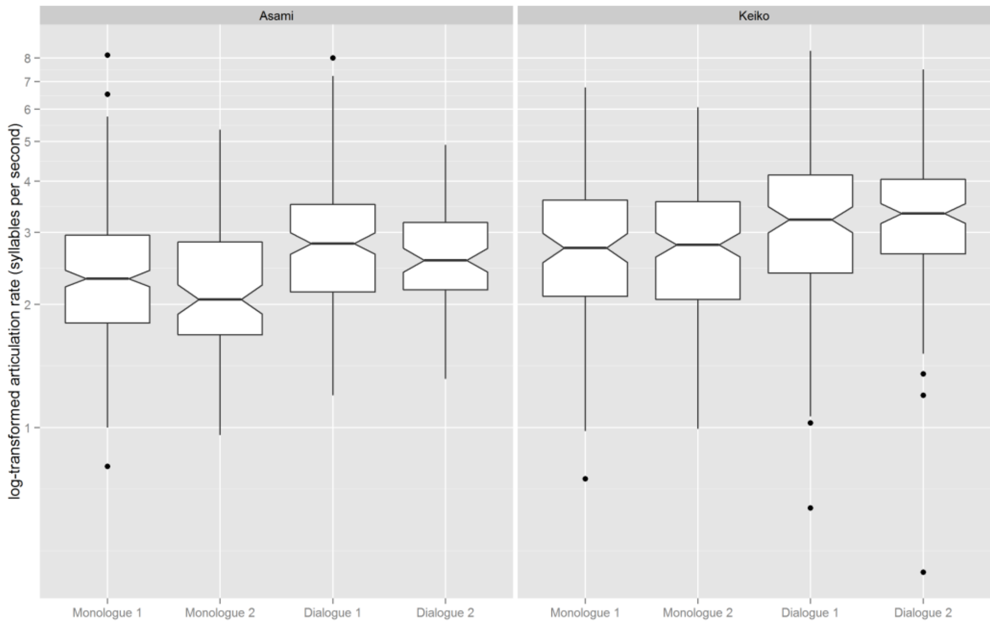

Figure 5.9 Median articulation rate (log-transformed) for Pear Story retellings. Notches represent 95% confidence intervals.

5.5 Speech rate

Speech rate is a much more common measure of fluency, and differs from articulation rate in that it takes into account silent pauses in the user's speech. Therefore, it is considered to be a more holistic measurement of monologic fluency (Towell et al., 1996). In monologic tasks, speech rate is typically calculated over the entire course of the monologue. In dialogue, the speech rate for a given speaker is calculated over the turns where that speaker has the floor. Pauses between turns are handled differently by different researchers. In the present study, speech rate was first calculated over the entire duration of each Pear Story retelling, ignoring backchanneling by the interlocutor, in order to more fairly compare the monologues and dialogues.

For Asami, the dialogues, with 91 and 95 wpm, had higher overall speech rates than the monologues, with 74 and 72 wpm (Figure 5.10). In all four retellings, the speech rate was "slower than normal", according to Tauroza and Allison (1990, p. 102), where the cut-off for "moderately slow" in the conversation genre is 160 wpm.

Part of this was due to the increase in articulation rate found in the dialogues. Another possible factor could be the length of silent pauses, which will be looked at below.

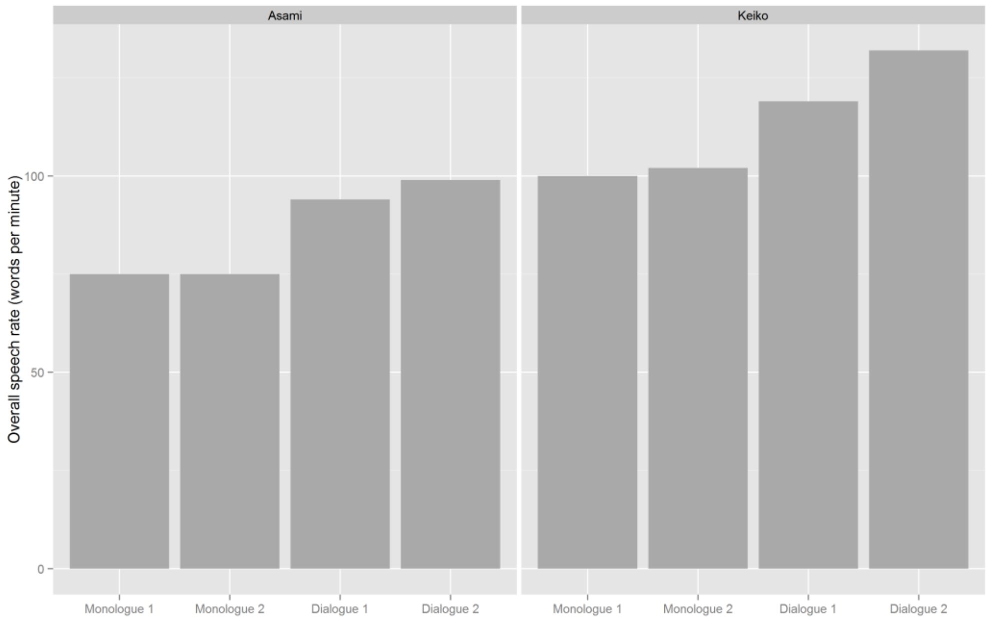

Figure 5.10 Overall speech rate for each Pear Story retelling, comparing both subjects.

As with articulation rate, Keiko showed similar trends to Asami for speech rate, with Monologue 1 and 2 having speech rates of 98 and 100 wpm, and Dialogue 1 and 2 having speech rates of 113 and 130 wpm (Figure 5.10). These values are still below normal native speaker speech rates (Tauroza & Allison, 1990), but significantly higher than the lower-proficiency learner, Asami. In fact, Keiko's lowest speech rate of 98 wpm in the first monologue is approximately the same as Asami's highest speech rate of 95 wpm in her second dialogue.

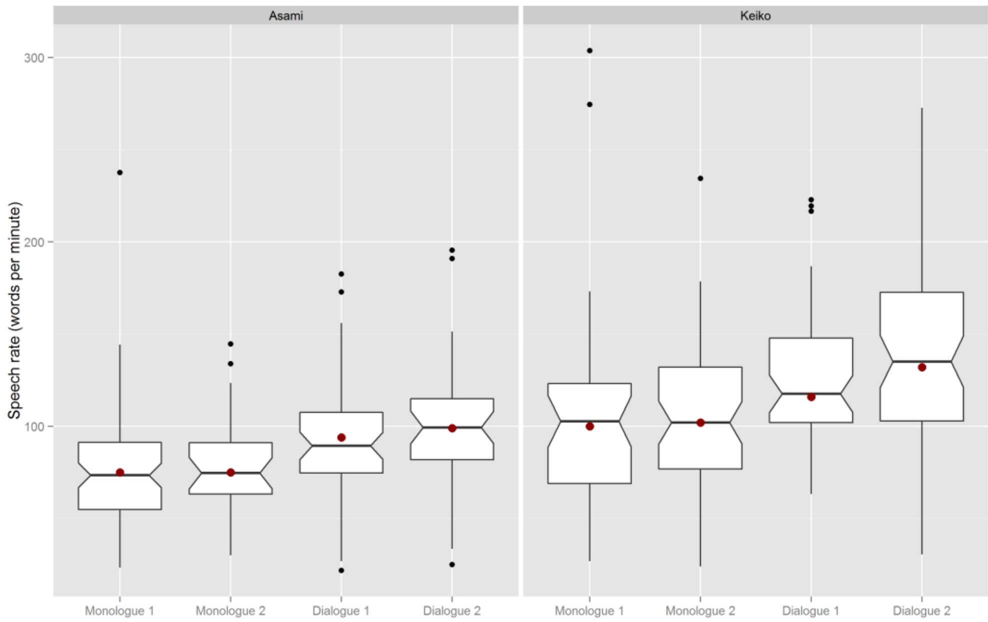

Figure 5.11 Speech rates for Asami's and Keiko's retellings calculated over each C-unit. Red dots represent overall speech rate for the given retelling.

Speech rate was also measured separately over each C-unit. Pauses between C-units were credited to the following C-unit in order to be able to account for all pauses over the course of the narrative. Distributions of speech rate per C-unit for Asami were marginally normal (Shapiro-Wilk test for Monologue 2: W = 0.956, p = .294; Dialogue 1: W = 0.956, p = .063; Dialogue 2: W = 0.939, p = .053), with the exception of Monologue 1 (W = 0.937, p < .001), which was due to a single outlier, W = 0.9857, p = .56 when removed. The outlier was the common phrase *yeah i don't know why* uttered very quickly, which could be formulaic. All four retellings had equal variances (Levene's test, F = 1.18, Df(3,181), p = .32). As with articulation rate, dialogues showed significantly higher performance (Figure 5.11), with median speech rates of 89 wpm (MAD = 25) for Dialogue 1 and 101 wpm (MAD = 20) for Dialogue 2, as compared to 73 wpm (MAD = 27) for Monologue 1 and 74 wpm (MAD = 18) for Monologue 2, H(3) = 20.03, p < .001. The significant differences were between Monologue 1 and both Dialogues, and Monologue 2 and Dialogue 2, but no significant differences between the monologues, between the dialogues, or between Monologue 2 and Dialogue 1.

Again, Keiko showed similar trends to Asami. The distributions of speech rates were all normal, $W > 0.96$, $p > .6$, with the exception of Monologue 1, $W = 0.90$, $p = .003$, which is due to two high speech rate utterances, (1) *i'm gonna talk about a movie*, and (2) *i don't know the name of*, $W = 0.9807$, $p = .77$ with the two outliers removed. Variances were equal, Levene's test, $F(3) = 1.64$, $p = .18$. Dialogue 2 (median = 130 wpm, MAD = 44) had significantly higher speech rates than Monologue 1 (median = 102 wpm, MAD = 39) and Monologue 2 (median = 102, MAD = 38), $H(3) = 21.2$, $p < .001$, however, there were no significant differences between the monologues and Dialogue 1, Monologue 1 and Monologue 2, or Dialogue 1 and Dialogue 2, determined by post-hoc Wilcoxon rank sum tests.

For both Asami and Keiko, speech rates appear to be affected more by the dialogic context than by practice, although the effect was not as pronounced as for articulation rates. The effect of higher articulation rate could be mitigated by longer pauses in the dialogues (to be examined below). This could also be the result of (1) the lower number of samples in the statistical calculations (i.e., fewer C-units than utterances), or (2) the result of differences in the speech rates of different types of C-units, i.e., non-clausal vs. clausal C-units.

For Asami, non-clausal C-units, not surprisingly, tended to be much shorter than clausal C-units (Figure 5.12). There was no significant difference in the length (in words) of clausal C-units between retellings, $H = 3.87(3)$, $p = .27$. In terms of speech rate, non-clausal C-units tended to have a wider range of speech rates in all retellings (Figure 5.13), while clausal C-units, which account for 70 to 78% of the C-units in the four retellings, showed more stability in terms of speech rate. The above significant differences between speech rates in monologue and dialogue were therefore due to the differences in the speech rates of clausal C-units, $H = 21.7(3)$, $p < .001$, which were significant between all combinations of monologues and dialogues by post-hoc Wilcoxon rank sum tests, and as is apparent in Figure 5.13.

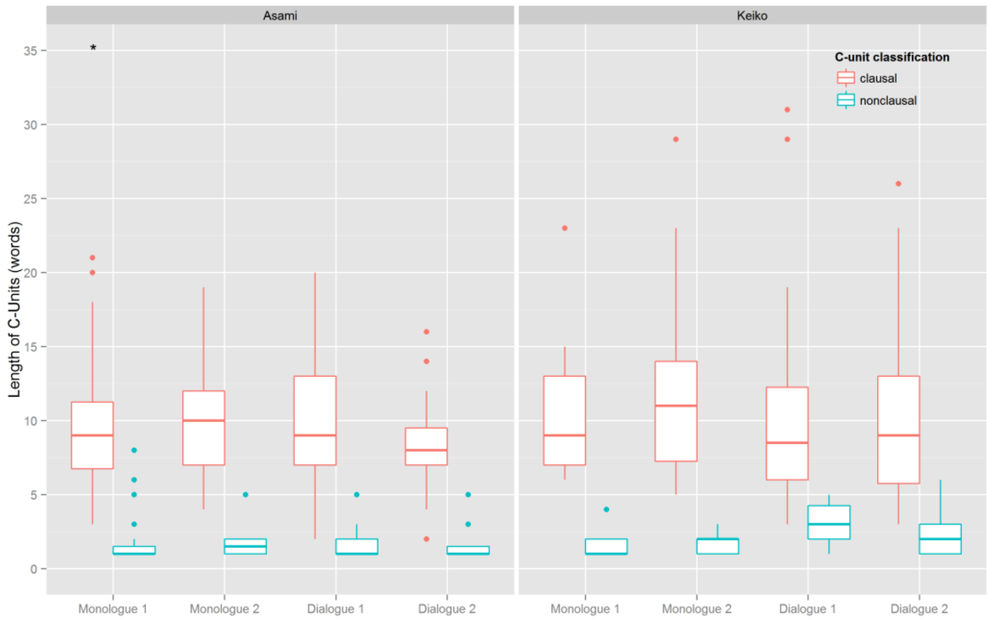

Figure 5.12 Median length of C-units by retelling for both subjects. The asterisk in the upper left corner represents an extreme outlier, a 45-word clausal C-unit from Asami's Monologue 1. This was removed from the plot in order to make the rest of the data more clearly visible.

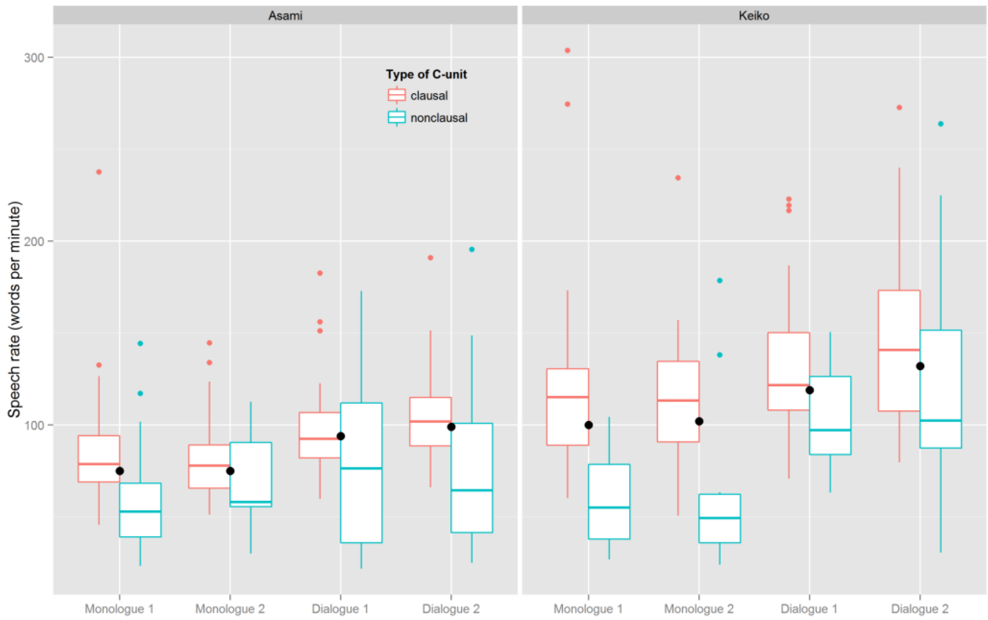

Figure 5.13 Median speech rate for C-units compared between retellings of both subjects. Black dots represent the overall speech rate for the given retelling.

Similar to Asami, in Keiko's retellings, clausal C-units were more common than non-clausal C-units, comprising 71 to 81% of the units, and longer in terms of words (Figure 5.12), and there were no significant differences in length of C-units between the retellings, $H(3) = 4.25$, $p = .23$. Unlike Asami, the differences in Keiko's speech rates were due to higher speech rates for both clausal, $H(3) = 15.5$, $p = .001$, and non-clausal C-units, $H(3) = 15.1$, $p = .002$. That is, non-clausal C-units also tended to have higher speech rates than the monologues (Figure 5.13).

Between the two case studies, Keiko's median speech rate (112 wpm, MAD = 38) was significantly higher than Asami's speech rate (median 81 wpm, MAD = 29), $W = 10576$, $p < .001$, $r = .4$, showing a medium effect. This is as expected since Keiko is the higher proficiency learner. Overall, both learners had higher speech rate performance in dialogues, although this effect was not as pronounced as the differences in articulation rate. In the next section, I will begin to look at the effect of silent pauses on these variables.

5.6 Phonation-time ratio

Another common temporal variable relating speech and silence is *phonation-time ratio* (PTR). This is the ratio of speech to the overall time spent as the speaker, i.e., including silent pauses. A PTR of 1 would mean a continuous stream of speech with no silent pauses, and a PTR of 0 would mean only silence.

For Asami, the dialogues had a 14-16% higher PTR than the monologues (Table 5.4). The length of the narrations is shorter in the second session recordings, and the total number of words was particularly low for Monologue 2. This was probably due to the fact that, as she said in post-recording interviews, she did not enjoy doing the monologue recordings as much as the dialogues, and was probably trying to finish it quickly. Interestingly, this possible effort resulted in a much shorter narrative with fewer words, but did not affect the speech rate or articulation rate of the narration.

Table 5.4 Phonation-time ratio for Asami and Keiko Pear Stories

	Asami				Keiko			
	words	*pauses*	*duration*	*PTR*	*words*	*pauses*	*duration*	*PTR*
Monologue 1	578	291	07:43	0.58	295	106	02:57	0.62
Monologue 2	224	102	03:00	0.62	550	164	05:25	0.68
Dialogue 1	372	172	03:57	0.66	447	137	03:46	0.70
Dialogue 2	230	83	02:20	0.71	491	137	03:43	0.74
overall				0.62				0.69

In the case of Keiko, comparing the first session recordings and second session recordings, the dialogues both have 9-10% higher PTR than the monologues (Table 5.4). Like articulation rate and speech rate, Keiko's overall PTR (0.69) is 11% higher than Asami's (0.62). Speech rate is related to articulation rate and phonation-time ratio mathematically, by the formula, *SR = AR x PTR*, when articulation is converted to words per minute. Therefore, an increase in either articulation rate or phonation-time ratio would result in a proportional increase in speech rate. However, PTR is also affected by the length of silent pauses, which will be examined in the next section.

5.7 Silent pause durations and frequency

The length and distribution of silent pauses is another important measure of utterance fluency, and as with the above measures, was not expected to vary significantly between retellings of the same story by the same subject, but would be expected to vary with the proficiency of the subject. However, as with articulation rate, speech rate, and phonation-time ratio above, comparisons of silent pauses in the Pear Story retellings of the two case studies revealed some unexpected differences.

Pause lengths for all the retellings of both subjects were log-normally distributed, $W > 0.97$ and $p > 0.1$ for all eight retellings, and had equal variances within each subject (for Asami, see Figure 5.14, and for Keiko, see Figure 5.15).

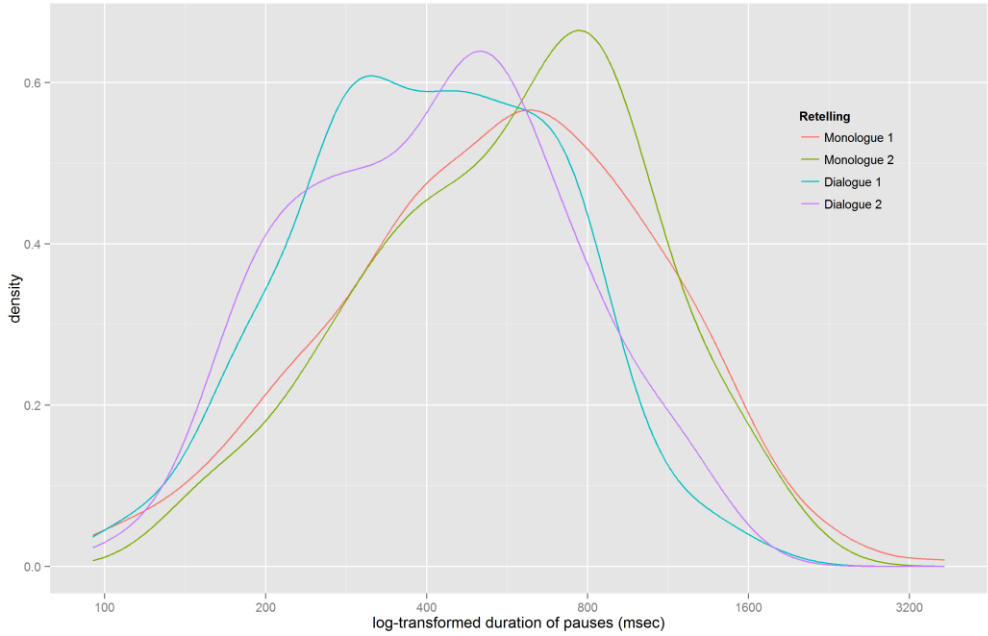

Figure 5.14 Density of log-transformed duration of pauses for Asami's Pear Story retellings.

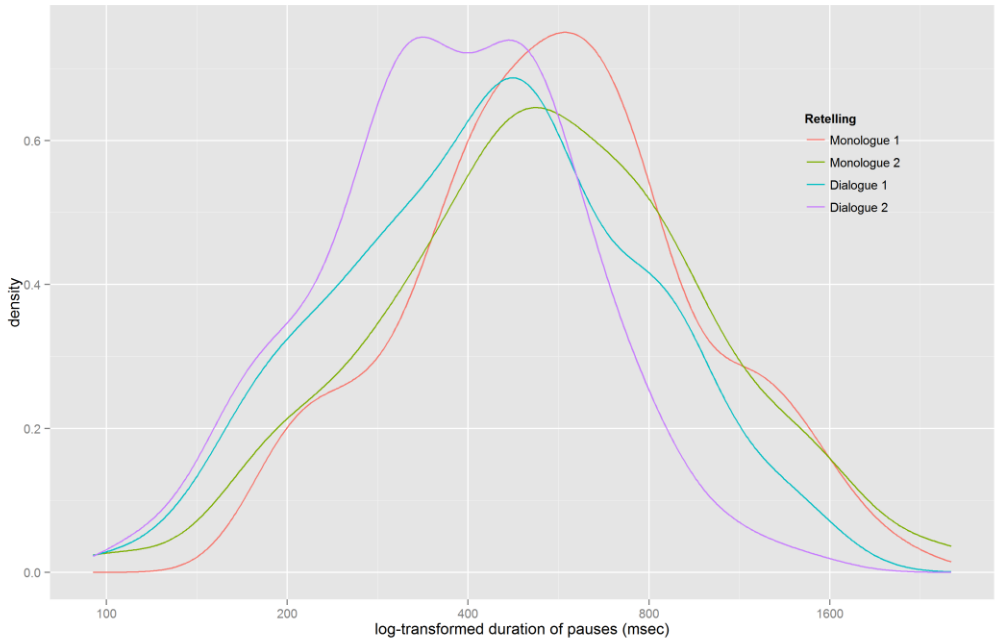

Figure 5.15 Density of log-transformed pause durations for Keiko's Pear Story retellings.

All of the retellings for both subjects had minimum pause lengths of close to 100 msec, with the exception of Asami's Monologue 2 and Keiko's Monologue 1, which had somewhat longer minimum pause lengths (Table 5.5). For both subjects the maximum pause lengths were longer in the monologues than dialogues, probably reflecting a lack of pressure to avoid silence in the monologues. Likewise, mean pause lengths were significantly shorter in the dialogues than the monologues for both subjects—for Asami, $F(3, 642) = 12.79$, $p < .001$, and for Keiko, $F(3, 547) = 14.04$, $p < .001$. Post-hoc tests showed significant differences for both subjects between the monologues and dialogues but not between the two monologues or between the two dialogues. These differences are small to medium effects, r = 0.22 for the difference between Asami's Monologue 1 and Dialogue 1, and r = 0.27 for Monologue 2 and Dialogue 2, and r = 0.23 for Keiko's Monologue 1 and Dialogue 1, and r = 0.29 for Keiko's Monologue 2 and Dialogue 2.

Table 5.5 Summary statistics for pause durations in Pear Story retellings.

	Asami				Keiko			
	Min.	*Mean (SD)*	*Max*	*N*	*Min*	*Mean (SD)*	*Max*	*N*
Monologue 1	95	544 (1.96)	3721	291	185	555 (1.71)	2125	106
Monologue 2	140	571 (1.82)	1859	102	97	534 (1.86)	2547	164
Dialogue 1	100	404 (1.75)	1675	171	95	436 (1.77)	1531	143
Dialogue 2	110	409 (1.77)	1350	82	110	376 (1.64)	1470	138

All values except N are in msec. Standard deviations (SD) are multiplicative rather than additive.

On the other hand, there was no significant difference in pause lengths between the two subjects, which is surprising given the more fluent performance of Keiko in terms of utterance length, articulation rate, speech rate, and phonation-time ratio. For monologues, Asami's mean duration of 551 msec (SD = 1.92) was not significantly higher than Keiko's 542 msec (SD = 1.80), t(615.11) = 0.338, p = .74. Similarly for dialogues, Asami's mean of 407 msec (SD = 1.75) was not significantly different from Keiko's 400 msec (SD = 1.70), t(518.403) = 0.362, p = .72. Of the temporal variables discussed here, only articulation rate, utterance length (in words), pause duration, and the number of utterances/pauses are independent variables. All others, such as speech rate, phonation-time ratio, overall duration of the retelling, are mathematically derived from those variables. A great deal of previous research has conflated these measures by treating them as independent measures. Mathematically, the longer utterances observed in Keiko's retellings should result in higher speech rate and phonation-time ratio, both of which were observed. However, Keiko's higher articulation rate should result in an additional increase in speech rate (somewhat less than proportional to the increase in articulation rate) but a decrease in phonation-time ratio. Since the difference in utterance length between Keiko and Asami is greater than the difference in articulation rate, this could explain the observed slightly higher phonation-time ratio in Keiko's case, and the significantly higher speech rate[21].

Figure 5.16 shows the relative numbers of pauses according to the two classification schemes for Asami's Pear Story retellings. As one would expect, pauses at clause junctures tended to be accompanied by a prosodic boundary tone (97.2% overall), while phrase-internal pauses tended to be unplanned (94.4% overall). Pauses at phrase junctures within clauses were more evenly distributed between planned and unplanned, however, the ratio of unplanned to planned phrase juncture pauses was highest in Monologue 1 and lowest in Dialogue 2 (Figure 5.16). This could be the result of practice. There also seemed to be evidence of practice in that the second retellings had fewer unplanned pauses between clauses and fewer planned phrase-internal pauses, implying more control over prosody and the grammatical structure of the narrative.

[21] Also, overall, Keiko had fewer utterances, which would result again in a slight increase in speech rate and in phonation-time ratio, but this effect is very small and considered negligible.

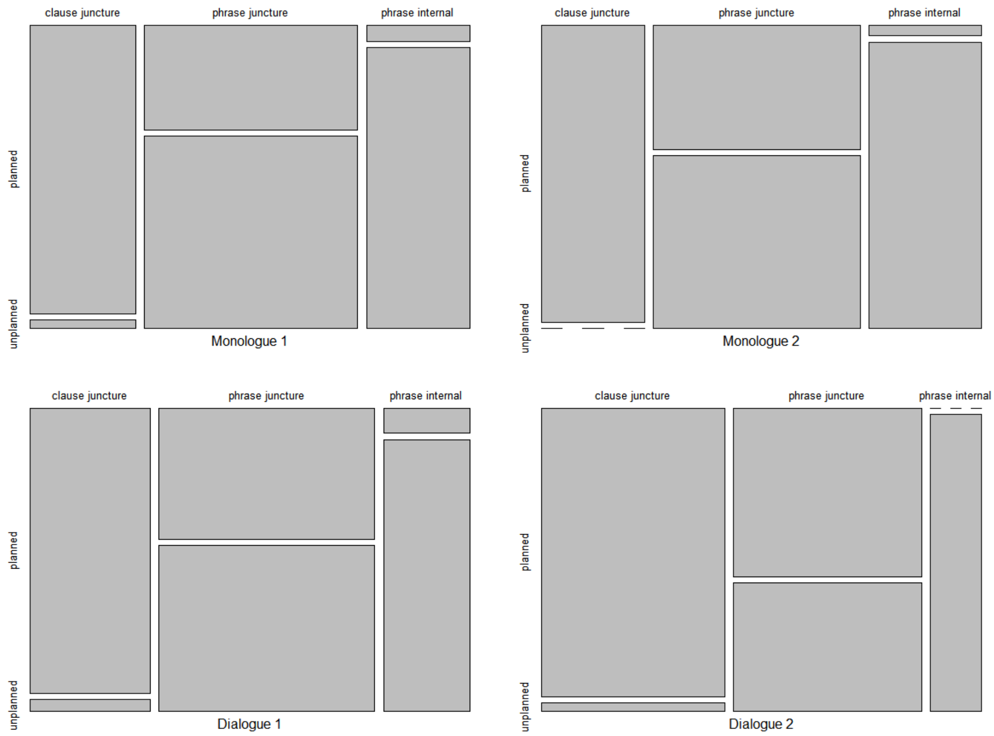

Figure 5.16 Mosaic plot showing the relative frequency of pauses by category in Asami Pear Stories.

Looking at the distribution of pauses by classification in Keiko's retellings (Figure 5.17), the overall trends were similar to those of Asami. For example, Asami had 22% dysfluent pauses, and Keiko had 20%. Asami had 51% unplanned pauses, and Keiko had 47%. There again seemed to be a practice effect observable in that there were fewer phrase-internal pauses in the second retellings. However, in Keiko's case, the dialogues have a comparatively higher percentage of phrase-internal pauses than the monologues. This could be a result of reduced planning time for the utterances due to the reduced overall pause lengths and higher articulation rate in the dialogues, or perhaps more care taken in the construction of her speech in the monologues.

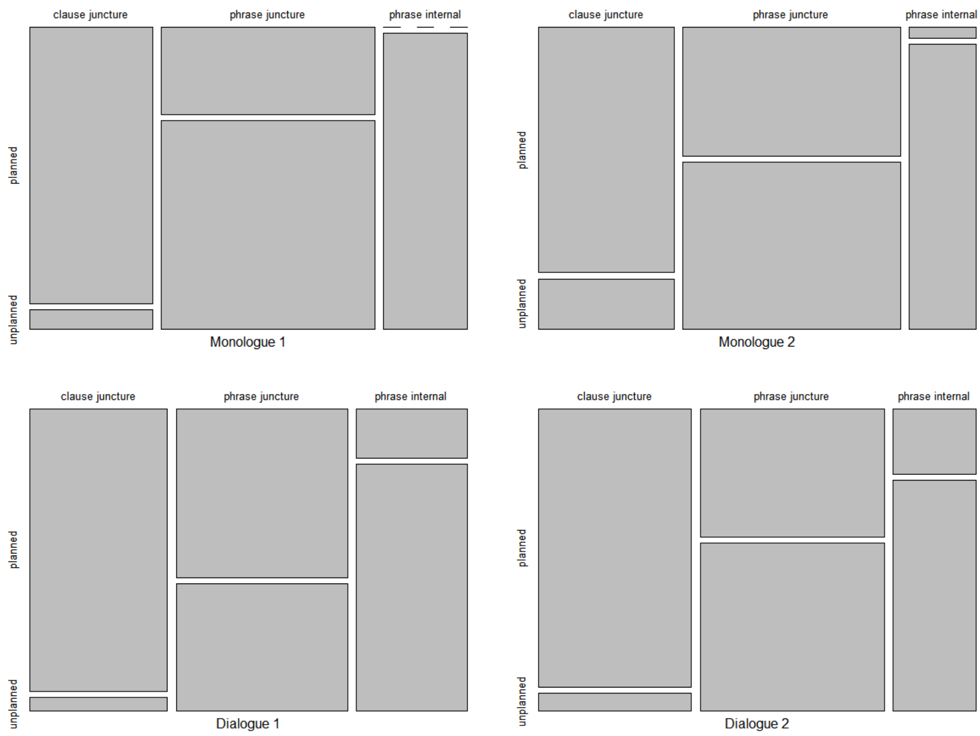

Figure 5.17 Mosaic plot showing the relative frequency of pauses by category in Keiko Pear Stories.

In Asami's retellings there were eight cases of "planned" phrase-internal pauses—that is, phrase-internal pauses that were accompanied by a prosodic boundary tone. Two of these occurred after the words *during* and *when*, which appeared to be used as discourse markers, similar to *and then* in typical native speaker speech. Three of the cases occur after an inserted vowel at the end of a word, which is often used by Japanese speakers as a hesitation device (Carroll, 2005). For example, in Extract 5.2 she uses this device on three words in succession—*and, give,* and *some* (lines 331-333). Other cases seemed to be afterthoughts, adding to the content or grammar of the preceding completed phrase as in Extract 5.3. In general, most of these appeared to involve some difficulty with the expression of the language.

Extract 5.2 Asami Pear Story Dialogue 1

```
329 ASAMI the boy <&s-> [//]
330       (320)
331       say thank you ændʊ@u [: and] gɪvu@u [: give].
332       (180)
```

```
333      samu@u [: some].
334      (295)
335      pears.
336      (355)
337      ┌to   ┐ those
338 NS1  └mmhm,┘
339      (480)
340 ASAMI boys?
```

Extract 5.3 Asami Pear Story Monologue 1

```
276 ASAMI anyway he look her.
277      (1211)
278      um
279      (285)
280      deeply,
```

For Keiko, there were 13 cases of planned phrase-internal pauses. Many of them were appeals to the listener for confirmation of word choice or content, or afterthoughts. Like Asami, she also once used the phrase *while it* as a discourse marker, which appeared to be deliberate and therefore not really dysfluent. Overall, as expected for a higher-proficiency learner, Keiko appeared to have a better ability to involve the listener to compensate for difficulties with the expression of the story.

Both subjects also had examples of unplanned clause juncture pauses, i.e., pauses between C-units that were preceded by a flat or cut-off intonation contour. For both subjects, these mostly occur after *uh* or *um* between clauses, which are normal hesitations. A few cases for both subjects follow abandoned clauses.

Pauses were analyzed by classification (location in syntactic structure and by preceding prosody) in order to determine if any particular type of pause was responsible for the lower average pause lengths in the dialogues for both learners. For Asami, planned pauses, with a median duration of 513 msec, tended to be longer than unplanned pauses, with a median duration of 467 msec ($\chi^2 = 3.99, p = 0.04$). The median duration of pauses between clauses were significantly longer (573 msec) than pauses between phrases (462 msec) and phrase-internal pauses (455 msec), confirmed by a Kruskal-Wallis rank sum test ($F(2, 645) = 8.0, p < 0.001$), but pauses between phrases were not significantly different from phrase-internal pauses in length. This corresponds to previous research showing that pauses at clause junctures tend to be longer (Pawley & Syder, 2000).

For Keiko, pause lengths showed similar distributions to those in Asami's retellings. Planned pauses, with a median length of 512 msec, were longer than unplanned pauses, with a median of 418 msec, $t(542) = 4.1$, $p < 0.001$. Also, pauses at clause junctures, with a median length of 567 msec, were significantly longer than pauses at phrase junctures (429 msec) and phrase-internal pauses (405 msec), $F(2, 541) = 17.0$, $p < .001$, but phrase juncture pauses were not significantly different from phrase-internal pauses. Again, these trends are similar to those observed with Asami.

The overall effect of shorter pauses for dialogues compared to monologues does not appear to be the result of any particular type of pauses. Multiple linear regression was used to test if log-transformed pause durations were predicted by subject, genre, pause location, pause prosody, and session, as well as for possible interactions between those variables. There were significant main effects for genre and pause location, which accounted for 11% of the variance ($R^2 = 0.11$, $F(3, 1188) = 46.77$, $p < .001$), and no significant effects for pause prosody, subject, or session. It was found that dialogue had significantly shorter pauses than monologue, and that phrase juncture and phrase internal pauses were significantly shorter than clause juncture pauses (Table 5.6). There were no interactions between the pause location or prosody and genre, therefore, the reasons for the significantly shorter pauses in dialogue must be due to some other factor, such as a faster or more even rhythm in the dialogic retellings. This rhythmic possibility will be looked at in the next chapter.

Table 5.6 Mulitple linear regression of log duration of pauses predicted by genre and pause location

Model	$R^2 = 0.11$	$F(3,1188) = 46.77$	$p < .001$
Coefficients	*B*	*SE B*	*p-value*
Constant	6.51	0.03	< .001
Genre (dialogue)	-0.32	0.03	< .001
Location (phrase juncture)	-0.27	0.04	< .001
Location (phrase internal)	-0.30	0.05	< .001

Thus, it seems that in the case of pauses, there is no significant practice effect between the first and second retellings, but there is a significant difference between monologues and dialogues, with the dialogues having a shorter mean pause length. One possible explanation for this is that the speaker feels pressure to minimize silence when there is an interlocutor present. In the interview following the conversations, Asami made a comment that she was aware of the need to avoid long silences. Another possibility is that the presence of an interlocutor provides feedback through verbal and non-verbal backchannels, which facilitate the speaker by clarifying that communication has been effective and thereby reduce hesitation in the speaker. This will be discussed further in the next chapter. The *phonation-time ratio* for both monologues was also lower than that of both dialogues (Table 5.4), which also shows that there is less overall silence in the subject's speech in the dialogues.

Surprisingly, comparing the trends for silent pauses between the lower-proficiency Asami and the higher-proficiency Keiko, there was no significant difference between the subjects, who had very similar trends with silent pauses. Therefore, pauses do not seem to be a major factor in contributing to the differences in utterance fluency of these two cases based on proficiency. On the other hand, pauses do seem to be an important element in dialogic fluency.

5.8 Speech management phenomena

5.8.1 Pruned speech rate

Pruned speech rate is calculated by removing all filled pauses, truncated words, repetition, and reformulation. Pruned utterances, therefore, are naturally shorter and result in a lower speech rate than the original utterances, and reflect the number of filled pauses and performance phenomena associated with dysfluent speech. In Asami's case, pruned speech rates for each retelling were between 6.1 and 17.3% lower than her unpruned speech rates. Although there is still a visible difference between the speech rates of the monologues and dialogues, the percentage decrease between the normal speech rate and the pruned speech rates are somewhat larger for the first session retellings than the second session, implying that practice may reduce the number of speech management phenomena and/or filled pauses for this speaker (Table 5.7).

Table 5.7 Pruned speech rate over each Pear Story retelling for both subjects

	Asami			Keiko		
	Pruned SR (wpm)	Original SR (wpm)	Change	Pruned SR (wpm)	Original SR (wpm)	Change
Monologue 1	62	75	-17.3%	88	100	-12.0%
Monologue 2	64	75	-14.7%	84	102	-17.6%
Dialogue 1	80	94	-14.9%	106	119	-10.9%
Dialogue 2	93	99	-6.1%	121	132	-8.3%

In Keiko's case, pruned speech rates for each retelling ranged from 8.3 to 17.6% lower than her unpruned speech rates (Table 5.7). The percent changes are similar in range to Asami's. However, Keiko does not show signs of a practice effect as Asami does, but rather shows that the amount of speech management phenomena and/or filled pauses are somewhat less for dialogue than for monologue. For both subjects, the pruned speech rates were higher for the dialogues than the monologues. It might be expected that more unfilled pauses and other speech management phenomena would be required to maintain the higher speech rates in dialogue, however, this does not seem to be the case. The following sections will examine this more closely.

5.8.2 Filled pauses

Filled pauses are very common in native speaker speech. In the British National Corpus, the frequency of *uh (er)* is 7475 per million and is the 17th most common word, and the frequency of *um (erm)* is 5276 per million and is the 29th most common word. The corpora in the present study are much smaller and therefore their frequencies can only be compared within the case study.

For Asami's retellings, filled pauses are more common in monologue than dialogue (Table 5.8). Both *um* and *uh* combined are 8.8 per hundred words in Monologue 1 and 5.5 per hundred words in Monologue 2, but 4.5 per hundred in Dialogue 1 and 3.1 per hundred in Dialogue 2. This, as a lower number of hesitations, also seems to reflect a practice effect with the frequency being lower in the second retellings than the first. One might expect that in dialogue the speaker has more pressure to avoid silence, and Asami actually stated something to this effect in the post-conversation interviews. However, the evidence does not support this, with fewer filled pauses in dialogues, as above, and a lower ratio of filled pauses to all pauses (silent and filled combined) in dialogues. In Asami's dialogues, filled pauses account for 7.9 % (Dialogue 2) and 8.6% (Dialogue 1) of hesitations, while in the monologues, they account for 10.5% (Monologue 2) and 14.9% (Monologue 1). For Asami, filled pauses were almost always preceded by a silence, although there were 10 cases of *and uh*. Filled pauses were also most commonly followed by a silence, and 71% of the occurrences of *um* and 29% of the occurrences of *uh* were surrounded by pauses.

Table 5.8 Frequency of filled pauses and most frequent other single words in Asami's Pear Story Retellings

Monologue 1				Monologue 2			
Rank	*Word*	*Freq*	*Per hundred*	*Rank*	*Word*	*Freq*	*Per hundred*
1	the	59	10.3	1	and	15	6.9
2	um	34	5.9	2	a	15	6.9
3	yeah	27	4.7	3	uh	12	5.5
4	he	22	3.8	4	pears	9	4.1
5	boy	20	3.5	5	he	9	4.1
6	and	20	3.5	6	up	7	3.2
7	uh	17	3.0	7	of	6	2.8

Dialogue 1				Dialogue 2			
Rank	*Word*	*Freq*	*Per hundred*	*Rank*	*Word*	*Freq*	*Per hundred*
1	the	27	7.6	1	and	14	6.1
2	and	22	6.2	2	he	10	4.4
3	mm	14	3.9	3	mm	9	3.9
4	uh	13	3.7	4	a	9	3.9
5	so	13	3.7	5	pears	8	3.5
6	he	12	3.4	10	uh	6	2.6
32	um	3	0.8	49	um	1	0.4

Table 5.9 Frequency of filled pauses and most frequent other single words in Keiko's Pear Story Retellings.

Monologue 1				Monologue 2			
Rank	*Word*	*Freq*	*Per hundred*	*Rank*	*Word*	*Freq*	*Per hundred*
1	and	27	9.2	1	and	46	8.41
2	uh	20	6.8	2	uh	38	6.95
3	the	18	6.1	3	the	37	6.76
4	he	17	5.8	4	he	23	4.20
5	a	15	5.1	5	pears	18	3.29
6	is	9	3.1	6	a	15	2.74
19	um	3	1.0	21	um	5	0.91

Dialogue 1				Dialogue 2			
Rank	*Word*	*Freq*	*Per hundred*	*Rank*	*Word*	*Freq*	*Per hundred*
1	and	44	9.9	1	and	42	8.7
2	the	27	6.1	2	the	31	6.4
3	uh	18	4.0	3	just	18	3.7
4	is	17	3.8	4	he	17	3.5
5	a	16	3.6	5	yeah	14	2.9
6	he	14	3.1	10	uh	14	2.9
23	um	4	0.9	25	um	4	0.8

In Keiko's speech, as with Asami, filled pauses are more common in monologue than in dialogue. Both *um* and *uh* were approximately 7.8 occurrences per hundred words in Monologue 1 and 2, and 4.9 per hundred for Dialogue 1 and 3.7 for Dialogue 2 (see Table 5.9). Overall, these numbers are not very different from

Asami's, although for Keiko they are more consistent between the monologues, while Asami's first monologue has more filled pauses than her other retellings. Considering that Asami's first monologue is also much longer in number of words than her other retellings, this shows even more difficulty with the first retelling of the story. Keiko is also similar in that filled pauses account for a higher percent of hesitations in the monologues (17.8% for Monologue 1 and 20.8% for Monologue 2) than the dialogues (13.3% for Dialogue 1 and 11.5% for Dialogue 2). However, these numbers are higher than Asami's, such that over all of the retellings, filled pauses account for 11.7% of hesitations for Asami, but 16.1% for Keiko. In other words, this implies that Keiko is able to use filled pauses in a way that avoids some unnecessary silences. Although Keiko also has a tendency to precede both *um* and *uh* with a silence, many of these are linked to the following word. Only 37.5% of *um* cases and 21.1% of *uh* cases are surrounded by pauses, compared to Asami's 71% and 30%.

5.8.3 Repetition, reformulation, and false starts

While silent pauses and filled pauses comprise hesitation devices, repetition, reformulation, and false starts are traditionally assumed to be signs of difficulties with the planning and production of language. Presumably, this is something that would improve over the course of the retellings with practice, and would not be likely to vary by genre. This was corroborated with Asami's retellings, where the first retellings had more frequent cases of these speech management phenomena, with 6.4 per hundred words in Monologue 1 and 7.0 per hundred words in Dialogue 1, but 4.6 per hundred in Monologue 2 and only 2.6 per hundred words in Dialogue 2. Most of these cases involved repetition and self-corrections, often apparently due to vocabulary choice or grammatical corrections.

Keiko, on the other hand, was more consistent, with 3.1 cases of these speech management phenomena per hundred for Monologue 1, 3.4 per hundred for

Dialogue 1, and 3.1 for Dialogue 2. Monologue 2 is slightly higher than the others with 5.3 per hundred. Thus, in Keiko's case, practice did not seem to be a significant effect. The higher frequencies in Monologue 2 could be due to the fact that she added more detail to that story compared to the others, including, for example, the section about the man who passes the farmer leading a goat. This was left out of the other retellings. Overall, as would be expected for a more proficient speaker, Keiko had overall fewer speech managment phenomena (3.8 per hundred) than Asami (5.7 per hundred).

5.9 Analysis of formulaic sequences

It has long been suggested that formulaic sequences are beneficial, and perhaps essential, for fluent speech (Pawley & Syder, 1983). In the present study, for both participants, the retellings were compared side-by-side to look for repeated sequences of words that would suggest that particular sequences are formulaic for the subject. Some sequences are repeated word for word, or nearly so, for each retelling. For example, the phrase, *i'm gonna talk about a movie*, is used at the beginning of all four of Keiko's retellings in exactly that form, spoken with an articulation rate that is much higher than the mean for each retelling, and resulting in a longer utterance. It was difficult to identify these sequences in an automated way (e.g., by searching for n-grams), because of variation in some of the recurrent phrases. Furthermore, some phrases varied in a way that suggests that the phrase was learned during the course of the retellings. For example, the phrase *is trying to pick up some pears on a tree* is used by Keiko in all four retellings, but the first time, in Monologue 1 (see Extract 5.4), it is spoken with three internal pauses. The longest pause (1070 msec) occurs with repetition of the word *pick,* which is likely due to lexico-grammatical choice. In the second retelling (Extract 5.5), this is reduced to two shorter internal pauses, and *in a tree* is changed to *on a tree.* In the third retelling (Extract 5.6), there is only one pause, and in the fourth (Extract 5.7), it is spoken as a continuous phrase.

Extract 5.4 Keiko Pear Story Monologue 1

```
013 KEIKO and a man,
014       (310)
015       is trying to
016       (737)
017       <pick> [/]
018       (1070)
019       pick up
020       (215)
021       some pears, in a tree,
```

Extract 5.5 Keiko Pear Story Dialogue 1

```
034 KEIKO an(d) he is
035       (326)
036       u:h picking up pears.
037       (190)
038       on the tree.
```

Extract 5.6 Keiko Pear Story Monologue 1

```
019 KEIKO u:h is trying to pick up
020       (560)
021       some pears on the tree.
```

Extract 5.7 Keiko Pear Story Dialogue 2

```
026 KEIKO is trying to pick up some pears on the tree.
```

For Keiko’s retellings, approximately 15-17% of the utterances appear to be formulaic in that they are repeated in two or more retellings—excluding the first attempts at these which are broken up by dysfluent pauses. Of these formulaic runs, almost all of them are longer than the mean length of run for the given retelling, and spoken faster than the mean articulation rate.

Asami shows much less use of formulaic sequences. Repeated phrases of more than two words account for only 2% of the first retelling (Monologue 1) and 6% of the second (Dialogue 1). The third and fourth retellings (the second session) showed 10-12% recurrent expressions. Similar to Keiko, many of the recurrent expressions are in longer utterances—particularly in Dialogue 2, which is the fourth retelling. This seems to imply a practice effect. Furthermore, only about half of the recurrent expressions are spoken at a higher than average articulation rate, with the exception of Dialogue 2 again, where almost all of them are faster than the average. Again, this seems to imply that these expressions are being learned over the course of the four retellings. Keiko, on the other hand, seems to start with a higher degree of internalized formulaic language, which would be expected for a more proficient speaker.

5.10 Discussion and conclusions

Contrary to expectations, there were substantial differences in temporal measures of fluency—those aspects of utterance fluency associated with monologic performance—that differed depending on the genre. There were also expected differences between the two subjects, based on their proficiency in speaking. The principle findings of this chapter are summarized here.

- The dialogue tasks were associated with higher articulation rates, higher speech rates, shorter silent pause lengths, and a higher phonation-time ratio, implying that the dialogues were temporally more fluent than the monologues.
- Dialogues also had fewer filled pauses and other speech management phenomena, which may suggest that the presence of an interlocutor helps the speaker in ways that reduce the need for these.
- Mean length of run did not vary by genre, implying that MLR may not be influenced by the presence of an interlocutor in narrative tasks, and may therefore be a better measure of cognitive fluency.

- Keiko, the higher-proficiency learner, had longer utterances, higher articulation rates, higher speech rates than Asami. She also appeared to use more formulaic language, or at least to have more stability in her language production than Asami.
- Keiko used more filled pauses than Asami, but fewer restarts, repetitions, and false starts.
- Practice over the course of the retellings appeared to result in longer utterances and fewer repetitions, restarts, and false starts. These effects were stronger with Asami than Keiko, implying that at Keiko's level of proficiency, practice was not as needed for this task.

Dialogue appears to be significantly more fluent than monologue. This is consistent with previous research (Bell, 2003; Ejzenberg, 2000; Riggenbach, 1989; Witton-Davies, 2012), however, in previous research, the monologic and dialogic tasks were different. For example, Witton-Davies (2012) used a picture-based storytelling task for the monologue task, and a discussion for the dialogue task. As the present study used the same storytelling task for both monologue and dialogue, it reinforces this body of research showing that dialogue is more fluent than monologue even when the task involves the same content.

In summary, having an interlocutor present appears to give a significant performance improvement even for utterance fluency. However, the reasons for this difference are not apparent from the monologic fluency variables alone. In the next chapter, I will investigate dialogic aspects of fluency, looking at how the confluence of the conversational narrative is constructed. I will also examine features such as rhythm in speech and the role of the interlocutor, and whether these features might help to explain the differences between monologue and dialogue found in this chapter.

6 Dialogic and co-constructed aspects of fluency in two parallel case studies

This chapter continues the parallel case studies of two Japanese learners of English, in tasks involving retellings of the Pear Story, either in monologue (alone) or in dialogue (with an interlocutor). The previous chapter focused on monologic features of fluency and traditional measures of fluent performance, such as speech rate and the length of utterances. Traditionally, these temporal variables have been used in research as measures of fluency in speech. As discussed in Chapter 2, most research has looked exclusively at the speaker, using these temporal variables as measures of fluency. Chapter 5 showed that measurements of temporal variables may vary depending on whether the narrative is taking place in a monologic or dialogic context. Although there are some studies that have noted similar effects, conversation is still often conceptualized as alternating monologue, and in a test situation, for example, the interlocutor would not usually be considered as altering the outcome of a learner's speaking assessment.

This chapter will investigate fluency from a dialogic point of view, by looking at dialogic aspects of the conversations that have been associated with fluency and with the confluence of the conversation as a whole. This will include looking at potential turn boundaries for signs of fluency in terms of minimizing gaps and avoiding overlap between speakers, as well as the use of prosody at these points. Also, the use of smallwords will be considered.

Another aspect of conversational interaction that has been proposed as being important to co-constructed fluent conversation is the alignment of rhythm and tempo. It was hypothesized in Chapter 3 that even with a lower articulation rate and shorter utterances, alignment of rhythm between the speakers could be what allows a learner who is of lower proficiency in terms of monologic fluency, to create an impression of higher fluency, as seemed to be the case with Asami in the pilot study of Chapter 3. In the current case study, there is the opportunity to look at these features by comparing them across the four retellings, giving more insight than might be gained from only one isolated case of conversation. The use of non-verbals will also be looked at, from both the speaker—with the use of gestures to complement her speech, and from the listener—particularly with the use of head nods.

The factors to be analyzed in this chapter, by virtue of their dialogic nature, are necessarily bound up with the particular interlocutor. Furthermore, many aspects, such as the alignment of rhythm, are expected to vary over the course of the conversation. Therefore, in this chapter, a more qualitative approach will be taken in order to highlight the details of the co-constructed fluency over the course of each conversation. In particular, this chapter employs primarily an approach informed by Conversation Analysis in that it looks at particular examples in detail throughout the data rather than generalizations or summaries of the data. In addition to this, it fits with other major tenets of CA in that it considers conversation to be an "emergent collectively organized event" (ten Have, 2007, p. 9). On the other hand, it differs from most CA in that the data is not truly naturally-occurring. The conversations were elicited using a task presented to the participants. That said, telling a story multiple times to different people is not entirely an unnatural activity, and the resulting conversations are naturalistic, i.e., not rehearsed or controlled, within the context of the task. In this study, I use the term *discourse analysis* with regard to the methodology to refer to this method of analysis which is informed by Conversation Analysis, but does not strictly follow CA methodology (for example, by using TCU's as the unit of transcription).

In order to make this analysis more manageable and comparable, one segment of each Pear Story, covering the same portion of content, was selected from each of the eight retellings. This segment occurs in the first half of each story, but not at the beginning, and is included in all eight retellings. It is the *complication*, or *complicating action*, segment of the narrative (Labov & Waletzky, 1967), which involves the description of a boy arriving on a bicycle and stealing a basket of pears. This allows comparisons between the retellings by the same subject with different interlocutors, and between subjects. This type of discourse analysis approach to the case studies facilitates identification of the factors and their interactions in the conversation that result in the co-construction of fluent performance. Furthermore, as it has been suggested that fluency needs to be described as a complex system (Segalowitz, 2010), with a multitude of interacting factors, an in-depth case study approach is advantageous since quantitative methods will have a tendency to eliminate odd cases from the data, which might be crucial to understanding the operation of such a complex system (Dörnyei, 2011).

The approach used in the majority of this chapter used a *comprehensive data treatment* method of data selection (ten Have, 2007). The eight segments, selected to cover the same content of the story retelling, were first analyzed line by line, looking at features relevant to fluency as elaborated in Chapter 3. Then, features that emerged were compared between similar and dissimilar cases between subjects and within the two retellings by the same subject, and with the monologue versions of the retellings.

Turn boundaries are considered to be extremely important points in conversation (Schegloff, 1996). The present research involves the retellings of stories, which are *narratives* rather than *elicitation* (Chafe, 1994), and *non-collaborative information provision* (Carter, 2004), where one speaker keeps for floor for an extended period of time, rather than a back-and-forth style of conversation, with a more equal distribution of the floor between speakers. Therefore, the listeners in the retellings rarely have turns beyond one or two-word responses, however, there are still *transition relevance points*, where a turn exchange would be possible, and where the backchannel is a minimal turn acting as a *continuer* (Schegloff, 1982). Whether backchanneling counts as a turn exchange or not is debatable (McCarthy, 2003), but this is somewhat irrelevant to the present study, which takes the utterance as the basic unit of speech, rather than the turn, and therefore looks at all pauses between utterances as a kind of turn boundary, or transition relevance point. In the present study, I consider any utterance to be a turn for the purposes of looking at turn boundaries.

In the next section, I will look at the principle of avoiding overlap and minimizing gaps (Stivers et al., 2009), using the same quantitative techniques used in the pause analysis of Chapter 5, but considering the conversations as co-constructed by taking into account the participation of the interlocutor. This will include the interlocutor's verbal backchannels, as well as head nods—considered to be non-verbal backchannels. The second section will take a discourse analytic approach, as I look more closely at what is occurring at these points in the retellings in terms of the use of prosody and vocabulary/grammar by the speaker, and verbal and non-verbal backchanneling by the interlocutor.

In Chapter 5, pauses were annotated and analyzed only with regard to the subject telling the story—that is, backchanneling or other speech by the interlocutor was simply ignored. In the present chapter, the conversation is considered to be co-constructed by both participants, and therefore, pauses created by both speakers together were analyzed. This was done first by measuring the length of silent gaps created when neither of the speakers was speaking. Second, head nods were included as communicative acts, and the empty gaps created when there was no speech or head nod were analyzed. These silent gaps and empty gaps were compared to the quantitative measures of pauses in Chapter 5. Also, these were compared by count. The number of pauses in the storyteller's narrative is by definition equal to the number of utterances. However, the number of silent gaps and empty gaps is lower since there are also cases of latching, where there is no space between the end of the speaker's utterance and the backchannel of the listener. The learner's speech and the native speaker interlocutor's speech were independently segmented along the wave form in ELAN, focusing only on the single speaker's speech. Small pauses (generally less than 35 msec) created by gaps between the annotations on separate tiers in ELAN in these cases were deleted if the researcher perceived no gap and it appeared that the space was an artifact of the segmentation process.

For both subjects, silences when the listener's backchannels were included tended to be shorter in duration than when the speaker was taken alone, Welch's $F(2,444) = 19.4$, $p < .001$ for Asami's dialogues, and Welch's $F(2,523) = 38.1$, $p = .004$ for Keiko's dialogues. This implies that the listeners' responses tended to occur during the speakers' pauses. In Asami's case, silent gaps (mean 304 msec) were shorter than the speaker's pauses (mean 405 msec), $p < .001$, corrected by the Benjamini-Yekutieli method (Benjamini & Yekutieli, 2001), $r = .23$, representing a medium effect. If head nods are included as backchanneling, gaps between speakers were also shorter (mean 302 msec), $p < .001$, $r = .24$, representing a medium effect. Keiko's case is similar with silent gaps (mean 284 msec) shorter than the speaker's pauses (mean 400 msec), $p < .001$, $r = .28$, representing a medium effect, and gaps including head nods (mean 267 msec) similarly shorter, $p < .001$, $r = 0.32$, representing a medium effect. For both speakers, there was no significant difference between the mean lengths of gaps with or without head nods taken into account.

Looking at the overall amount of silence in the retellings, as was expected, the interlocutors' backchannels and head nods act to reduce the amount of silence or non-communication in the narrative (Table 6.1). Taken alone, Asami's narrative in her dialogic Pear Story retellings are composed of 30.8% silence. With the interlocutors' backchannels, this is reduced to 25.3%, and the addition of head nods reduces this to 18.0%. Similarly, Keiko's narrative is composed of 27.5% pauses, which is reduced to 21.3% with the listeners' backchannels, and 17.4% with head nods.

Table 6.1 Mean and total length of gaps and overlap for Asami and Keiko's dialogic Pear Story retellings

	Asami			Keiko		
	Mean (SD)	*N*	*Total (sec)*	*Mean (SD)*	*N*	*Total (sec)*
Pauses	405 (1.75)	253	119.6 (30.8%)	400 (1.70)	274	125.8 (27.5%)
Silent gaps	304 (1.92)	265	98.3 (25.3%)	285 (1.90)	284	97.5 (21.3%)
Empty space	302 (1.91)	192	69.9 (18.0%)	267 (1.92)	247	79.6 (17.4%)
Overlaps	220 (2.18)	39	11.2 (2.9%)	225 (2.25)	80	23.7 (5.2%)

In narrative-type conversations like these, the listener generally avoids taking turns beyond backchannels, and therefore chances for overlap would likely be less than in a conversation with more equal turn exchanges, however, it is still possible that backchanneling could overlap the speaker's narrative. In Asami's retellings, 2.9% of the narrative had overlapping speech from the interlocutors, and in Keiko's retellings, 5.2% contained overlapping speech (Table 6.1). Since overlap is generally avoided in conversation (Stivers et al., 2009), the cases of overlap could be due to problems on the part of the speaker (for example, not using prosody to signal continuation) or the idiosyncracies of the listener. This will be examined in more detail below.

There also appear to be differences in the number and length of gaps and overlaps in the retellings according to the particular interlocutors. For Asami, the second dialogue retelling has fewer gaps but much more overlap than the first dialogue retelling (Table 6.2), however the mean lengths of overlaps and gaps are similar, with the exception of silences between speakers (significant at $t(148) = 3.2$, $p = .001$, $r = .2$), which are shorter in the second retelling (Figure 6.1). This seems to imply more active verbal backchanneling on the part of the interlocutor in the second dialogue.

Table 6.2 Total length and ratio of gaps and overlaps in Asami and Keiko's dialogic Pear Story retellings

Asami	first retelling		second retelling	
	N	*Total length (sec) and ratio*	*N*	*Total length (sec) and ratio*
Pauses in narrative	171	80.3 (32.8%)	82	39.3 (27.3%)
Silent gaps	180	71.6 (29.2%)	85	26.6 (18.5%)
Empty space	148	55.2 (22.5%)	44	14.7 (10.2%)
Overlaps	12	2.9 (1.2%)	27	8.3 (5.8%)
Keiko	**first retelling**		**second retelling**	
	N	*Total length (sec) and ratio*	*N*	*Total length (sec) and ratio*
Pauses in narrative	137	67.6 (28.9%)	137	58.2 (26.1%)
Silent gaps	145	55.7 (23.8%)	139	41.8 (18.7%)
Empty space	132	45.9 (19.7%)	115	33.7 (15.1%)
Overlaps	36	11.9 (5.1%)	44	11.8 (5.3%)

In Keiko's case, on the other hand, the mean lengths of gaps are somewhat shorter ($t(281) = 3.2$, $p = .002$, $r = .19$ for silences between speakers, and $t(238) = 2.1$, $p = .03$, $r. = .13$ for gaps when head nods are included) for the second dialogue (Figure 6.1). However, it should be noted that these trends, both in Asami's and Keiko's cases, are difficult to interpret, and although they show overall tendencies, do not shed much light on what is happening in the actual retelling for the given speaker. This again is support for using a case study methodology to look at the particular cases within the retellings for evidence of features that affect these temporal variables. However, these data show that the segments chosen for the case study analysis do, in fact, reflect the overall trends observed in Chapter 5.

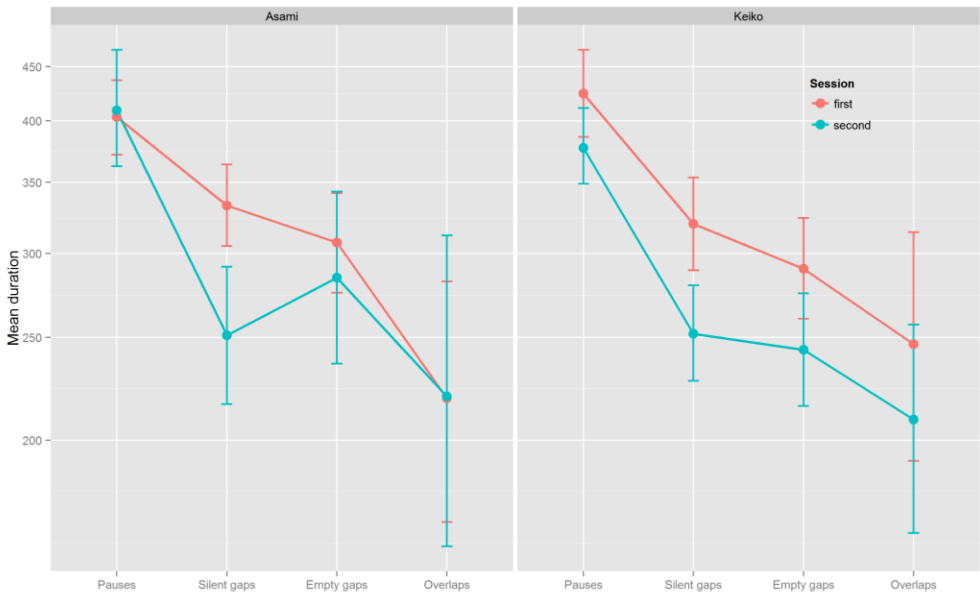

Figure 6.1 Mean log-transformed duration of gaps and overlap for Asami and Keiko's dialogues. Pauses are silences in the learner's narrative. Silent gaps are silences between both participants. Empty gaps are silences with no head nods.

It was observed that the dialogues tended to have higher articulation rate and shorter pauses than the monologues. This was independent of the types of pauses. The length of utterances did not vary significantly across retellings, although there may be a slight trend for longer utterances due to practice. In order to investigate the reasons for this and to examine other interactive factors contributing to the co-constructed fluency of the conversations, I compared the *complicating action* segment of all four of Keiko's Pear Story retellings.

6.1 Keiko Monologue 1

Keiko's first dialogic retelling is similar in profile to her second, however, the pauses and gaps in her first dialogic retelling appear to be slightly longer in mean duration (Figure 6.1) and overall duration (Table 6.2). This could be a result of practice or an effect of the particular interlocutor. For the complication segments of Keiko's retellings, the monologues have lower speech rates (113 and 101 wpm) compared to the dialogues (128 and 141 wpm)—corresponding to the trends found in Chapter 5.

As explained above, the segment used in this analysis was selected based on the content of the retelling of the story—the segment of the Pear Story where a boy arrives on a bicycle and steals a basket of pears from a farmer who doesn't notice while picking pears in a tree. In both retellings by both subjects, this segment comprises something similar to a "paragraph" of spoken discourse, which are generally preceded by longer than normal pauses (Chafe, 1994, p. 138). All four of the segments begin with the phrase *and then,* which is the second most frequent two-word phrase in her retellings at 1.5 occurrences per hundred words, preceded by long pauses. All four retellings use very similar words to describe the boy arriving on the bicycle, suggesting that these expressions could be formulaic for Keiko. Monologue 1 (lines 048 to 061) is shown below. After a long 1540 msec pause at the beginning of the segment, she begins with the discourse marker, *and then*, to introduce the complicating action of the story. In this segment of this monologue, this is the only discourse marker that she uses.

The three stressed words in Keiko's first two utterances in this extract (lines 049-051) create a rhythmic pattern with a tempo of approximately 92 bpm. The short 190 msec pause in line 050 does not result in any missed beats (Figure 6.2). Traditionally, this pause would not have been annotated at all, being less than 200 msec, however, it is preceded by a prosodic boundary tone (in this case a fall-rise continuing tone). This highlights the difficulty of annotating speech in terms of intonation units (here the two boundary tones would suggest separate intonation units), and of setting arbitrary minimum limits on pause durations (which traditionally would have resulted in annotation as a single intonation unit).

Extract 6.1

```
048          (1540)
049    KEIKO and `then,
050          (190)
051          a `boy is <`coming. [92 bpm]
052          `(664)
053          on> [//] `uh `coming.
054          (353)
055          `riding on a `bicycle?
056          `(559)
057          an(d)
058          `(680)
059          <`he:>[/]
060          `(748)
061          he just `stop in `front of `tree, [107 bpm]
```

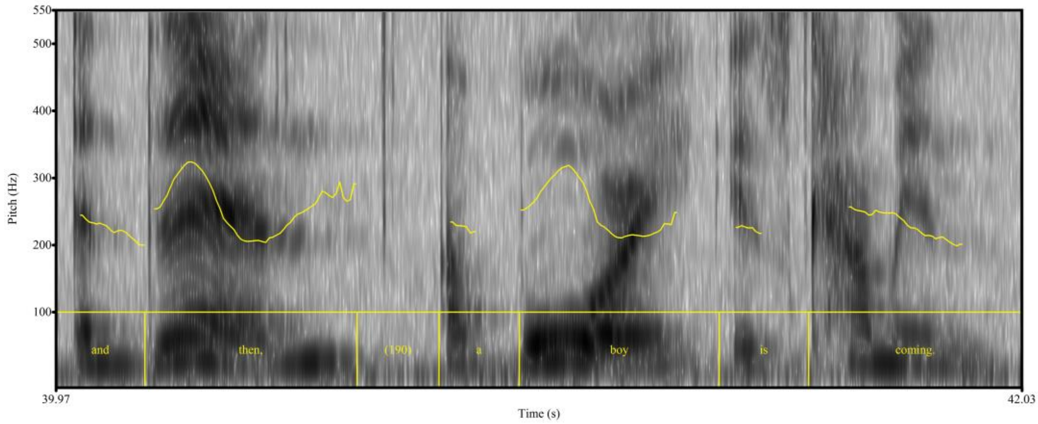

Figure 6.2 Pitch contour and spectrogram for Keiko Monologue 1 lines 049-051. High pitch transients can be seen on *then, boy,* and *coming,* with low-rise boundary tones on *then* and *boy,* and a high-falling boundary tone on *coming.* Pitch peaks can be seen to occur at approximately equal intervals, in spite of the short silence after *then.*

This rhythmic pattern is interrupted for two beats—one during the 664 msec pause, and one after *on* at the false start in line 053—but is then picked up again on the next beat with *coming* in line 053, and the two stressed words in line 055, resulting in three continuous beats at 92 bpm. Her falling boundary tone at the end of the utterance in line 051 suggests that she has completed the clause, however, after a 664 msec pause, she attempts to add to it, and then immediately self-corrects to her previous ending on *coming*. The self-correction appears to fall on the previously established beat after two missed beats, and then is followed by line 055 where she may be using *riding* as a correction of the previous attempt at using *coming* as the verb to use with the word *bicycle.* The two content words in line 055 appear to maintain the previous rhythm, although these words carry only a very slight initial pitch stress.

The next beat is skipped, and the *and* in line 057 starts between this and the following beat, resulting in two missed beats in lines 056-058. This word is neither stressed nor lengthened, and ends with a cut-off intonation contour, which may imply that she intended it to follow it immediately with another word. The lengthened *he* in line 059 begins on the next beat, but is followed by another long pause resulting in another missed beat. This is followed in line 061 by restarting with the word *he* and completing the phrase, ending with a fall-rise boundary tone. This utterance is spoken at a tempo of 107 bpm, defined by the stressed words. This increase in tempo could be an attempt by Keiko to catch up after the 2.3 seconds of hesitation in lines 058-060.

The next phrase (lines 063-069), Keiko speaks in one or two-word utterances with relatively long pauses between them, causing the rhythmic gestalt to be lost. Lines 062-064 result in three missed beats. However, *pick* in line 065 and *basket* in line 067 appear to be spoken at the same tempo established in line 061. It may be the case that the speaker continues to perceive the rhythm herself through these long hesitations, but the long hesitations could make this difficult for a listener to perceive. Line 069 also fails to establish a rhythmic pattern due to a restart.

Extract 6.2

```
062          `(550)
063    KEIKO an(d)`
064          `(593)
065          he `pick
066          `(405)
067          a `basket.
068          `(535)
```

```
069        <`full of a-> [//] uh full of pears,
```

The phrase in lines 071-077 starts with an elongated *and* followed by a 839 msec pause—the longest in this segment. There is another hesitation in line 073 with a cut-off phrase that is restarted in the following line. From line 075, Keiko seems to figure out what she wants to say or how to say it, and establishes a rhythmic pattern, based on the stressed words, with a tempo of about 90 bpm—almost identical to the rhythm established at the beginning of this segment. This rhythm is continued after a single silent pause with *bicycle* in line 077, and then after another single silent pause with line 079. In sequences of utterances such as these, Keiko shows a more proficient ability to stress content words rather than function words, and to use these to establish a rhythmic pattern. However, there are some subtle signs of difficulty. For example, she holds *didn't* in line 079 which causes *realize* to feel slightly delayed from the established rhythm (Figure 6.3).

Extract 6.3

```
070        (585)
071   KEIKO an(d)
072        (839)
073        uh <he try> [//]
074        (429)
075        he `tries to `carry the `basket. [90 bpm]
076        `(761)
077        on the `bicycle.
078        `(645)
079        but the `man didn't `realize it.
080        `(605)
081        `an(d)
082        `(245)
083        `he (i)s just `riding
084        `(416)
085        on the `bicycle,
```

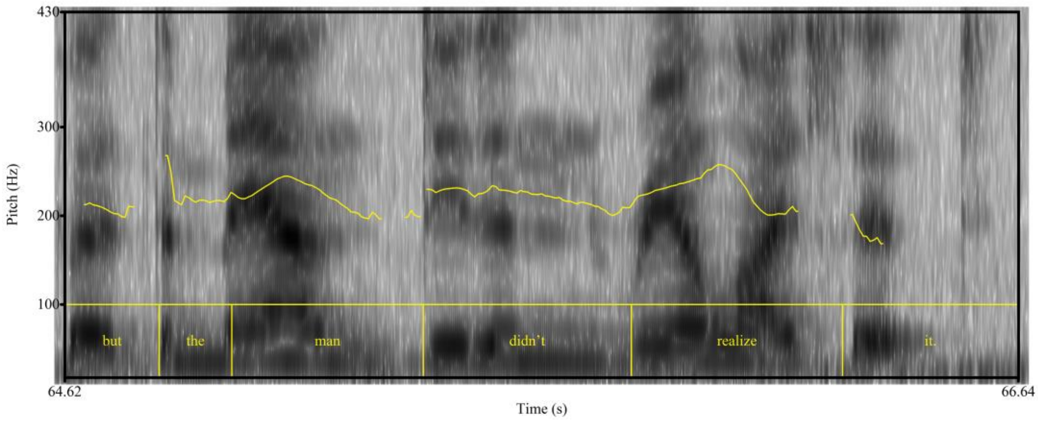

Figure 6.3 Pitch contour and spectrogram from Keiko Monologue 1 line 079. Stress can be seen by the pitch transients on the content words, *man* and *realize.* Keiko holds the /n/ in *didn't* which causes *realize* to be slightly delayed in rhythm.

The final phrase of this segment appears to continue the rhythm established in lines 075-079 with word *riding* after *and* in line 081 comes late. *Bicycle* in line 085 seems to be somewhat rushed. This would probably usually come on the beat following *riding* in line 083, but she misses that beat, and *riding* occurs before the following beat. Perhaps this is also a sign of her awareness of her own hesitations and trying to catch up the story.

6.2 Keiko Monologue 2

As with all of Keiko's retellings, she begins with the discourse marker *and then*. The first C-unit after the discourse marker is from line 089 to 093. No clear rhythmic pattern is created with the single stressed words in line 087 and 089, due to a hesitation in the production of line 091. The first sign of a rhythmic pattern in this segment begins to emerge with the utterance in line 095 with the phrase *`and `<u>stop</u> in `front of the `<u>tree</u>* (the beats are marked with slashes), at a tempo of approximately 105 bpm, although the first word, *and*, is not stressed. This rhythmic pattern may, in fact, begin with the previous utterance with the stressed words in *a `<u>bi</u>cycle is `<u>com</u>ing*, and one silent beat during the 778 msec pause in line 094.

Extract 6.4

```
086         (1406)
```

```
087   KEIKO an:(d) then,
088         (358)
089         a boy,
090         (610)
091         uh riding on
092         (107)
093         a `bicycle is `coming. [102 bpm]
094         `(778)
095         `an(d) `stop in `front of the `tree.
```

The utterance in line 097 is the longest in this segment and appears to continue the previous rhythmic pattern at a tempo of 105 bpm with beats on *the `man `picking up `pears `didn't `realize it*. It is also articulated at 4.13 syllables per second, which is higher than the average of 3.31 for multi-syllable utterances in this retelling and therefore evidence for a more fluent portion of the retelling. In most cases in Keiko's retellings, stress falls on content words rather than function words, but in this particular utterance, her speech is something between syllable-timed and stress-timed (Figure 6.4).

Extract 6.5

```
096         `(849)
097   KEIKO the `man `picking up `pears `didn't `realize it.
098         (737)
099         an(d)
100         (180)
101         the boy,
102         (407)
103         u:h just put
104         (407)
105         his bicycle on the ground,
106         (862)
107         a:n(d)
108         (683)
109         u:m
110         (350)
111         an(d)
112         (430)
113         pick
114         (626)
115         a `basket `full of `pears, [86 bpm]
116         `(969)
117         `a:nd `then, [90 bpm]
118         (407)
119         u:h <`put> [/]
120         (510)
121         `put it
122         (186)
123         `on his `bicycle.
124         `(837)
```

```
125         `an(d) jus(t)
126         `(154)
127         he has `gone,
```

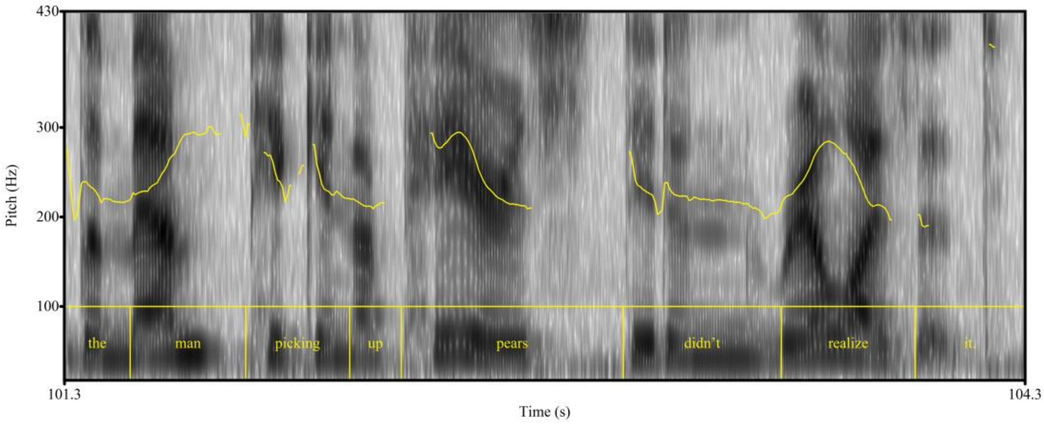

Figure 6.4 Pitch contour and spectrogram from Keiko Monologue 2 line 097. This utterance shows an example of Keiko's rhythm being between stress-timed and syllable-timed. There are clear pitch transients on the content words, *man*, *pears*, and *realize*, and less clear transient on the second syllable of *picking*. The rhythmic beats occur on all of these words as well as on the function word *didn't* which doesn't appear to be stressed.

This is followed by a long 737 msec pause (line 098) and a slight hesitation after *an(d)* in line 099 that results in three missed beats and delays the next stressed word *boy*. The following utterances also show signs of difficulty, with a filled pause at the beginning of line 103 and a short silence immediately afterward. In line 105 she stresses two words which could be interpreted as establishing a tempo of 90 bpm—slower than the previous tempo, but this is difficult to justify since three beats are required to clearly establish a pattern (Auer et al., 1999). This is followed by a long 862 msec pause (line 106) which could contain one silent beat, with *and* in the next utterance breaking the silence on the next beat. The same thing occurs with the following silence and filled pause *u:m* in lines 108-109. However, the following *an(d)* and stressed word *pick* in lines 111-113 come early and would disrupt this rhythm.

The following utterance *a `basket `full of `pears* has two stressed words, but appears to contain three beats at a tempo of 90 bpm. This is followed by a 969 msec pause which contains two silent beats. The next word *a:nd* occurs just after the beat, and is lengthened noticeably, which delays the stressed word *then* in line 117. The rhythm is re-established with a tempo of 90 bpm with beats in lines 119-123 as follows: *u:h `put `put it `on his `bicycle*. The next beat is missed but it is possible that the rhythmic pattern continues weakly in lines 125-127 with *`an(d) jus(t) `(154) he has `gone.*

Keiko's rhythm at times appears to be carried on the stressed words, and at other times seems to be carried on some unstressed words as well—similar to Asami, as will be shown below. However, in these cases, the stressed words are much more clearly intonated with pitch contours than Asami's generally flatter intonation.

In the last C-unit of this segment, she seems to intend to utter a longer noun phrase—possibly *the man picking up pears* since she cuts off a word beginning with /p/—and then she breaks up the utterance into two parts with *the man* in line 129 and *on the tree* in line 131. After this restart, the following two utterances do not seem to create a clear rhythmic pattern. Although the words *man, tree,* and *realize* are all stressed, the tempo they suggest is approximately 45 bpm, which is perceptibly very slow, and below the lower bound for rhythm (Auer et al., 1999). However, this could also be interpreted as a continuation of the previous 90 bpm tempo, with intervening single silent beats in lines 130 and 132.

Extract 6.6

```
128   KEIKO (470)
129         but <the man &p-> [//] u:m the `man,
130         `(459)
131         on the `tree,
132         `(447)
133         didn't `realize it.
```

6.3 Keiko Dialogue 1

Similar to Keiko's other retellings, in Dialogue 1, she begins with a long pause—in this case, a 485 msec and 725 msec pause surrounding a lengthened *and* (line 085-087), which is collectively a 2170 msec hesitation. She begins this part of the story with the discourse marker, *and then,* as in both of her monologues, and with almost the same first clause. However, in this case, she uses *coming*, rather than *riding*, with *on a bicycle,* and has a self-correction of *the* to *a* in lines 094 and 096.

Extract 6.7

```
085        (485)
086   KEIKO a:n(d)
            * <- K looks down & right
087        (725)
088        and `then
089        (95)
090        a `boy, [90 bpm]
           * <- K resumes eye contact
091        (255)
            1==
092   NS3   mm`hm,
            =====
093        (65)
           ===+
094   KEIKO is `coming <on the:> [//]
                 * <- gestures right hand move from
                      right side pointing
                      inward to meet left hand in front
                        * <- K looks away briefly
095        ``(205)
096        on a `bicycle.
                  * <- 2 hands together beat down
097        `(340)
             2== <- nod preceding by eyes up & left
098   NS3   mm`hm,
            =====
099        (163)
           =====
100   KEIKO the `bicycle is `really `big, [100 bpm]
                 * <- gesture 2 hand palms inward
                                     * <- hands spread outlining
                                          a big round shape
            ==+
101        (215)
102   NS3   &=laughs
103   KEIKO `bigger than `him,
             * <- 2 hand beats on syllables, move right on "him"
                         2=== <- nod with smile
104        (140)
           =====
105   NS3   mm`hm,
            ====+
106   KEIKO `and `he:` just `put the
                  * <- K looks up & left
             * <- hands return to front of chest together
107        (185)
```

```
108         `bicycle on the `ground,
             * <- K resumes eye contact
             * <- 2 hands palms inward beat
                            * <- 2 hands downward her right side
                     1==========+
109         (235)
110   NS3   mm ┌`hm┐
111   KEIKO    └  a┘nd
```

A rhythmic pattern with a tempo of approximately 90 bpm is established by the speaker’s stressed words and the interlocutor’s backchannel in lines 088 to 092, and is continued with Keiko’s *coming*. The lengthened *the* in line 094 results in a missed beat when the speaker self-corrects *the* to *a.* The following beat is also missed, with the stress on the word *bicycle* in the correction (line 096) falling somewhat after the next beat. However, the listener’s response *mmhm* falls on the following beat aligned with the previously established rhythm. The utterance in line 100 contains three beats, *the `bicycle is `really `big,* at a faster rhythm of 100 bpm. Keiko may be rushing to catch up after the hesitation. Line 103 starts on the next beat after one beat of silence, where the listener laughs. This laughter starts on that silent beat. The interlocutor’s backchannel once again falls on the following beat, at the newly established tempo. The *and* in line 106 starts just before the next beat, which is not really abnormal considering that the word is not stressed. This results in one beat without a stressed syllable, which is followed with *he:* on the next beat. This hesitation on *he:* results in one missed beat, however, *put* in line 106 and *bicycle* and *ground* in line 108 re-establish the previous rhythm. The rhythmic alignment continues with the interlocutor’s backchannel in line 110. It appears that throughout this segment, which contains two minor hesitations (lines 094 and 106), that the speaker and the interlocutor are able to maintain a co-constructed rhythm at a tempo of 90-100 bpm, in spite of the minor hesitations. However, it also appears that the tempo may vary slightly over the course of the conversation, in this case possibly as Keiko tries to catch up from her less fluent moments.

The next part of the retelling contains what appears to be more difficulties, with appeals to the listener for confirmation of understanding, as evident by the rising boundary tones on the utterances in lines 113, 115, and 117, after a relatively long 555 msec pause. This long pause results in a missed beat from the previous rhythm. After this Keiko appears to slow down the tempo slightly to 90 bpm with the stressed words in lines 113-117. The interlocutor's reaction, *oh*, in line 118 occurs between beats, and overlaps slightly with the end of the word *pears* in Keiko's utterance in line 117.

Extract 6.8

```
109          (235)
110   NS3    mm┌`hm┐
111   KEIKO    └  a┘nd
              2====+
112          `(555)
113          maybe `stole? [90 bpm]
114          (220)
115          a `basket?
                * <- gestures 2 hands holding basket
116          `(320)
117          full of `pears?
                      * <- same gesture basket, beat downwards
                      * <- NS pulls head back and sidelong glance
118   NS3    oh!
119   KEIKO a `basket `yeah.
                 * <- another beat with same gesture
120   NS3    &=laughs
121          (100)
122   KEIKO an(d) `put
              * <- K looks down & left, lifts basket hands up
123          (170)
124          the `basket on the-
125          `(365)
126          bicycle?
             * <- K resumes eye contact, gestures basket down
127          `(534)
128   NS3    ┌`ah!           ┐
129   KEIKO └and then just┘ gone,
                               * <- both hand flick left
```

Generally, listener responses appear to occur only after utterances that end with prosodic boundary tones, and utterances that end with flat intonation contours seem to not prompt a listener response. In lines 113-117, there are three utterances with high-rising boundary tones, but the listener response comes only after the third in line 118. A possible explanation for this is that line 114 is not long enough to open up a beat for a listener response. There is, however, an empty beat in the pause in line 116, but at this point, rather than verbally responding, the listener pulls her head back and gives a sidelong gaze as if to show surprise. NS3 then verbally responds in line 118.

The rhythm is picked up again in line 119 on *basket* and *yeah*, and then in line 122 on *and* and *put*. The rhythm is maintained with the word *basket* in line 124, but the following beat, which is where *bicycle* should fall, is missed due to the hesitation after *the.* This hesitation is odd in that Keiko has already formed a gesture modeling the picking up of a basket, and therefore seems to have decided the content of that phrase. Furthermore, she cuts off *the* in line 124 with a /b/ phoneme, in preparation for the word *bicycle*. On the other hand, during this and the previous utterance, Keiko is looking slightly downward and to her left, as if she is trying to remember something. Eye contact returns when she says *bicycle* in line 126, but rhythm misses two beats. The interlocutor appears to maintain alignment with the rhythm by placing her reaction *ah* in line 128 on the second beat after line 124. Keiko, however, appears to have reset the rhythm at the same tempo when she says *gone* before the next beat in line 129.

In line 131, Keiko maintains the 90 bpm tempo from *gone* in line 129 after a single silent beat in the 723 msec pause in line 130, with *man* on the beat, but the words *picking* and *pears* appear to be delayed, suggesting a slight hesitation. However, the interlocutor's backchannel falls on the following beat in alignment with the previous rhythm, as does *realize* in Keiko's next utterance (line 133) and the interlocutor's reaction in line 135.

Extract 6.9

```
130          (723)
131   KEIKO but the `man, who was` picking up` pears, [90 bpm]
                      * <- right hand points to right
                                     * <- 2 hands rolling around
                                        each other then palms up
                                         holding pears on "pears"
                            1=====+             2=====
132   NS3   uh`huh,
            =====+
133   KEIKO he didn't `realize it.
             * <- hands together in front
134          (230)
135   NS3   `oh!
```

Throughout this segment, the listener seems to be able to align correctly to the rhythm established by the speaker, even when the speaker changes the tempo, using minimal listener responses. Listener responses in narratives tend to contribute in a reactive way—that is, not in a syntactic or lexical way—unless there are problems in the narrative (Carter, 2004). The listener responses in this segment do just that while also helping to maintain the established rhythm over pauses and minor hesitations.

All of the listener responses in this segment occur between phrases in Keiko's narrative and that are preceded by prosodic boundary tones. This is true with all of the five backchannels and three reactions that occur in this segment, as well as for the interlocutor's laughter. The other gaps, which are hesitations or signs of dysfluent performance, are generally left empty and result in a missed beat. For example, the hesitation after the cut off phrase in line 094 is not filled, as is the phrase-internal pause in line 107. In particular, the portion of the story from lines 111 to 117 contains three pauses of 220 to 555 msec, but no interlocutor backchannels, nods, or other reactions. Perhaps this is merely the listener waiting for the end of the phrase to react, or it may be a sign to the speaker that the current speech has not yet been understood, and to continue until the listener is able to understand what was meant. Lines 122 to 126 also have phrase-internal pauses left empty by the interlocutor. The long pause in line 130 seems to be an exception in that line 129 is a complete phrase with a low-rise boundary tone, however, the interlocutor does not fill the following silence, verbally or non-verbally. This 723 msec pause occurs after the end of the clausal C-unit that begins with line 106. Pauses between clauses tend to be longer than other pauses (see Chapter 5), and this particular point could be interpreted as the end of the scene. It is also possible that the interlocutor doesn't think this is the end of the story, and is waiting for a continuation before reacting, as she does in line 135. In Keiko's second dialogue, this same line, *and just gone*, is followed immediately by the interlocutor's reaction, *wow*, with the same continuation, *and the man uh picking up pears he didn't realize it.* In fact, three of the four retellings (not Monologue 1) contain the point using the word *gone*, and all four contain the point about the farmer not realizing it.

6.4 Keiko Dialogue 2

Keiko's second dialogic retelling is similar in profile to her first, however, the pauses and gaps appear to be slightly shorter in mean duration (Figure 6.1) and overall duration (Table 6.2). This second dialogue seems to show more fluent performance in terms of temporal variables than her first dialogue or her monologues. The narrative from lines 099 to 114 is made up of two chained clauses, with no pauses that could be considered dysfluent, and represents a highly fluent segment of the narrative. In Keiko's Pear Story Dialogue #2, this section is preceded by a 556 msec silence (line 099), which is substantially longer than the average for the pauses in Keiko's speech of 377 msec. As with the other retellings, she begins the segment with *and then*, and uses a fall-rise intonation contour to signal continuation before a minimal 110 msec pause (line 101). The next utterance, *a boy*, is also followed by a fall-rise boundary tone, but then a longer 220 msec pause, which is broken by both the listener's backchannel *mmhm* and Keiko's continuation in line 105, resulting in a 145 msec overlap. The next utterance, *uh riding on a bicycle*, ends with rising intonation, which normally would express a question or doubt, but Keiko says this with eye contact and a confident facial expression, which suggests that she may be using this boundary tone simply as a sign of continuation or as an appeal to the listener for confirmation of understanding. This incongruity seems to be recognized by the interlocutor, who lifts her head in a questioning manner during the 205 msec pause in line 106. The interlocutor remains silent, however, possibly due to the syntax of the relative clause in line 105 implying continuation, until what appears to be the end of the clause with *is coming* in line 107. This could be interpreted as the ending of the clause because of the falling intonation and the grammatical structure, although Keiko continues it in her following utterance, *and stop in front of a tree*, in line 111. The falling boundary tone in line 107 is followed by a long 625 msec pause in the speaker's narrative, which seems to imply that that the continuation in line 111 was an afterthought. In terms of minimizing gaps in conversation, the interlocutor's backchannel, *uhhuh*, results in two very short silences of 165 and 95 msec, mitigating the silence of the speaker's long 625 msec pause.

Extract 6.10

```
099          (556)
100   KEIKO and `then, [82 bpm]
101          (110)
102         a `boy,
               * <- right hand palm up out to right side
103          (220)
104   NS4   mm[`hm,]
105   KEIKO   [uh ] riding on the `bicycle?
                                        * <- 2 hands together
                                    move down palms downward
106          (205)
              * <- NS head rise, eyebrows raised
107         is `coming.
                * <- 2 hands move from right side inward to chest
108          (165)
            3====
```

```
109   NS4   `uhhuh.
             ====+
110          (95)
111   KEIKO an(d) `stop in `front of the `tree, [118 bpm]
                      * <- 2 hands down beat, roll around each other,
                            palms inward beat down on "tree"
112          (200)
            1====
113   NS4   `mmhm.
             ====+
114   KEIKO an(d) he: just `put his `bicycle on the `ground.
                                                         [95 bpm]
             * <- hands together     * <- 2 fists move downwards
                               hands down right side -> *
                                                            4===
115   NS4   `mmhm mmhm.
             ========+
116          (140)
```

Taken on its own, the overlap in line 104 and 105 might be construed as a sign of a misunderstanding or problem with the flow of the conversation. However, looking at the rhythm of the speech, Keiko puts stress (annotated with underlining) on *then* and *boy* which sets up a rhythmic pattern with a tempo of 82 bpm with no missed beats. The following silence is broken by Keiko's *uh* on the next beat following the stress on *boy*, thereby keeping closely aligned with the previous rhythm. There seems to be rhythmic alignment between the participants in that the stress of the interlocutor's backchannel, *mmhm*, stressed on the second syllable, is rhythmically aligned with Keiko's *uh* to break the silence. In this way, the overlap demonstrates the alignment between the speakers rather than a problem with the conversation—that is, both speakers are working towards filling the silences. Furthermore, this suggests that the statement that silences are minimized may be an oversimplification, and that the rhythm of the conversation is important to determine when those silences are filled. The next beat in the rhythm falls on *bicycle*, followed by *coming* after a 205 msec silence. After *coming*, there is the relatively long 625 msec pause in the speaker's narrative, discussed above, which is partially filled with the listener's response, *uhhuh*, stressed unusually on the first syllable, which is aligned with the rhythm created by the preceding four beats. The next beat again follows this rhythm and occurs on *stop*, and then *tree*, in line 111. The final statement of this extract, in line 114, follows a similar pattern. The exact intervals between the beats are not perfectly equal, some being longer and some being shorter than the 750-760 msec period set up by the first two utterances. This is normal and this analysis is done by listening to the audio and perceiving the rhythmic gestalt of the conversation, rather than relying wholly on objective measures, as per Auer et al. (1999). It should be noted that the listener responses in this extract do not always occur on the strong beats set up by the speaker. In the case of this segment, the listener responses in lines 104 and 109 occur on strong beats, but the one in line 113, for example, occurs between strong beats by the speaker.

The next segment of the retelling is shown below, which consists of a single clause, broken into 6 utterances and appears less fluent than the preceding segment in that it includes phrase-internal pauses (lines 122 and 127), and a hesitation with an elongated vowel on *he* (line 119). The interlocutor only responds when the subject pauses following a low-rise boundary tone (lines 123-125), and reacts to the story after the clause finishes with a falling boundary tone (lines 128-130). In other words, the interlocutor does not fill the silences that could be considered signs of dysfluency.

Extract 6.11

```
117   KEIKO `an(d)
               * <- Keiko looks to her right as if thinking
               * <- hands together
118         `(570)
119         i `think that `he:
120         `(440)
121         jus(t)
122         (295)
123         `try to `steal, [90 bpm]
                   *   <- Keiko resumes eye contact
124         (315)
            1====
```

```
125   NS4   ┌mm`hm.┐
126   KEIKO └a     ┘  `basket `of
                        * <- hands open palms up, beat down
             ====+
127         `(441)
128         the pears.
             * <- hands roll and palms up on "pears"
129         (70)
130   NS4   `oh!
             * <- NS looks up & left
```

The rhythm of the subjects' speech is not as clearly delineated at this point, but the tempo is still approximately 82 bpm, which would imply that the rhythm and tempo have been maintained from the previous segment. However, listening to this extract, the word *jus(t)*, which does not appear to be stressed, is somewhat delayed, which weakens the rhythmic gestalt of the segment. It is possible that rhythmic problems here are a sign of less fluent performance on the part of the speaker. In Keiko's Dialogue 1, lines 111-117, she also has difficulty with this portion of the story, which is spoken in short utterances of 2-3 words, with rising intonation on each utterance, apparently to appeal for confirmation from the speaker. From her facial expression in the current segment, Keiko appears to be somewhat uncertain of what she is saying, but regains confidence and continues the story smoothly once the interlocutor has given an appropriate surprised response to the information about the boy stealing the basket (line 130). In both of the monologue versions of Keiko's story, she does not use the word *stole*, but rather simply states that the boy picked up the basket of pears and put it on his bicycle. In this second dialogue retelling (and the fourth retelling overall), she states that she initially misunderstood the boy's intention (lines 156-167), which may be the reason she chooses to just describe the actions of the story rather than the intentions at this point in the monologue versions. However, similar to her dialogue retellings, the descriptions of the boy stealing the basket of pears are less fluent. In Monologue 1 (lines 057-077), this is composed of 11 utterances, separated by fairly long pauses of 405 to 839 msec, and mostly without a preceding boundary tone, implying unplanned hesitations. Monologue 2 is similar with short utterances and long pauses during this portion of the story (lines 105-125).

In general, it appears that the interlocutor avoids backchanneling or responding during pauses that occur in the middle of phrases, such as those in lines 118, 120, and 122. However, the pause spanning lines 124-125 is internal to the phrase *steal a basket,* but is filled with a head nod and backchannel by the interlocutor. On one hand, the pause in this example is preceded by a low-rise boundary tone, signaling a planned pause, as well as the return of the speaker's gaze toward the interlocutor. During lines 117-123, where the interlocutor does not backchannel, the speaker is looking to her right as if considering what she is saying or trying to remember the story.

The next segment is shown in below, which shows signs of relatively more fluent performance. The utterances are longer: Line 131 is a 7-word utterance, line 137 is 8 words, and line 148 is 12 words—Keiko's longest utterance of all four retellings. This particular line is composed of phrases that are repeated through her four retellings—particularly the phrases, *the man picking up pears* and *he didn't realize it.* In addition, the tempo of this segment (lines 131-137) is approximately 95 bpm, which is faster than the preceding segments. The first utterance (line 131) ends with a low-rise boundary tone, and is immediately followed by a listener response which is accompanied by a head nod. The second utterance (line 134) prompts a backchannel from the listener immediately after the word *pears*, which is syntactically the end of that clause, although the listener response ends up overlapping Keiko's *and*. This backchannel is immediately preceded by a head nod.

Extract 6.12

```
131   KEIKO yeah `there are some `basket in `there, [100 BPM]
                   * <- hands together, eyes downward
                                   * <- hands open palms down,
                                         resume eye contact
                                             1=====
132   NS4   `mmhm,
             ===+
133         (110)
134   KEIKO uh `full of `pears ┌ an(d)┐
135   NS4                      └`mmhm.┘
                 * <- hands roll & palms up on "pears"
                                 1===+   * <- K looks up & right,
                                                 hands together
136         (415)
137   KEIKO `he `tried to `pick up `one of them.
                             * <- K resumes eye contact
                     hands palms in grab -> *
                    a basket and lift up
138         `(625)
             1==+
139         `an(d)
140         (357)
141         `put [118 bpm]
             * <- gesture continues beat in front of face
142         (245)
143         `on his `bicycle
                      * <- gesture continues down & left
                         3== <- raises head
144   NS4   `uhhuh.
             ====+
145   KEIKO `an(d) jus(t) `gone.
                             * <- 2 fists push out to left
146          (65)
147   NS4   `wow.
148   KEIKO `yeah and the `man uh `picking up `pears ┌he `did┐n't
`realize it.
                            * <- right hand points out to right
                         right hand palm up -> *
                  right hand rapid shake on "didn't" -> *
149   NS4                                          └mm`hm, ┘
                                                    1====+
150         (140)
151         `oh o┌`kay. ┐
             * <- NS leans back
152   KEIKO     └yeah. ┘
153         (485)
```

The third utterance (line 137) is followed by a relatively long silence. This is after a falling prosodic boundary tone, implying that the current thought is finished, in spite of the fact that the subject continues the C-unit in line 139. Although there is no verbal response at this point, the listener gives a very short single head nod on the beat during this 625 msec pause (line 138). The following short utterances in lines 139-143 are dysfluent from a syntactic point of view, and also seem to correspond to a slower tempo of approximately 60 bpm. During this period the listener does not backchannel or nod, but then finally responds after the completion of the phrase in line 143 with a response showing comprehension accompanied by a strong head nod in line 144. This occurrence of *uhhuh* is spoken with strong intonation, implying more than a simple continuer. This is reasonable following the two phrase internal pauses, and helps to show the speaker that the story has been understood, which allows the speaker to move on. The lack of response during the more dysfluent portions of the speech may conversely let the speaker know that she has not been understood and needs to continue. The rhythmic pattern in lines 139 to 147 can also be interpreted as occurring at a faster rhythm of 120 bpm, with a missed beat in line 140, and stress on *on* in line 143. At this tempo, the listener responses in lines 144 and 147 fall on beats during Keiko's pauses. This 120 bpm interpretation seems more plausible because the following more fluent segment is also at 120 bpm.

The longest utterance in this extract is line 148, which has a listener backchannel at a natural pause point in the syntactic structure and on the rhythmic beat—although the speaker does not actually pause at that point. Following the utterance, the listener's reaction *oh okay* and the speaker's *yeah*, overlap such that the tempo appears aligned. Lines 148-152 appear to continue the previous rhythm with a tempo of approximately 120 bpm. The rhythm is set up with the long utterance in line 148—although the filled pause *uh* in this line causes *picking* to be delayed and thus restarts the rhythmic pattern at the same tempo slightly later. Alignment is apparent in the listener's response in line 151 (with beats on *oh* and the second syllable of *okay*) and the speakers' response in line 152.

In the next few utterances, Keiko explains that she initially thought that the boy in the story was trying to help the farmer, but then realized he was stealing the pears. Again, the interlocutor does not respond during pauses that are internal to the syntactic phrases and without a prosodic boundary tone, as in lines 154-159. NS4 responds during the the 655 msec long pause in line 161 (after the falling boundary tone in line 160) with a head nod. This pause is quite long—even longer than the average duration of a pause between clauses for Keiko This could be how it is interpreted by the listener due to the falling boundary tone preceding the pause. This prompts both speakers to break this silence at the same moment (lines 162 and 163).

Extract 6.13

```
154   KEIKO an(d)
              * <- K looks left
155         (235)
156         i thought at firs(t)
157         (445)
158         u:h `he [120 bpm]
159         (295)
160         `tried to `help him.
                       * <- K resumes eye contact
161         `(655)
             3====+ <- opens mouth as if "ah", pulls head back
162         ┌to `car┐ry the ┌basket but i think yeah.┐ <- rushed
163   NS4   └uh`huh.┘       └oh i see i see.        ┘ <- rushed
              2====+        * <- NS looks upwards briefly
164         (320)
165   KEIKO `just he `try to [90 bpm]
166         (320)
167         `steal it.
168         (85)
169   NS4   oh.
170   KEIKO `yeah.
              2===
171         (164)
             ====+
```

It is difficult to perceive a rhythmic gestalt during this portion of the narrative (lines 154-160). Lines 154-156 are not aligned with the previously established rhythm due to hesitations in lines 153 and 155, even though line 156 contains a clearly stressed word. The rhythm is re-established in lines 158-160 the words *he*, *tried*, and *help*, at the same tempo of 120 bpm. The next beat is left empty during the long pause in line 161, which could be what prompts the response from the listener, which is still aligned with the rhythm such that the strong syllables of *to carry* in line 162 and *uhhuh* in line 163 occur on the following beat. The remainder of the utterance in line 162 is spoken very quickly, as if Keiko is trying to compete or catch up during the overlapping speech. This appears to break up the previous rhythmic gestalt, and the final clause of this segment of the story (lines 165-167) is uttered at a slower tempo of about 95 bpm. This rhythm, however, is established with slight hesitations on *he* in line 165 and the pause in line 166, which causes the phrase to sound choppy. The listener's response in line 169 is not aligned with this rhythm.

Keiko uses few discourse markers or smallwords in this segment. They include *and then* and *yeah. And then* is used to begin this "paragraph" level segment of the story in all four retellings (as well at other similar points in the story). She uses *yeah* in line 131, following the interlocutor's expression of surprise at the boy stealing the basket of pears, *oh*, and likewise in line 170. Similarly, *yeah* is also used in line 148, following the interlocutor's *wow,* and in line 152, following *oh* (which overlaps with *okay*—perhaps showing a misinterpretation of the *oh* in *oh okay* as a standalone *oh.* She also uses *yeah* at the end of line 162, again likely responding to the interlocutor showing that she understands the story.

Keiko consistently uses gestures to help her tell her story—almost continuously throughout the retelling. For example, in line 020, she gestures the shape of the brim of a hat as she says *hat*, and then gestures from her neck down when she says *overalls* in line 023. Shortly thereafter, she first says the word *ladder* in line 032, which is repeated in line 034. After the first *ladder* the interlocutor does not react, so Keiko repeats the word accompanied by an up-and-down gesture, as well as rising intonation. This is followed by the interlocutor saying *okay* with a head nod to show that she has understood, which allows Keiko to continue the story.

Extract 6.14

```
016   KEIKO an(d) a man,
017         (350)
018         u:h
019         (480)
020         who is wearing hat,
021         (270)
022   NS4   mm[hm.]
023   KEIKO   [and] overalls,
024   NS4   mmhm mmhm,
025         (108)
026   KEIKO is trying to pick up some pears on the tree.
027         (295)
028   NS4   okay,
029         (155)
030   KEIKO an(d)
031         (300)
032         uh there is a <ladder.> [/]
033         (482)
034         <ladder?> [/]
035         (305)
036   NS4   okay.
```

To summarize, Keiko appears to be able to create rhythm in her speech, which the listener is able to align to. In the monologues, this appears to be hampered by her own hesitations and the lack of support from the listener. She also can use gestures to support her story in the dialogues.

6.5 Asami Monologue 1

Asami's complication segments reflect the overall trends observed in temporal variables of fluency with a few exceptions. The complication segments show mean utterance lengths very close to the overall mean utterance lengths for the respective retelling, with the exception of Dialogue 2. In Dialogue 2, Asami's mean utterance length is higher at 3.35 words compared to the overall 2.84 for that retelling. For this segment, the length of utterances appears to improve slightly with practice over the course of the retellings.

The speech rates are higher for the dialogue segments (94 and 115 wpm) than the monologue segments (81 and 80 wpm), and are similar to the speech rates for the entire retellings, except again that in Dialogue 2 the segment has a higher speech rate (115 wpm) than the overall speech rate for that retelling (99 wpm).

Pause lengths, however, show less difference between monologue and dialogue in this segment than in the overall retellings. Mean pause length in the monologues is lower (454 and 516 msec) in the segments than overall (544 and 571 msec), and mean pause length in the dialogues is higher (429 and 447 msec) than overall (405 and 411 msec).

Articulation rates follow the same trends and levels of the overall retellings, with the dialogues having higher rates than the monologues, however, in Dialogue 2 this segment has a higher articulation rate (3.10 syllables per sec) than the overall retelling (2.69 syllables per sec). This difference is also shown in the median articulation rates per multi-syllable utterance, implying that this difference is not due to single-syllable utterances being especially short.

To summarize, the complication segments of Asami's retellings generally follow the trends noted in Chapter 5, with mean length of utterance improving somewhat over the course of the retellings (presumably due to practice), speech rates and articulation rates being higher in the dialogues, but with a smaller distinction between the monologues and dialogues in pause lengths. However, the exception to this is Dialogue 2, which has a particulary high articulation rate and mean length of utterance (and therefore also speech rate), implying more fluent performance compared to the rest of the Dialogue 2 retelling.

The complication segment of Asami's Monologue 1 is much longer than the other retellings. This could show an uncertainty with what amount of explanation is needed to convey the story clearly. It could also show more difficulty with the construction of the story itself. The segment starts with a hesitation on the word *and* before repeating it in the phrase *and then* (lines 075-077). The first sequence that appears to establish a rhythm is line 083, where all three words in *maybe he is* are stressed equally with flat intonation, at a tempo of 108 bpm (Figure 6.5). This pattern could actually start at line 079, which is aligned with this rhythm. The rhythm continues through a silent beat during the 670 msec pause in line 084, and then continues with *eight* in line 085. The word *nine* in line 085 is drawn out and followed by a long 1273 msec pause, which loses the previously established rhythm.

Extract 6.15

```
074   ASAMI (435)
075         <and> [///]
076         (285)
077         and `then, [108 bpm]
078         ``(919)
079         a `boy, [///]
080         (110)
081         a `little `boy,
082         (170)
083         `maybe `he `is
084         `(670)
085         `eight or `ni:ne,
086         (1273)
```

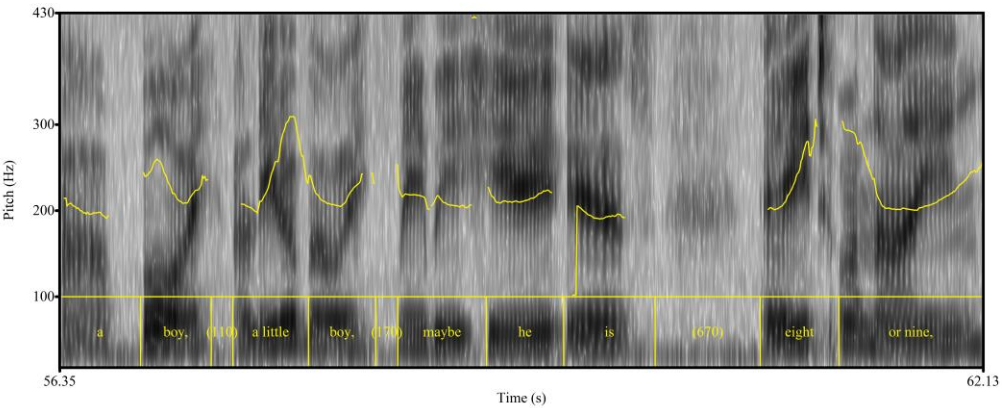

Figure 6.5 Pitch contour and spectrogram from Asami's Monologue 1, lines 079-085. Almost every word contributes to the rhythm, with the exception of articles and *or*, as can be seen by the nearly equal spacing. There is one silent beat in the 670 msec pause, but the other two short pauses do not result in missed beats. High-pitch accents appear to occur on *boy*, *little*, and *eight.* Low-rising boundary tones occur on both instances of *boy* and *nine.*

The following several utterances are spoken one or two words at a time, which precludes establishing any kind of rhythmic pattern until line 097. Here, Asami stresses the first syllable of *bicycle*, and *bigger* and *than* in line 099, but this does not appear to establish a clear rhythmic pattern. Furthermore, the 575 msec pause in line 100 and 570 msec pause in line 102 break up any potential pattern. Lines 105 to 111 vary in tempo and do not manage to set up a rhythmic gestalt. Up to this point, Asami appears to stress almost every word except for articles and occasionally *is*.

Extract 6.16

```
087   ASAMI a:nd <he:> [/]
088          (1145)
089          he
090          (220)
091          ride,
092          (375)
093          a bicycle.
094          (680)
095          actually,
096          (275)
097          the bicycle
098          (155)
099          is bigger than
100          (575)
101          the boy?
102          (570)
103          yeah.
104          (426)
105          an(d)
106          (185)
107          he found
108          (225)
109          the pears: bag,
110          (466)
111          yeah.
```

Lines 111-115 establish a rhythm of 105 bpm. This continues through single silent beats during the pauses in lines 116, 118, and 120. This is maintained until three beats are missed in lines 121-124. After the cut-off *uh* in line 127, Asami begins a self-correction, presumably of the grammar mistake in line 121, at a faster rhythm of approximately 115 bpm, which continues until the 1249 msec pause in line 134. Throughout her speech up to this point, pauses contain one silent beat, but the 1249 msec pause in line 134 contains 2 silent beats, and the silence is broken later between the second and third beats, which seems to show a breakdown in the rhythmic pattern. The next few utterances are spoken at a similar pace to the previous ones, but with a more choppy rhythmic feel, particularly with the restart in line 137. Line 141 is spoken more quickly, but includes a hesitation on *a:re*, which breaks up the potential rhythmic pattern.

Extract 6.17

```
111   ASAMI yeah.
112         (293)
113         `actually [105 bpm]
114         (278)
115         the `man `who
116         `(477)
117         are `taking
118         `(403)
119         `pears `during the `movie
120         `(433)
121         <`is
122         `(770)
123         u:m`
124         `(165)
125         `three `bags> [//]
126         (466)
127         uh-
128         (598)
129         the `man `has [115 bpm]
130         `(337)
131         `three `bags
132         `(307)
133         `to:
134         ```(1249)
135         yeah. put in
136         (240)
137         &θə- [///]
138         (360)
139         many pears,
140         (367)
141         yeah. so there a:re three bags.
```

The following several lines contain what appear to be difficulties for the speaker as evidenced by the cut-off *the* in line 145 and the self-correction in line 151, as well as a lack of rhythmic continuity. There is a long 1112 msec pause in line 152, which is then followed by almost entirely single-word utterances to complete the clause at line 165.

Extract 6.18

```
142   ASAMI (514)
143         and
144         (210)
145         the-
146         (488)
```

```
147        boy, [///] a little boy,
148        (350)
149        fin(d)- [//]
150        (618)
151        found,
152        (1112)
153        the:
154        (649)
155        bags,
156        (1388)
157        and,
158        (794)
159        he:
160        (190)
161        robbed.
162        (634)
163        the one of
164        (513)
165        the bags.
166        (644) <- sharp inhalation
```

The next part of the story begins with an in-breath that I interpret as expressing that the next part of the story is a funny or interesting part. This portion of the narrative appears to have a weak rhythmic pattern at a tempo of 114 bpm in lines 167-173, but with single silent beats in the intervening pauses (lines 168 and 170). However, after laughing while saying the word *and* in line 175, she slows down and stresses every word, but fails to establish a clear rhythm. The phrase from line 181 to 191 contains multi-word utterances, but these are spoken at an articulation rate much lower than the average for this retelling, and these contain long pauses (up to 1197 msec) and repetition. This appears to be a difficult part for her. In fact, in Dialogue 1 and Monologue 2, she avoids this statement altogether, but in Dialogue 1 states *yeah it's difficult to say but* (line 187) at this point in the story. In Dialogue 2 she does explain this part again, but also has difficulty with it (see lines 076 to 080 of Dialogue 2).

Extract 6.19

```
167  ASAMI but,
168        (606)
169        `yeah. [115 bpm]
170        `(500)
171        the `bicy`cle `is `bigger `than
172        (130)
```

```
173          the `boy,
174          (330)
175          and [=! laughing]
176          (385)
177          the bag of pears
178          (940)
179          is so heavy,
180          (825)
181          so
182          (414)
183          <it's:
184          (306)
185          looks hard to> [/]
186          (774)
187          um
188          (931)
189          it's looks hard to:
190          (1197)
191          um carry.
192          (1317)
193          yeah.
194          (218)
```

She finishes the complication segment with another sequence of mostly single-word utterances without establishing any rhythmic gestalt. Her difficulty with the story may also be shown by her hesitation on an appended vowel to *ride* in line 208, which is common among novice Japanese speakers of English (Carroll, 2005). In Asami's case, this strategy is rarely used in her retellings—that is, she both rarely appends vowels to final consonants, and rarely uses them as hesitation points.

Extract 6.20

```
195    ASAMI but
196          (395)
197          he robbed it,
198          (640)
199          and
200          (190)
201          then
202          (669)
203          <the &mæ-> [//]
204          (230)
205          the boy,
206          (1073)
207          &=laughs
208          ɾaidu@u [: ride]
209          (970)
210          a bicycle,
```

Asami uses few discourse markers in this segment. She begins the segment with *and then*, which is common in her stories. She also frequently uses the marker *yeah.* In line 103, this occurs at a change in the focus of the story. A similar use of *yeah* seems to occur in line 111. In line 135 and 141 *yeah* seems to be self-directed talk, i.e., a statement that she has decided what she wants to say. This could also be true for the case in line 103 and 111. The examples in lines 169 and 193 also seem to support the latter interpretation.

6.6 Asami Monologue 2

Asami starts the complication segment of Monologue 2 similarly to the others, with a long 775 msec pause, and the discourse marker *and* and *next.* The first phrase of this segment consists of three utterances, and seems to establish a rhythm of approximately 116 bpm, with every word being spoken on a beat and single silent beats in the intervening pauses (lines 015-019).

The following phrase (lines 021-023) is uttered at a faster rhythm of approximately 120 bpm—again, with every word except *the* occurring on a beat. However, the next phrase *as big as the boy* is slower (approximately 102 bpm), which may reflect some difficulty with this expression. This expression is used in Dialogue 1, where it seems to be used with more difficulty, and in Dialogue 2, where it is used more fluently. This demonstrates that there is some improvement with practice over the course of the retellings.

Extract 6.21

```
ASAMI (775)
     an(d)
     (150)
     next
     (570)
     a `small `boy, [112 bpm]
     (700)
     `showed `up, `with
     `(588)
     `riding `bicycle,
```

```
020         `(540)
021         the `bicycle `is very
022         (255)
023         big.
024         (495)
025         `as `big `as the `boy. [102 bpm]
026         (910)
```

The next phrase begins with the discourse marker *and* and a very long 1794 msec pause. Asami appears to have some difficulty with determiner selection in line 029, where she corrects *the* to *a.* The previously established rhythm is lost due to these hesitations, and is not established again until lines 031-033, with a tempo of 102 bpm again. This pattern continues through the next utterance (line 035) until it is interrupted by a 570 msec hesitation that occurs in the middle of the word *cannot.* This hesitation is likely due to lexical choice between the words *watch* and *look*. In this retelling, she self-corrects *watch* to *look*. She made this same self-correction in Dialogue 1 (she did not mention this point in Monologue 1), where she also had difficulty with the phrase *back of him.* She uses this phrase again here but with a phrase-internal pause (lines 041-043). In fact, she uses a falling boundary tone on *back*, which would imply that she intended to finish at that point. This may explain why the pause in line 042 is so long (600 msec).

Extract 6.22

```
027   ASAMI a:nd
028         (1794)
029         <the-> [//] a fat man
030         (780)
031         `is
032         (235)
033         `gathering `pears `on the `ladder, [102 bpm]
034         (296)
035         `so <`he can`no^(570)t
036         (718)
037         watch> [//]
038         (706)
039         he cannot look
040         (379)
041         back.
042         (600)
043         of him,
044         (792)
```

The next phrase begins with a discourse marker and filled pause *and uh* in line 045, as well as a long 988 msec pause. She also seems to have some difficulty here, and uses *uh* to fill some of the silence, but then has a 741 msec pause in the middle of the word *find*, which she makes into two syllables by adding a vowel to the end of the word. She also has a self-correction for determiner choice (lines 051-053) and repetition to clarify content (lines 053-055). These utterances are interspersed with pauses and too short to create a rhythmic pattern. The utterance in line 053 is spoken at a tempo of 82 bpm, but the following utterance in line 055 is faster at a tempo of 105 bpm, precluding the establishment of a rhythmic gestalt.

Extract 6.23

```
045   ASAMI a:nd uh
046         (988)
047         a small boy,
048         (1363)
049         u:h &fain^(741)dʊ: [: find]
050         (314)
051         a:
052         (348)
053         some box of <pears,> [//]
054         (394)
055         gathered pears,
056         (569)
```

The phrase consisting of lines 057 to 065 is spoken very much one word at a time at a rhythm with a tempo of 102 bpm, but with a silent beat in line 060, and two missed beats in lines 063-064, which may lose the rhythmic pattern. The utterances in lines 067 and 069 establish a rhythmic pattern of 115 bpm, starting from *fat man* in line 067 and with one silent beat in the 550 msec pause of line 068. This pattern is maintained somewhat with *he* in line 071, although *because* starts slightly late, leaving the 155 msec pause. The falling intonation on *that* in line 069 would indicate that she intended to stop, but she does not take a breath in the short pause, perhaps showing that she changed her mind as she was speaking. After one pause of silence, the second syllable of *cannot* in line 073 falls on the beat, but the rhythm is then lost with the delay of the verb in line 075 after a 791 msec pause.

The final phrase of this segment (line 077-085) starts again with the discourse marker *and* followed by a filled pause *uh*. No clear rhythm is established with the three utterances of this C-unit, due particularly to the long 782 msec pause in line 082.

Extract 6.24

```
057   ASAMI `and `he: [102 bpm]
058         (185)
059         `robbed
060         `(649)
061         `a: `box
062         `(285)
063         `of
064         ``(991)
065         `pears.
066         (381)
067         but a `fat `man, [115 bpm]
068         `(550)
069         `didn't `notice `that.
070         (155)
071         be`cause `he
072         `(430)
073         can`not
074         (791)
075         look.
076         (749)
077         and,
078         (520)
079         u:h
080         (348)
081         a small boy,
082         (782)
083         left
084         (393)
085         there.
```

In Monologue 2, Asami only uses the discourse marker *and* to begin topical sections of the segment. She uses filled pauses infrequently—only three times, one following *and*, one beginning an utterance, and one alone.   Asami seems to vary the stressed syllables of words depending on the situation or maybe depending on the rhythm that she is speaking at. Generally, all words are spoken on a beat, but sometimes articles are not, and sometimes more than one syllable in a word is given its own beat.

6.7 Asami Dialogue 1

As with Keiko's retellings, Asami starts this segment of the story with a long pause of 1355 msec, although this is partly filled with the interlocutor's backchannel from the previous turn and an extended head nod, beginning on the backchannel, and continuing through Asami's *and then* in line 061 to begin the segment.

The first C-unit is from lines 061 to 069, and is composed of four utterances. She seems to have some difficulty producing this first part, as evidenced by the fact that she breaks eye contact in line 061 and looks slightly to her left, and her self-correction of *little* to *small* in the following utterance. She makes eye contact again briefly when she says *little* in line 063, and then makes continuous eye contact from *bicycle* in line 067. During this time the interlocutor does not backchannel or give any feedback until after line 067, where he makes a single, slow head nod that encompasses the pause and part of the following utterance. It appears that the listener does not nod until eye contact is established. Following the final utterance in the C-unit (line 069), there is a long 1130 msec pause. In the middle of this pause, the interlocutor nods twice to show comprehension of what has been said up to this point.

Extract 6.25

```
060         (535)
061   ASAMI and then,
                  * <- A looks away
062         (345)
063         uh: <a: little> [//]
                    * <- A resumes eye contact
064         (400)
065         a `small `boy, [120 bpm]
                * <- A looks away
                * <- hands palms down move down
                     * <- left hand wave to left side
066         (100)
```

```
067         `with `bicycle,
                  * <- A resumes eye contact
                  * < - gestures bicycle handles toward left
068         `(660)
             1====
069         `showed `up.
             =====+
070         ``(1130)
               2==+
```

This segment begins with no clear rhythm being established until line 065, where rhythm with a tempo of 120 bpm is established with four continuous beats in lines 065-067. Every word is stressed with the exception of the indefinite article in line 065. There is an empty beat in the 660 msec pause in line 068, and then the rhythm is continued with the two words in line 069. Unlike Keiko, who has more control over stressing content words in her speech, and de-emphasizing and reducing the non-stressed words, Asami seems to be stressing almost every word in her speech. The listener does not verbally respond during this section, but the beginning of the head nods in lines 068 and 070 appear to be timed with the rhythm established by the speaker.

The next phrase (lines 071-075) are also uttered at about 120 bpm, continuing the same rhythm as the previous phrase, after two silent beats during the 1130 msec pause in line 070. Again, Asami stresses every word except the definite article in line 073. Additionally, in line 071 she stresses the vowel added to the end of *and* to make two stressed syllables here. This is a common technique for hesitations by novice Japanese speakers of English (Carroll, 2005), but to my knowledge has not been documented as being used to maintain rhythm. In this case, Asami is looking to her left throughout this phrase until the word *bad* in line 075, which could be a sign of hesitation, however, the added vowel is not lengthened as would be expected for a hesitation. The restatement of *bad boy* in line 077 begins on the next beat of the previous rhythm after a single silent beat in line 076. The listener responds here with another head nod starting on the next beat in the rhythm in line 078.

Extract 6.26

```
071   ASAMI `æn`dʊ@u [: and]
               * <- looks away
072          (230)
073          the `boy `is
074          `(595)
075          `bad.
              * <- resumes eye contact
076          `(530)
077          bad `boy,
              * <- Asami nods type 2 leaning forward
078          (210)
              1===
079          be`cause,
             ======+
080          `(350)
081          um
             * <- looks away to left
082          `(360)
083          `duɾiŋ`gu@u [: during] [130 bpm]
084          `(395)
085          the `fat `man;
                   * <- right hand up in prep for next gesture
086          (465)
087          `gæða`dʊ@u [: gathered];
              * <- resumes eye contact
              * <- right hand gestures picking pears on 2 beats
088          (265)
089          `pea`zu@u [: pears],
090          `(705)
              2====
091          `he
              ==+
092          `(620)
093          `kʊdɪn`tu@u [: couldn't];
              * <- looks away down and left
094          (300)
095          <`look-> [//]
               * <- gesture pointing down and behind her
               * <- regains eye contact
096          (270)
097          `watch
              * <- same a previous gestures repeated twice
098          (520)
099          `uh
              * <- gesture with both hands pointing down and behind
              * <- looks down behind her right side
100          (385)
101          °nan@s:jpn darou@s:jpn.°
102          (730)
103          back of uh
              * <- NS leans forward and turns head
                   moving left ear closer
104          (450)
```

```
105          &hɪ- [//] him?
                        * <- quick single head nod
             3==+ <- NS as if to show comprehension or "ah!"
106          (265)
107    NS1    mmhm,
              2===+
108          (220)
```

The long filled pause *um* in line 081 results in three missed beats in lines 080-082, which appears to lose the previous rhythm, since the following four utterances establish a new rhythmic pattern with a tempo of approximately 130 bpm with the final vowels added to *during* in line 083 and *pears* in line 089 being stressed, and single silent beats during the pauses in lines 084, 086, and 088 (Figure 6.6). It should also be noted that three of these four utterances end with boundary tones in spite of the fact that the pauses are all phrase-internal. This may show that Asami's non-native speech is being actively constructed, and therefore each utterance (or intonation unit) of the four, although comprising one phrase and one "new idea" (Chafe, 1994), for a non-native speaker could really be considered to be four "new ideas", with each phrase being a lexical or grammatical accomplishment for the speaker. Immediately upon completing the word *pears* in line 089, the listener nods, which shows that he was able to correctly predict the speaker's turn completion point. This point is accompanied by syntactic and prosodic completion. The listener's response is again timed with the next beat in the rhythmic pattern.

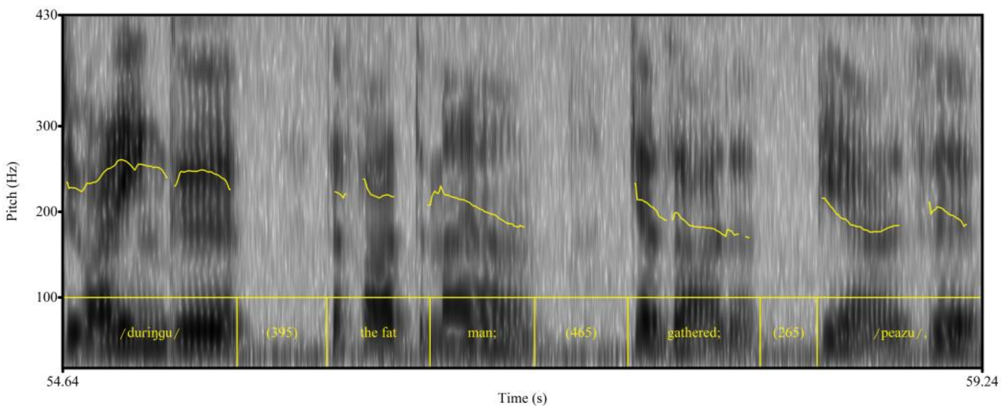

Figure 6.6 Pitch contour and spectrogram from Asami Dialogue 1, lines 083-089. Every word except the article *the* contributes to the rhythm as can be seen with the approximately equal spacings. The

three silent pauses each contain one missed beat. Importantly, the vowels appended to *during* and *pears* also contribute to the rhythm.

The next phrase in the segment shows some signs of difficulty. The previous rhythmic pattern is continued after the silent beat in line 090 until the Japanese aside in line 101. She uses gestures to support her speech by pointing with both hands behind her during lines 095-105 to demonstrate “behind him”, which she has difficulty producing. Asami finally comes up with *back of uh (450) hi- him*, finishing with a high rising intonation contour, in lines 103-105. The listener finally shows comprehension of this part with a nod in line 105, and then a verbal response of *mmhm.* This is the first verbal response in this part of the story, although the listener does respond with head nods, and these seem to be aligned with the speaker’s rhythm. This may be due to the rising boundary tone, which could signal an appeal to the listener.

In line 109, Asami uses *mm mm mm* as a marker, perhaps to assert to herself that what she has said was correct. Asami seems to briefly return to the earlier tempo of 120 with the stressed words in lines 113-117. This is also evidenced by the listener’s backchannel in line 119, which falls on the following beat in Asami’s pause. However, Asami’s response in line 121 is late and results in three missed beats. The rhythm is not clearly established during the rest of this C-unit.

In lines 121-131, Asami appears to have difficulty with the story, particularly with the word choice for *robbed*, which she speaks with falling intonation as if she has completed the phrase. Again, from the point of view of the learner focused on producing correct sentences, she has completed the task at hand, although not the phrase. This phrase also does not seem to establish a rhythmic pattern. There is a break in the flow of the utterances in line 123 with a restart, and the word *boy* is pronounced with two syllables, both stressed. Also, the Japanese aside in line 127 is spoken much faster than the rest. Even the longer utterance of line 131 has a slowly enunciated *of* that upsets the rhythm of the utterance. The listener's backchannel in line 133 comes after a 290 msec pause, which is a fairly long pause between speaker turns and shows no rhythmic alignment. Asami uses a two-hand pulling gestures to demonstrate *robbed* as she says that word in line 127.

Extract 6.27

```
109   ASAMI mm mm mm so,
             * <- A nods type 2
                       * <- looks away to left
                  2===+
110         (720)
111         <the-> [//]
112         (260)
113         `he: [120 bpm]
114         `(670)
115         `kʊdɪn`tu@u [: couldn't];
116         (240)
117         `noticed` the `boy,
             * <- resumes eye contact
                           * <- Asami head nod
118         (260)
119   NS1    mm`hm
             1===+
120         `(505)
121   ASAMI so
             * <- A looks away
122         (775)
123         <he-> [//] <the boy,> [//]
124         (305)
125         the small boy,
                 * <- resumes eye contact,
                 * <- gestures pointing downward
126         (908)
             * <- looks away
             * <- gestures taking something
127         °nan@s:jpn darou@s:jpn.° robbed.
                                    * <- resumes eye contact
128         (215)
129         the
130         (465)
131         `box `of `gathered [=! laughing] `pear, [100 bpm]
             * <- gestures big circle as shape of box
                       * <- open left hand palm in beat down
                 same gesture one more beat -> *
132         (290)
           2==== <- with smile & wince
133   NS1    mm┌hm, ┐
             =======
134   ASAMI   └&=la┘ughs
                  =====
135         (620)
             ===+
136         yes.
137         (572)
```

The next phrase (lines 138-147) begins with a high articulation rate utterance (line 140) which begins to establish a rhythmic pattern at a tempo of 130 bpm, but is lost due to the cut-off *big* at the end and the following Japanese aside. It should be noted that in this retelling (Asami's Dialogue 1), she uses Japanese seven times, with four of the occurrences being in this complication segment. This could be a sign of extraordinary difficulty explaining this segment, however, she does not use Japanese at all during any of the other retellings, including the previous retelling, Monologue 1. Another possibility is that she is having difficulty explaining herself to this particular listener. There is some evidence for this in that the listener's verbal responses tend to be fewer than the listener in Dialogue 2, and he suppresses a yawn during line 144.

Extract 6.28

```
138   ASAMI an(d)
             * <- looks away
139         (350)
140         <the bicycle is so big-> [//]
141         (300)
142         big tte@s:jpn iu@s:jpn ka@s:jpn ano@s:jpn
                1========================+
143         (340)
144         the `bicy`cle `is `as `big `æ`zʊ:@u [:as] [180 bpm]
                 * <- vague gesture, 2 hands in front, palms inward
                 * <- NS suppressed yawn
145         `(360)
146         `boy,
             * <- resumes eye contact
147         `(305)
             2====
148   NS1   mmhm,
             =====
149         (170)
             ====+
150   ASAMI so <it-> [//]
                 * <- looks away
151         (510)
152         °nan@s:jpn darou@s:jpn° `it `ʃɪmu`zʊ@u [: seems]
                                                   [130 bpm]
153         (175)
154         `diffi`cult `to `carry,
             * <- resumes eye contact
                            * <- gestures bicycle handles steering
155         (565)
```

```
156   NS1   `right ┌`yeah.┐
157   ASAMI        └`mm   ┘ mm mm.
                    * <- Asami rapid head nods
           2===================
158         `(430)
             ====+
```

The utterance in line 144 is one of the most temporally fluent in this segment with 6 words, 9 syllables, and an articulation of 3.81 syllables per second, which is higher the than average for the segment of 2.78. Asami establishes a rhythm with a fast tempo of approximately 150 bpm—with a syllable-timed rhythm, including a final vowel added to *as* in line 144. This rhythm continues after one silent beat to *boy* in line 146, and is maintained by the interlocutor’s backchannel in line 148, and the following utterance in line 150. Line 150 is cut-off, and after a fairly long pause, she interjects a Japanese phrase again. This phrase is probably self-directed, since she is not making eye contact at this point, only regaining eye contact when she continues in English with *difficult to carry* in the next utterance. These phrases, after the Japanese interjection, are spoken much more slowly than the previous tempo, and do not seem to establish a rhythmic pattern. In fact, there is some overlap with the listener’s *right yeah* and Asami’s *mm mm mm* in lines 156-157, which does not appear to be rhythmically aligned in any way. In this case, the overlap can be seen as a sign of the participants not being aligned rhythmically.

The following several lines show some difficulty again, with short utterances and signs of difficulty with lexical (pronoun) choice in lines 159-163. Following this is a series of one and two word utterances which fail to establish any clear rhythmic pattern. Asami also uses gestures extensively in this section, illustrating *robbed* in line 163, outlining the shape of a box in line 167, and finally demonstrating holding the box in line 169. There is one backchannel by the listener, but it slightly overlaps with the ending of *pears* in line 169, again showing a lack of rhythmic alignment.

Extract 6.29

```
159   ASAMI `but <`she> [//]
               * <- looks away
                     * <- hands move to left side prep
                          for taking, remain there
160        (195)
161        uh  <he-> [//]
                  * <- resumes eye contact quickly
162        (290)
163        <he-> [//] the `boy `robbed [95 bpm]
                               3=======+ <- NS quick head raise,
                                     preceded with lean forward
                                  * <- gesture taking something
164        (125)
165        `the
166        (295)
167        `big box
            * <- gesture 2 hands outline big box shape
168        `(360)
            2====
169        `of pea┌`zu@u [: pears],┐
            =+ * <- gestures 2 hands holding box
170   NS1          └                mm┘hm,
                                    1===+
171        (355)
```

Similarly, the next phrase (lines 172-175) is spoken in short utterances that break up the clausal structure and fail to establish any clear rhythmic pattern, mainly due to the slow articulation rate and the long pauses, including a very long 1285 msec pause. Lines 176-178 may establish a rhythm at a tempo of 120 bpm again, with a silent beat in the phrase-internal pause in line 177. The listener responds with a head nod after the end of the phrase (lines 178-179), and both speakers come in on the next beat (lines 180-181), which may show rhythmic alignment. The utterance in line 181 is longer, and again is spoken in a syllable-timed manner at a faster tempo of about 150 bpm, with all syllables spoken with equal stress, although *the* is shorter, acting as a pick-up to the beat on the first syllable of *bicycle*, and a short middle syllable in *bicycle.* The following two syllables *and go* in line 183 continue this rhythm after two silent beats (at 165 bpm).

Extract 6.30

```
172   ASAMI so he
             * <- looks away
                * <- begins long gesture pick up box
173         (635)
174         bring
             * <- resumes eye contact
             * <- completes gesture put on bicycle
175         (1285)
176         `pears `box `on [110 bpm]
              * <- gestures bicycle handle and hold through
177         `(685)
178         `bicycle,
              * <- gesture bicycle handle downbeat
179         `(760)
             5==== <- intense then multiple small nods
180   NS1    ┌mmhm,┐
181   ASAMI  └`and ┘ `he `ride `on `the `bicy`cle, [150 bpm]
                         * <- RH short wave to right
                               * <- return to bicycle handles
                                         * <- gestures steering
                                              bicycle handles
             =====+
182         `(760)
               * <- hands move to tabletop
183   ASAMI `and `go.
                   * <- left index finger draws line on table
184         &=laughs
185         yeah [=! laughing]
            2=== <- smile & wince
186         (275)
            =====
187         yeah ┌it's dif┐ficult to say but
188   NS1        └okay.   ┘
            ========+
189         (103)
190   ASAMI mm.
             * <- looks down at table
```

The final phrase in this segment is made up of two utterances, with more clearly distinguishable stressed words (containing pitch transients). Lines 185-189 do not appear to have a clear rhythmic pattern due to the extended laughter. The overlap in lines 187-188 may be a sign of this lack of clear rhythm.

Asami's use of discourse markers in Dialogue 1 is similar to Monologue 1. She begins the segment with *and then*, and uses *and* and *so* to connect points of the segment. Similar to her monologue, in line 109 she uses *mm mm mm* as a kind of answer to her own doubt expressed by the rising intonation in line 105, which seems to be an appeal to check comprehension of the phrase *watch back of him* (meaning "see behind him"). This difficulty is clear from the Japanese aside in line 101, and is then responded to by the listener in line 107. She uses *yeah,* a common discourse marker for her, at the end of the segment to emphasize the main point (the stealing of the pears), and again in line 187 to introduce her explanation of the difficulty of explaining the story in the segment. Finally, she finishes the segment with *mm.*

6.8 Asami Dialogue 2

The complication segment of Asami's Dialogue 2 begins similarly to the other retellings with the discourse marker *and then*. All of these segments in her retellings begin with *and then*, except for Monologue 2, *and (150) next.* In this segment, not unlike her others, she uses *and* to introduce some sub-topics. She also uses *yeah* and *mm* as the only other smallwords or discourse markers. Unlike the monologues and Dialogue 1, in this retelling *yeah* and *mm mm mm* generally follow a response from the listener (see lines 082-085, 103-104, and 112-113), while in the other retellings these discourse markers appeared to be used more as a kind of self-confirmation of what she had said.

Extract 6.31

```
061         (229)
062   ASAMI and then,
             * <- A looks away
063         (271)
064         a `small `boy, [116 bpm]
                * <- A resumes eye contact
                     4===
065         `(315)
             =====
066         showed `up,
            ==========
067   NS2    mm`hm,
             =====+
068   ASAMI with `bicycle.
                   * <- A gestures bicycle handles
069   NS2    (203)
            2====
070         o`kay,
            =====
071         (103)
            =====
```

The first phrase of this segment appears to set up a rhythmic pattern at a tempo of 116 bpm with the stressed words in lines 064-068 and single silent beats in the intervening pauses. The backchannel by the listener in line 067 appears to be aligned with this rhythm and falls on the single beat between *up* in line 066 and the first syllable of *bicycle* in line 068. The response in line 070 also appears to fall on the next beat after a single silent beat.

The following utterance (line 072) is the longest run in this segment (eight words), and appeared difficult in Monologue 2 and Dialogue 1. In this retelling, the utterance is spoken more smoothly as a single run, but the rhythm is established with every word except articles spoken on the beat. This rhythm is at a tempo of approximately 94 bpm, slower than the previous tempo of 120 bpm. This may show that although there is some improvement here over the course of the retellings, it still shows signs of difficulty.

Extract 6.32

```
072   ASAMI the `bicycle `is `as big `as the `boy, [94 bpm]
               * <- looks away     * <- resumes eye contact
                                   * <- 2 hands beat down together
                        2 hands wave to left -> *
             ===============+                4==
073   NS2    (310)
             =====
074          `mm!
             ===
075          (110)
            =====
076   ASAMI `so <it `seems
               * <- A looks away
             ============+
077          (530)
078          uh to-> [//]
079          (185)
080          it seems hard to carry,
             * <- A resumes eye contact
                * <- 2 hands small downbeat
                              * <- gestures bicycle handles
                      5============
081          (110)
            =====
082   NS2    &=laughs
            =======+ <- nod intensifies during laughter
083   ASAMI &=laughs
084          (320)
```

```
085          yeah
086          (175)
```

After this utterance, there is a single silent beat, and then the listener responds with *mm*, which is spoken with a rise-fall pitch transient. The peak of this pitch transient falls on the next beat, which lends more evidence to the alignment of rhythm between the speaker and listener. The listener does a sharp upward nod (during a long sequence of mild nods) at the point of the silent beat in line 073. From the video, I interpret this nod as a non-verbal version of *oh!* said for surprise and to show understanding. This example appears to show that head nods and gestures may also be rhythmically aligned.

Line 076 appears to continue the previous rhythm, however, after this, the rhythm appears to be lost due to the filled pause *uh* and the restart in line 078. Line 080 is uttered slowly and does not appear to establish a rhythmic gestalt.

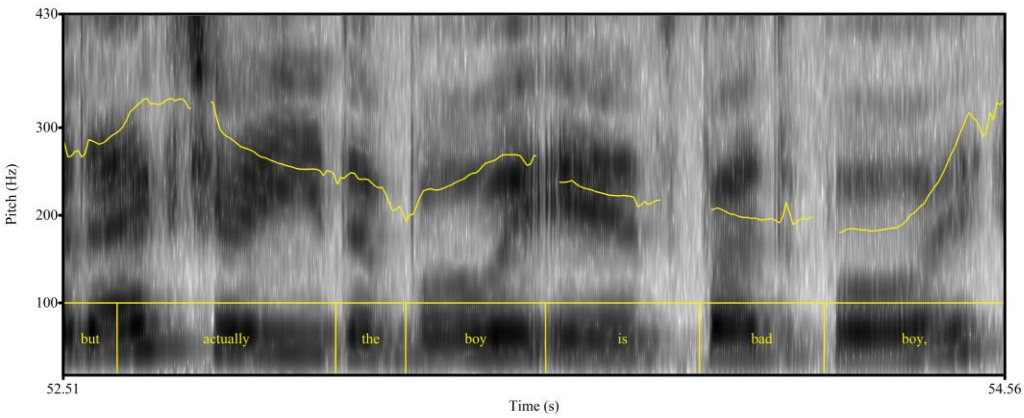

Figure 6.7 Pitch contour and spectrogram from Asami Dialogue 2 line 087. This example is relatively more fluent in temporal terms, and illustrates more stress-timing. The rhythm here is maintained by *actually*, *boy*, and *bad.* The first two have high pitch accents.

The next utterance (line 087) is a relatively fluent one in terms of length (seven words), and contains three stressed words, which sets up a rhythmic pattern at a tempo of 90 bpm. In this case the stressed words have more clearly visible pitch contours (Figure 6.7). *Bad boy* seems to be spoken as a single word, with the first syllable stressed. This is continued in the next utterance (line 089) with the stressed syllable in *because*. The listener also responds with *oh*, which is aligned with the beat, and therefore overlaps Asami's *because*. It is also interesting to note that the listener's head nod begins earlier than the verbal response, but exactly as Asami's previous utterance ends—marked with a low-rise boundary tone. In this example, the head nod appears to begin at the point of the silent beat after the utterance in line 087, and ends with the listener's verbal response in line 090, which may show a rhythmic alignment of head nodding as well. Most of the head nods by the listener appear to begin at the same time as stressed words (see lines 064, 072, 080) or immediately after a previous utterance ends, corresponding to a silent beat (lines 069, 088). In this particular case, the prosodic boundary tone may signal a TRP to the listener, who then responds on the following beat. Asami, by not leaving the space open, ends up overlapping.

Extract 6.33

```
087   ASAMI but `actually the `boy is `bad boy, [90 bpm]
                                   * <- left hand point toward table
                                            * <- repeat same gesture
088         (271)
           3====
089         be┌`cause,┐
090   NS2     └`oh!   ┘
           ======+
091         (260)
092   ASAMI `he: [//]
              * <- looks away to left
093         ``(773)
094         he `robbed the `box of `gathered `pears,
            * <- left hand point down towards table
                 * <- resumes eye contact, gesture 2 hands
                                   pull away to right side
                            * <- gestures outline of box
                       3===================+    3===
095   NS2    (250)
             =====
096         o`kay,
            =====
097   ASAMI but a `fat man [90 bpm]
                    * <- gestures 2 hands down to right side
             * <- looks up to left
            ====+
098         `(443)
099         `couldn't
100         `(622)
              * <- resumes eye contact
             2==
```

```
101          `notice.
              =======
102   NS2    (99)
              ====
103          m┌`m:.  ┐((.pt))
104   ASAMI   └`mm mm┘ mm mm ┌ mm.┐
105   NS2     *              └`i  ┘ see. <- NS wince;
                                     A rapid nods type 2
             ==========================
106          `(325)
              =====
```

In line 092, *he* continues the previous rhythm. However, she lengthens this word as a hesitation, misses two beats, and then continues on the same rhythm with *robbed* in line 094. The listener responds on the next beat following the utterance of line 094 with *okay*, stressed on the first syllable. Line 097 involves a slight hesitation on *but* which delays the rhythmic pattern, i.e., it restarts the rhythm at the same tempo slightly later. The single-word utterances in lines 099 and 101, *couldn't* and *notice*, continue on the same rhythmic pattern with single silent beats intervening in the pauses of lines 098 and 100.

The head nods in lines 094-097 also seem to be aligned with the rhythm. The listener raises her head with a peak corresponding to the beat where Asami says *box*, and then falls on the next beat, corresponding to the first syllable of *gathered*. The following nod rises to peak when the listener pronounces the first syllable of *okay.*

After Asami says *notice* with falling intonation in line 101, the listener backchannels on the next beat with *mm*. After a single silent beat, the listener adds *i see* on the next beat (stressing the first word).

Extract 6.34

```
107   ASAMI `and he `ride `bicycle `and leave `there,
              =+           * <- A gesture bicycle handles
                            left hand points -> *
                            to left side
108   NS2    (215)
109          oh ┌`no.  ┐
110   ASAMI     └`mm mm┘ mm.
                   * <- A nods type 2
```

```
111   NS2   (224)
112         huh.
113   ASAMI yeah.
```

The last substantive utterance of the segment is line 107, where the beat seems to be established with non-content words in *`and he `ride `bicycle `and leave `there*. These do not all appear to be stressed (that is, showing a pitch transient), and the last three words, *and leave there*, are pronounced with equal syllable timing. This seems to continue the previous rhythm, maintained by the listener's responses. The rhythmic pattern from line 107 is maintained by the first word of the listener's response *oh no* in line 109, and then by the last syllable of Asami's *mm mm mm* in line 110, and then again with *yeah* in line 113, which is almost (but not) aligned with the listener's short *huh* in line 112.

Overall, this retelling seems to show much more signs of rhythm in the speaker's narrative and of rhythmic alignment between the speakers. There also seems to be more rapport between the participants than in Asami's other dialogue, where the listener was less reactive and seemed bored.

6.9 Summary

This chapter examined each of the eight Pear Story retellings line by line, in order to examine how factors of dialogic and monologic fluency interact in the co-construction of each narrative. The next chapter will extract themes from this data in order to move towards a model of dialogic fluency.

7 Trends in the co-construction of fluency

7.1 Introduction

In this chapter, I would like to bring together the findings of Chapters 5 and 6, in order to draw conclusions about the research questions posed in Chapter 3. To review, the main purpose of this study was to examine the extent to which co-constructed features of fluency should be taken into account in definitions and frameworks of fluency, and whether concepts such as McCarthy's *confluence* are necessary for a more complete definition of fluency. Therefore, this study investigated, through a detailed case study approach, how different factors interact to create a fluent (or confluent) conversation between a learner of English and a native speaker. These research questions were a result of the observation that Asami, the lower-proficiency subject in the parallel case studies, was able to create a strong impression of fluency on native-speaker interlocutors in spite of having relatively weaker values of temporal measures traditionally associated with fluency, such as speaking rate and length of uninterrupted runs. Therefore, by comparing Asami with a higher-proficiency learner with higher measures of monologic fluency, the factors that Asami employs to compensate for her weaker monologic fluency were examined. This chapter will look at aspects of dialogic fluency, such as the use of discourse markers, non-verbal communication, rhythmic alignment, and speech management strategies, and how they function within a whole system to create a fluent, co-constructed conversation.

7.2 Rhythmic alignment

In general, it appears that Keiko was able to establish rhythmic gestalts more consistently then the lower-level learner, Asami. For both of Keiko's monologues, approximately half of the segment featured discernible rhythmic patterns. On the other hand, for Keiko's dialogues, almost the entire segment was associated with discernible rhythm, although the rhythm was sometimes restarted after short delays, and the tempo was occasionally changed. Keiko's speech tended to be at tempos of 90 bpm, with some jumps to 100-107 bpm. An exceptionally high tempo occurred during the second half of the segment from Dialogue 2, where Keiko spoke at a tempo of 120 bpm. This was achieved, not by increasing her speech rate or articulation rate, but rather by using more frequent accents—something between strict syllable-timing and stress-timing. Unlike speech rate and articulation rate, which were lower in the monologues than the dialogues, the tempos did not appear to vary according to the genre.

Asami, on the other hand, appeared to have somewhat more difficulty establishing clear rhythmic patterns in her narratives. Her first monologue had only two sections where a rhythmic pattern was maintained for more than two or three utterances, and the majority of the segment had no discernible rhythm. Her second monologue was more rhythmic, but was still less than half associated with a discernible pattern. This was due to the tendency toward short utterances of one or two words, and long silent pauses. Less than half of her first dialogue was associated with a discernible rhythm, however, rhythm was maintained through most of the segment in the second dialogue. This could reflect the practice that Asami had over the course of these retellings, with the first one, Monologue 1, being the least rhythmic and the last one, Dialogue 2, being the most rhythmic. Like Keiko, there is also a tendency for the dialogues to have more consistent rhythm. Asami's pace was faster than Keiko's, at tempos of 102-108 bpm, 112-120 bpm, and occasional forays up to 130 and 150 bpm (in Dialogue 1). This was surprising in that her speech rate and articulation rate were significantly lower than Keiko's. This seems to be reflected by the manner in which the two speakers created rhythm. Asami tended to stress more words in each phrase than Keiko. This could be described as more of a "syllable-timed" manner of speaking, however, she tended to stress only the main syllable of each word. Generally, articles were not stressed, but stress on other function words seemed to vary—some words being stressed in one utterance, but not in another. The following example illustrates Asami's stress pattern, with beats marked with grave accents before the relevant syllable. Every word except for the articles and auxiliary verbs are stressed.

Extract 7.1 Asami Monologue 1

```
115 ASAMI the `man `who [105 bpm]
116       `(477)
117       are `taking
118       `(403)
119       `pears `during the `movie
120       `(433)
121       <`is
122       `(770)
123       u:m`
124       `(165)
125       `three `bags> [//]
```

In the pilot study of Chapter 3, it was found that Asami's speech was often aligned rhythmically with the interlocutor, even though the two spoke at quite different speech rates. To illustrate, Asami might produce a three-word utterance with all three words stressed at a tempo of 100 bpm, while the interlocutor might produce a nine-word utterance with three words stressed at the same tempo of 100 bpm. This results in comparatively longer runs and higher speech rates for the interlocutor, while at the same time maintaining the same aligned rhythm. Although the interlocutors in the Pear Story narratives spoke very little, the alignment of their responses supports this hypothesis from the pilot study.

In contrast to Asami, Keiko stressed fewer words, as shown in the following example, where the rhythm was maintained with single silent beats in the pauses. Keiko's tempo in this example was 90 bpm, with a speech rate of 128 words per minute, while Asami in the example above had a higher tempo of 105 bpm, but a much lower speech rate of 90 words per minute.

Extract 7.2 Keiko Monologue 1

```
075 KEIKO he `tries to `carry the `basket. [90 bpm]
076       `(761)
077       on the `bicycle.
078       `(645)
079       but the `man didn't `realize it.
080       `(605)
081       `an(d)
082       `(245)
083       `he (i)s just `riding
084       `(416)
085       on the `bicycle,
```

The relative lack of rhythm in the monologues appears to be a mainly caused by short utterances and long pauses. Longer utterances tend to result in longer sequences of stressed words which facilitates the establishment of a rhythmic gestalt. Keiko tended to have longer runs (mean length of run is 3.16 words for Monologue 1 and 2.89 words for Monologue 2) compared to Asami (mean length of run is 1.94 for Monologue 1 and 2.26 for Monologue 2). Length of run is not the entire story, however. Very short pauses (generally less than 200 msec) may fall between beats, allowing multiple short utterances to add up to a longer continuous rhythm, as in the following extract from Asami's Monologue 1, where lines 079 to 083 result in six continuous beats at 108 bpm, in spite of the two short pauses.

Extract 7.3 Asami Monologue 1

```
079 ASAMI a `boy, [///]
080       (110)
081       a `little `boy,
082       (170)
083       `maybe `he `is
084       `(670)
085       `eight or `ni:ne,
086       (1273)
```

The presence of restarts, cut-off words, and long pauses seemed to account for many of the times that the rhythmic pattern is lost. Again, in the above example, the rhythmic pattern was disrupted due to the lengthening of the vowel in line 085 and the 1273 msec silence in line 086. On the other hand, this example also demonstrates that the presence of apparent dysfluencies may not necessarily result in a breakdown of the rhythm. Asami handled the restart in lines 079-081 without missing a beat. Furthermore, as in both examples above, single silent beats did not appear to disrupt the rhythm.

Throughout the retellings in both monologue and dialogue, both subjects showed fluctuations in the tempo established by the rhythm of their speech. Bohle (2007; cited in Selting, 2013), noted that changes in tempo can occur with the start of new turn construction units. This appears to be true for Keiko, where changes in tempo tend to coincide with the beginning of new C-units. The exceptions seem to be associated with difficulties with the production of the narrative, as in the following example from Dialogue 1 where she slowed from 100 bpm to 90 bpm (line 113) while using high-rising boundary tones apparently to appeal for help or confirmation from the listener. This could also be an intentional slowing down to signal uncertainty to the listener, as she slowed down at this same point in Dialogue 2.

Extract 7.4 Keiko Dialogue 1

```
106   KEIKO `and `he:` just `put the [100 bpm]
                   * <- K looks up & left
             * <- hands return to front of chest together
107         (185)
108         `bicycle on the `ground,
             * <- K resumes eye contact
             * <- 2 hands palms inward beat
                             * <- 2 hands downward her right side
                      1==========+
109         (235)
110   NS3   mm┌`hm┐
111   KEIKO   └  a┘nd
             2====+
112         `(555)
113         maybe `stole? [90 bpm]
114         (220)
115         a `basket?
               * <- gestures 2 hands holding basket
116         `(320)
117         full of `pears?
```

Asami tended to stress more words than Keiko and generally had higher tempos. Her monologues showed only small variations of between 102 and 115 bpm. On the other hand, her first dialogue showed much more variation in tempo, from 100 to 180 bpm. Difficult parts seem to be associated with higher tempos and the stressing of more syllables. The fastest tempo in Dialogue 1 of 180 bpm comes right after speaking a rapid phrase in Japanese to herself, where she stresses almost every syllable in the phrase (line 144 below)—including a vowel added to the word *as*. Otherwise, tempo changes also seem to occur at junctures between C-units. In Dialogue 2, Asami is fairly consistent with tempo throughout the segment.

Extract 7.5 Asami Dialogue 1

```
140   ASAMI <the bicycle is so big-> [//]
141           (300)
142           big tte@s:jpn iu@s:jpn ka@s:jpn ano@s:jpn
                  1=========================+
143           (340)
144           the `bicy`cle `is `as `big `æ`zʊ:@u [:as] [180 bpm]
```

In spite of the variation in tempo that occurs frequently within the narratives, these data show that listeners are able to maintain rhythmic alignment with the speaker and keep up with these changes in tempo. In the following extract, from Keiko's Dialogue 1, a tempo of 90 bpm is established from line 088 to line 098, with aligned listener responses in lines 092 and 098. When the speaker speeds up the tempo in line 100, the listener responds in accordance with this new tempo in line 105 and 110.

Extract 7.6 Keiko Dialogue 1

```
086   KEIKO a:n(d)
             * <- K looks down & right
087         (725)
088         and `then
089         (95)
090         a `boy, [90 bpm]
            * <- K resumes eye contact
091         (255)
             1==
092   NS3   mm`hm,
             =====
093         (65)
             ===+
094   KEIKO is `coming <on the:> [//]
                 * <- gestures right hand move from
                      right side pointing
                      inward to meet left hand in front
                        * <- K looks away briefly
095         ``(205)
096         on a `bicycle.
                  * <- 2 hands together beat down
097         `(340)
             2== <- nod preceding by eyes up & left
098   NS3   mm`hm,
             =====
099         (163)
            =====
100   KEIKO the `bicycle is `really `big, [100 bpm]
                  * <- gesture 2 hand palms inward
                                     * <- hands spread outlining
                                          a big round shape
             ==+
101         (215)
102   NS3   &=laughs
103   KEIKO `bigger than `him,
              * <- 2 hand beats on syllables, move right on "him"
                        2=== <- nod with smile
104         (140)
            =====
105   NS3   mm`hm,
             ====+
106   KEIKO `and `he:` just `put the
                   * <- K looks up & left
             * <- hands return to front of chest together
107         (185)
```

```
108          `bicycle on the `ground,
              * <- K resumes eye contact
              * <- 2 hands palms inward beat
                              * <- 2 hands downward her right side
                       1==========+
109          (235)
110   NS3    mm┌`hm┐
111   KEIKO    └  a┘nd
              2====+
```

To summarize, both Asami and Keiko are able to establish rhythmic gestalts in their speech that can be aligned to by the listeners. The higher-proficiency learner here has an advantage because of the ability to produce longer utterances. Rhythmic gestalts can persist across some pauses and through other hesitation phenomena in speech, although points of difficulty are often the source of the rhythm being lost. Finally, it appears that the way learners establish rhythm can vary depending on the proficiency of the learner (with more advanced learners stressing fewer words and less advanced learners stressing more words), as well as within a given speaker's speech depending on the difficulty they experience with language production from moment to moment. This is in line with Arvaniti's (2009) conclusion that the difference between stress-timing and syllable-timing is a continuum rather than a dichotomy.

7.3 Discourse markers

Keiko and Asami appear to use some discourse markers differently. Both of them begin major segments with similar markers, such as *and then*, however, their use of *yeah* in particular is quite different. Keiko uses this discourse marker in response to engagement and information receipt listener responses, such as *oh* and *wow* (below, lines 118-119).

Extract 7.7 Keiko Dialogue 1

```
113   KEIKO maybe `stole? [90 bpm]
114          (220)
115          a `basket?
                * <- gestures 2 hands holding basket
```

```
116         `(320)
117         full of `pears?
                     * <- same gesture basket, beat downwards
                     * <- NS pulls head back and sidelong glance
118   NS3    oh!
119   KEIKO a `basket `yeah.
               * <- another beat with same gesture
```

There was only one case of these types of listener responses that was not verbally acknowledged by Keiko. In Dialogue 1 line 129, she does not acknowledge *ah* by the listener (see below), but in this case the listener response overlaps her own speech. Keiko does not use these discourse markers at all in her monologues, which confirms their use as acknowledgements of listener responses.

Extract 7.8 Keiko Dialogue 1

```
124   KEIKO the basket on the-
125         (365)
126         bicycle?
127         (534)
128   NS3   ┌ah!           ┐
129   KEIKO └and then just┘ gone,
```

Asami, on the other hand, uses *yeah* many times during her first monologue, although not at all in her second monologue. She also uses *yeah* in both of her dialogues, as well as *yes* and *mm mm mm*, which appear to be used similarly. Most of the uses appear as breaks between minor sections or points of the story, such as in the extract below from her first monologue (lines 103 and 111).

Extract 7.9 Asami Monologue 1

```
097   ASAMI the bicycle
098         (155)
099         is bigger than
100         (575)
101         the boy?
102         (570)
103         yeah.
104         (426)
105         an(d)
106         (185)
107         he found
108         (225)
109         the pears: bag,
110         (466)
111         yeah.
112         (293)
113         `actually [105 bpm]
```

```
114        (278)
115        the `man `who
116        `(477)
117        are `taking
118        `(403)
119        `pears `during the `movie
```

However, she also seems to use these at points of difficulty in the story. In Monologue 1 (lines 133-134 below), there is a long hesitation with a lengthened vowel on *to* and a 1249 msec pause. In line 135, she seems to have decided how to complete the phrase and begins the utterance with *yeah*. Again, in line 141, she uses *yeah* before restating the previous difficult phrase. It may be that these uses of *yeah* are essentially all used by Asami to assert herself in the narrative, either in terms of completing a C-unit, or of completing a difficult phrase in the story.

Extract 7.10 Asami Monologue 1

```
129        the man has
130        (337)
131        three bags
132        (307)
133        to:
134        (1249)
135        yeah. put in
136        (240)
137        &θə- [///]
138        (360)
139        many pears,
140        (367)
141        yeah. so there a:re three bags.
```

In Dialogue 2, there are several engagement and information receipt listener responses, but unlike Keiko, there is often no acknowledgement from Asami or Asami's response appears to be overlapped with the listener response, implying that it is not a reaction to the listener response. There are a few cases that could be acknowledge of the listener's reaction. She uses *yeah* in two cases after the listener laughs and after the listener response *huh* at the end of Dialogue 2.

To summarize, it seems that Asami uses these discourse markers in more of a self-directed way and perhaps as a filler between phrases, and sometimes when she has worked out difficult parts of her narrative. Keiko, on the other hand, uses this discourse marker in a way that interacts more with the listener. This use of discourse markers for interaction, interacting more closely with the other participant, is likely a feature of more advanced speakers of English. Generalizations cannot be drawn from a case study such as this, but it seems clear that this is a feature that Keiko is able to use to create more alignment in the conversation, to a much better extent than Asami.

7.4 Speech management strategies

Filled pauses are an important method of filling silence at less fluent points in speech, and have been associated with warning the listener of upcoming hesitations (Fox Tree, 2001) and of upcoming complex constructions (Watanabe, Hirose, Den, & Minematsu, 2008). Most of Asami's uses of *uh* and *um* are associated with long pauses in the narrative, where the filled pause is used to break and partially fill what would be an extended silence. There are also several cases of very long pauses that are not filled at all, which would be expected of a lower-proficiency learner. However, generally, Asami uses filled pauses to prevent these silences when trouble arises. That said, all of her filled pauses are associated with difficulties such as in the following example from her first dialogue. She is having difficulty choosing the words to express "see behind him" in lines 095-105. She uses a filled pause in line 099 to break what would be a 1485 msec silence. This filled pause signals the difficulty she is having completing the phrase, further shown by her Japanese aside where she asks herself, "How do you say that?" The completion in line 103 also contains a filled pause warning the listener of the difficulty with completing the prepositional phrase in line 105. Her difficulties finding the expression she needs may be a sign of weakness in cognitive fluency, and could negatively affect her perceived fluency, however, her use of filled pauses is not unlike a native speaker (Tottie, 2011), and could benefit her impression of fluency on a listener.

Extract 7.11 Asami Dialogue 1

```
091   ASAMI`he
            ==+
092        `(620)
093        `kʊdɪn`tu@u [: couldn't];
            * <- looks away down and left
094        (300)
095        <`look-> [//]
            * <- gesture pointing down and behind her
            * <- regains eye contact
096        (270)
097        `watch
            * <- same a previous gestures repeated twice
098        (520)
099        `uh
           * <- gesture with both hands pointing down and behind
           * <- looks down behind her right side
100        (385)
101        °nan@s:jpn darou@s:jpn.°
102        (730)
103        back of uh
           * <- NS leans forward and turns head
                moving left ear closer
104        (450)
105        &hɪ- [//] him?
```

Similar to Asami, most of Keiko's filled pauses are associated with some type of difficulty with the production of the narrative, and used to break up long pauses. However, she also uses them in most cases where restarts occur in her speech, as shown in the following example from her Monologue 1. In line 069, the cut-off phrase *full of a* is immediately following by a filled pause before the correction, *full of pears*. This is a typical use of filled pauses that is very nativelike (Fox Tree, 2001).

Extract 7.12 Keiko Monologue 1

```
063   KEIKO an(d)`
064        `(593)
065        he `pick
066        `(405)
067        a `basket.
068        `(535)
069        <`full of a-> [//] uh full of pears,
```

Keiko also seems to be able to use filled pauses to warn the listener of upcoming complex syntax, as was found in Watanabe, et al. (2008), although that study was done on Japanese rather than English. In both Keiko's Monologue 2 and Dialogue 2, she uses *uh* to introduce an embedded clause in the same place in the narrative. In Monologue 2 this is realized as *a boy, (610) uh riding on (170) a bicycle is coming*, and in Dialogue 2 this is realized as *a boy, (220) uh riding on the bicycle? (205) is coming*. Therefore, the filled pauses in these two examples are not unlike native speaker usage. On the other hand, the silent pause in the monologue example is in a disruptive location (Lennon, 1990), while the silent pause in the dialogue example does not break up a phrase. It would be more likely for a highly proficient speaker of English to use a filled pause, rather than a silence, in the disruptive location of the monologue example (Kjellmer, 2003). In summary, Keiko appears to have the ability to use filled pauses not only to deal with difficulties but also as a signal to the listener of complex grammatical constructions, while Asami seems to use them only in problem spots. However, both speakers seem to be aware of the need to avoid long silences, and employ filled pauses to facilitate that.

In terms of other speech management devices, at points of restarts, repetition, and self-correction, Keiko tends to begin the corrected phrase at the beginning of the clause or phrase, as in the previous example (line 069). She also uses filled pauses in most of these locations. Restarting at the beginning of the clause or phrase is typical of native-speaker restarts (Biber et al., 1999), although non-native speakers tend to overuse filled pauses at these points (Götz, 2013).

Asami has restarts that are very similar, as in the following example from her second dialogue, where she restarts from the beginning of the clause, and uses a filled pause to signal difficulty.

Extract 7.13 Asami Dialogue 2

```
076   ASAMI `so <it `seems
               * <- looks away to left
                        * <- left hand palm down downward to left
              ============+
077          `(530)
078          uh` to-> [//]
079          (185)
080          `it seems `hard to `carry,
              * <- resumes eye contact
                  * <- 2 hands palms down downward beat
                                 * <- gestures bicycle handles
                        5============
```

Although traditionally considered to be a sign of dysfluency, restarts are common in native speaker speech, and Götz (2007) has suggested that learners' lack of repetition counterintuitively hurts their perceived fluency. There are some points in Asami's retellings where difficulties are not mitigated by any speech management devices, as in the following example from her second dialogue. Here one clause is broken into three utterances with intervening silent beats internal to the phrase. This is a case where a hesitation with a filled pause, and a repetition of the entire phrase could help the speaker to appear more fluent.

Extract 7.14 Asami Dialogue 2

```
097   ASAMI but a `fat man [90 bpm]
                     * <- gestures 2 hands down to right side
              * <- looks up to left
             ====+
098          `(443)
099          `couldn't
100          `(622)
               * <- resumes eye contact
              2==
101          `notice.
             =======
```

Götz (2013) found in her corpus-based study of fluency that there were three broad groups of native speakers, and three broad groups of non-native speakers, based on how they achieved more fluent (or not fluent) speech. The largest native speaker group achieved high temporal fluency with frequent formulaic language, and relatively low numbers of unfilled pauses, filled pauses, repeats, and discourse markers. Another native speaker group achieved high speech rate with the use of repetition, discourse markers, smallwords, and many pauses (i.e., shorter utterances), but relied less on formulaic language. Finally, the lowest temporally fluent group used few of any of these features.

The non-native speaker groups seemed to fall into two main groups, either achieving their fluency through frequent use of formulaic language, or through frequent use of repetition, filled pauses, and discourse markers, but not both. Even the third, smaller, highest fluency group relied mostly on formulaic language, and less on fluency enhancement strategies.

Keiko appears to be a learner who relies heavily on formulaic language, which appears to be what enables her more fluent moments throughout her narratives. To reiterate, in this study, formulaic language is being defined as language that appears to be stored whole for the individual learner, rather than formulaic as defined by frequency in a native speaker corpus. This seems to enable Keiko to avoid disruptive pauses internal to phrases more than Asami. Asami, at a lower level of proficiency, does not really fit into any of Götz's groups, but appears to be more similar to the group that used fluency enhancement strategies more than formulaic language.

7.5 Listener responses

If dialogic fluency, or confluence, is the smooth flow that results from the co-construction of the conversation by all the participants, then turn-exchanges are key points in the conversation. Particularly in a narrative-style conversation such as these Pear Story retellings, where the onus is on the speaker to continue the narrative, listener responses serve to give feedback to the listener that the narrative is understood, and thereby help the narrative to be smoother and more efficient, as well as create alignment and affiliation between the listener and speaker. In order for the listener responses to properly mesh with the narrative to create a harmonious duet rather than clash, the listener must rely on cues from the gaze, gestures, prosody, and syntax of the speaker, as well as pragmatic features of the discourse. This section will examine the listener responses in the Pear Story dialogue retellings, and how they interact with these other speaker features in the conversation.

Listener responses can be verbal or non-verbal. Verbal listener responses can be divided into four categories, based on the framework developed by O'Keeffe and Adolphs (2008). The first category is *continuers*, which are minimal responses such as *mmhm*, that signal the speaker to continue speaking. The second is *information receipt tokens* such as *okay*, which signal that the speaker has been understood. The third are engagement response tokens such as *wow*, which convey an emotional response such as surprise. Finally, there are *convergence tokens* such as *yeah*, that are used to show agreement and convergence. All of these, except perhaps *continuers*, imply that the speaker has been understood, and engagement and convergence tokens would seem to be especially important for creating affiliation and general alignment between the listener and speaker.

Non-verbal listener responses include head nods and facial expressions or other gestures. In the present study, head nods and facial expressions or other apparently meaningful movements by the listener were annotated in line with the transcript for the complication segments analyzed in Chapter 6. There were five categories of head nods based on the framework of Knight and Adolphs (2008). Small head nods can occur as a single nod (Type 1) or as multiple nods (Type 2). More intense head nods can also occur as a single nod (Type 3) or as multiple nods (Type 4). Type 5 nods are a combination of small and intense head nods.

Although there were important differences between the four native speakers in how they responded to the narratives, some general trends did appear. Verbal listener responses tend to occur after prosodic boundary tones and at points of syntactic completion. In the following example, the listener response follows a boundary tone (lines 123-125), but that pause is not at a strong point of syntactic completion. It has been shown that listeners use both syntactic and prosodic information to predict turn endings (De Ruiter et al., 2006), so this example may not be unusual. Furthermore, it is possible that the use of *i think*, as well as *just*, may be showing doubt to the listener, who may then be responding to help or assure the speaker to further the narrative.

Extract 7.15 Keiko Dialogue 2

```
117   KEIKO `an(d)
               * <- Keiko looks to her right as if thinking
               * <- hands together
118         `(570)
119         i `think that `he:
120         `(440)
121         jus(t)
122         (295)
123         `try to `steal, [90 bpm]
                   *   <- Keiko resumes eye contact
124         (315)
           1====
125   NS4   ┌mm`hm.┐
126   KEIKO └a     ┘  `basket `of
                          * <- hands open palms up, beat down
             ====+
127         `(441)
128         the pears.
             * <- hands roll and palms up on "pears"
```

Listener responses are also closely tied to eye contact. In this study eye contact was only inferred from the video and not precisely measured, so further research is needed to confirm these observations, however, it was not difficult to determine from the video when the speaker was looking away. When the speaker's gaze is directed away from the listener, the listener does not respond, even if this results in missed beats in the established rhythm. At the same time, however, for both speakers, when they were not looking at the listener, they did not use prosodic boundary tones when they paused, and tended to make eye contact before completing a clause. Therefore, it is difficult to determine a cause and effect relationship between a particular action and the presence or absence of a listener response, but there was a strong tendency in all four dialogues, for verbal listener responses to occur at points with prosodic boundary tones, syntactic completion, and the speaker's gaze toward the listener.

In the four complication segments of the Pear Story retellings, there were a total of 38 listener responses. Continuers comprised the largest group (47%), but engagement tokens (24%) and information receipt tokens (26%) were also common. There was only one convergence token in this subset of data. Knight and Adolphs (2008) found in a graduate supervisory session that most listener responses were continuers and convergence tokens. Therefore, the type of discourse may have an influence on the types of listener responses. It may be that the lower number of convergence tokens is due to the fact that the story being told was not the subject's own story, but a retelling of a movie. The nature of the task may have prioritized the understanding of the story by the listener, which is shown more through engagement and information receipt tokens.

As discussed above, verbal responses tended to occur at points of syntactic completion and prosodic boundary tones. In these segments, the non-minimal responses (engagement and information receipt tokens) tended to follow the completion of C-units, while continuers were used at breaks within the C-unit.

Non-verbal responses (head nods) were also very common, and in fact, more frequent than verbal responses with 44 head nods occurring in the four segments. Smaller nods, either single nods or multiple nods, were more frequent than more intense nods (Table 7.1).

Table 7.1 Overlap of head nods with listener responses for all dialogue retellings

	Head nods	Continuers	Engagement	Info receipt	Convergence	overlap
Verbal LR		18	9	11	1	79% of 38
Type 1	15	8				50%
Type 2	17	8	1	4	1	71%
Type 3	8		1	3		50%
Type 4	3	1	1	1		100%
Type 5	2	1				59%
overlap	62% of 45	100%	33%	73%	100%	

Head nods often accompanied listener responses, and when they did, they often started earlier than the verbal responses. All of the 18 continuers occurred with or during head nods, usually of type 1 or 2 (small nods). Other listener responses often occurred with head nods, giving an overall rate of 79%, but this number does not reflect the individual differences in the native speaker listeners that result in quite different profiles.

The listener in Asami's Dialogue 1 (NS 1), who was an American male in his 30's, responded much more often with head nods only rather than verbal responses. Eight head nods occurred without verbal responses, and all nine verbal responses were accompanied by head nods (Table 7.2). Most of the head nods were small (type 1 and 2), and most of the verbal responses were continuers. The more intense (type 3) head nods seemed to be performed in response to difficulties—restarts in particular—by Asami. In line 163 of the example below, a type 3 head nod occurs after *boy* when this was the result of three restarts, and therefore is likely meant to assist the speaker by showing comprehension after the point of difficulty.

Extract 7.16 Asami Dialogue 1

```
159   ASAMI `but <`she> [//]
                * <- looks away
                         * <- hands move to left side
                                prep for taking, remain there
160         (195)
161         uh  <he-> [//]
                    * <- resumes eye contact quickly
162         (290)
```

```
163        <he-> [//] the `boy `robbed [95 bpm]
                                   3======+ <- NS quick head raise,
                                       preceded with lean forward
                                     * <- gesture taking something
```

In the following example, the head nod in line 142 is exceptional in that it overlaps Japanese speech and occurs when she is not making eye contact. At this point in the retelling, NS1 suppressed a yawn as well, so this could be a sign of disengagement, or of trying to pretend to be engaged.

Extract 7.17 Asami Dialogue 1

```
140        <the bicycle is so big-> [//]
141        (300)
142        big tte@s:jpn iu@s:jpn ka@s:jpn ano@s:jpn
               1=========================+
143        (340)
144        the `bicy`cle `is `as `big `æ`zʊ:@u [:as] [180 bpm]
                * <- vague gesture, 2 hands in front,
                     palms inward
                * <- NS suppressed yawn
```

Table 7.2 Overlap of head nods and listener responses in Asami Dialogue 1

	Head nods	Continuers	Engagement	Info receipt	Convergence	overlap
Verbal LR		6	0	2	1	100% of 9
Type 1	5	2				40%
Type 2	9	3		2	1	56%
Type 3	2					0%
Type 4	0					
Type 5	1	1				
overlap	47% of 17	100%		100%	100%	

The listener in Asami's Dialogue 2 (NS2) tends to use more substantial responses, and has nine listener responses in this segment—five engagement responses, three information receipt responses, and only one continuer (Table 7.3). This participant also tends towards multiple and more intense head nods. All of the listener responses are accompanied with head nods except for the engagement tokens, *oh no* and *huh*. The type 3 head nods (one of which does not emcompass a verbal response) seem to act as an information receipt responses. The silent type 3 nod occurred in the middle of an utterance, at a point with no syntactic or prosodic completion, but after the word, *robbed* (line 094), which is the key word in this complication segment.

Table 7.3 Overlap of head nods and listener responses in Asami Dialogue 2

	Head nods	Continuers	Engagement	Information	Convergence	overlap
Verbal LR		1	5	3	0	78% of 9
Type 1	0					
Type 2	2		1	2		100%
Type 3	3		1	1		67%
Type 4	2	1	1			100%
Type 5	1					0%
overlap	75% of 8	100%	60%	100%		75%

In Keiko's Dialogue 2, the listener tended to use very small head nods (Table 7.4), and only uses them with continuers. Out of eight responses, one was an engagement token and two were information receipts, none of which were accompanied by head nods. While generally following the overall trends discussed above, there was one exceptional case. In the following example (line 131), the type 1 head nod occurs in the middle of the utterance, but after Keiko uses a low-rise prosodic boundary tone, which usually signals a pause, showing that small misalignments may occur as the two participants work to together to create the conversation.

Extract 7.18 Keiko Dialogue 1

```
131   KEIKO but the `man, who was` picking up` pears, [90 bpm]
                    * <- right hand points to right
                                   * <- 2 hands rolling
                                        around each other
                                        then palms up
                                     holding pears on "pears"
                             1=====+           2=====
```

Table 7.4 Overlap of head nods and listener responses in Keiko Dialogue 2

	Head nods	Continuers	Engagement	Information	Convergence	overlap
Verbal LR		5	1	2	0	62% of 8
Type 1	3	1				33%
Type 2	4	4				100%
Type 3	0					
Type 4	0					
Type 5	0					
overlap	71% of 7	100%	0%	0%		

In Keiko's Dialogue 2, the listener tended to respond a lot. She had more verbal responses than the other native speaker participants, approximately half of them being continuers, as well as the second highest number of head nods (Table 7.5). She did not use head nods with the engagement responses indicating surprise.

Table 7.5 Overlap of head nods and listener responses in Keiko Dialogue 2

	Head nods	Continuers	Engagement	Information	Convergence	overlap
Verbal LR		6	3	4	0	70% of 13
Type 1	7	5				67%
Type 2	2	1				50%
Type 3	3			2		67%
Type 4	1			1		100%
Type 5	0					
overlap	70% of 13	100%	0%	75%		

In summary, it appears that different listeners have different profiles with their use of verbal and non-verbal responses, although there were many common features. Among the common features were that listener responses tend to occur when the speakers are making eye contact, and that verbal responses tend to occur at points of syntactic, pragmatic, and prosodic completion, and tend to be aligned with the speaker's rhythm. Continuers tend to co-occur with head nods, although the head nods may start earlier than the verbal response. More substantial verbal responses tend to occur at the boundaries between C-units or more substantial segments of the narrative. Head nods also occur in the same locations as verbal responses, but may occasionally occur in middle of utterances, particularly at points where the speaker is having difficulty. In general, listener responses serve to combine with the speaker's narrative to make a confluent duet of discourse, but they may also serve to assist or reassure the speaker at times of difficulty. This generally results in turn boundaries that are smooth, with the listener response occurring during the speaker's pause, but they are not always perfect. Listeners sometimes may misinterpret the ending of a turn when syntactic and pragmatic completion seems imminent, but the speaker continues regardless, or when the continuing phrase is not necessary to give the clause syntactic or pragmatic completion.

From these data, it does not appear that aligned verbal responses from the listener are required for the maintainance of rhythm. In Asami's Dialogue 1, she was able to maintain rhythm without any verbal contribution from the listener. There were head nods, and while these fall at appropriate positions in Asami's narrative, they do not always appear to be aligned with the rhythm clearly, since they often start just before the end of utterances.

Even with a listener present, the tempo may be varied at times, perhaps reflecting difficulty on the part of the speaker, or other factors such as hurrying after a less fluent segment of narrative. With no listener present, there is no need for alignment of any kind, therefore, the speaker is free to start, stop, and change the rhythm at will. If we assume that alignment—rhythmic and otherwise—is one way to create a positive social connection between the participants, then alignment would be something that speakers strive for (in ideal situations). When rhythmic alignment occurs, the conversation may be felt to be smooth and easy from the point of view of the participants. Therefore, rather than being a necessary condition of a conversation, rhythmic alignment may be more of a sign of afflliliation, as suggested by Müller (1996), i.e., one of the many ways that speakers align both socially and linguistically. Furthermore, the result of rhythmically aligned speech is a conversation that flows with minimal gaps and overlap, avoiding the longer pauses that appear in the monologues. Therefore, this may be a major factor in the perception of Asami as a competent speaker from the point of view of interlocutors.

7.6 Turn boundaries

Turn boundaries are important points in conversation, and a key feature of a confluent conversation is the smooth negotiation of these points. This section will bring together the features in dialogue that facilitate smooth turn boundaries, as well as look at points were turn boundaries were not so smooth, and examine their causes.

The following example, from Asami's Dialogue 2, is an example of a smooth turn boundary for the lower-proficiency learner. There is no response from the listener after line 062, since Asami is not making eye contact. After returning her gaze toward the listener as she says *a small boy*, the listener begins nodding and Asami completes the utterance with a low-rise boundary tone. After one silent beat, she adds *showed up*, with another low-rise boundary tone. This point is syntactically complete and has a low-rise boundary tone (signaling continuation), and the listener responds on the following beat with a continuer. However, Asami adds *with bicycle* with a falling boundary tone, signaling the end of the C-unit. Her gesture imitating bicycle handles is timed with the stress on *bicycle*. The listener then responds with an information receipt token on the next beat. Listener responses other than continuers tend to occur at the end of C-units as in this example, but this response could also reflect the listener's updating the story as the extra information in line 068 is added. The listener's head nods continue through all of this until Asami looks away in line 072.

In this example, all of the speech from lines 064 to 070 is rhythmically aligned at a tempo of 112 bpm, with only two silent beats. Asami shows the ability here to establish a rhythmic gestalt, which the listener is able to align to, in spite of Asami's tendency to stress almost every word. She is also able to use boundary tones to signal continuation or completion appropriately.

Extract 7.19 Asami Dialogue 2

```
062   ASAMI and then,
              * <- looks away
063         (271)
064         a `small `boy, [116 bpm]
                 * <- resumes eye contact
                      4===
065         `(315)
             =====
066         showed `up,
            ==========
067   NS2    mm`hm,
            =====+
068   ASAMI with `bicycle.
                    * <- gestures bicycle handles
069   NS2    (203)
            2====
070         o`kay,
            =====
071          (103)
            =====
072   ASAMI the `bicycle `is as `big `as the `boy, [94 bpm]
               * <- looks away     * <- resumes eye contact
                                   * <- 2 hands beat down together
                      2 hands wave to left -> *
                ==============+               4==
073   NS2    (310)
             =====
074         `mm!
             ===
```

In line 072, Asami slows the tempo down as she says *the bicycle is as big as the boy*, which is a phrase that appears to be difficult for her to produce here and in the other retellings. This change in tempo does not adversely affect the listener, however, who is able to respond with an engagement token on the following beat. As in the previous example, the listener responds with multiple intense head nods after Asami resumes eye contact in line 072 and continues these nods over several lines.

Overlaps would traditionally be considered as signs of problems. It has already been demonstrated above that although gaps are generally avoided in dialogue, what fills them tends to occur according to the rhythm established at that point in the conversation, rather than trying to reduce the intervening silences to zero. Therefore, gaps should normally be present, and their length is dependent on the rhythm and the speech sounds occurring after the previous stressed syllable and before the following stressed syllable. On the other hand, overlaps, ideally, should not occur at all in a “perfect” conversation.

In the complication segments of Asami’s dialogues, there were several cases of overlap. In the following extract from Dialogue 2, shown below, Asami and NS2 overlap in lines 089-090. In this example, Asami sets up a 90 bpm rhythm in line 087, and ends this line with a low-rise boundary tone. She then continues in line 089 without missing a beat. This boundary tone, as well as the syntactic completion of the utterance in line 087 was likely interpreted by the listener as a TRP, such that she reacts with the engagement token, *oh,* on the next beat. Asami’s rushed continuation results in overlap, but she holds the floor and continues uninterrupted into the next utterance.

Asami’s non-nativelike features here are the use of *because* as a free-standing discourse marker with a boundary tone. Also, her use of a boundary tone without a pause, leaves no space for the listener. However, in this example, alignment is maintained, and the story continues smoothly in spite of the overlap. The overlap occurs, but the flow of the conversation is maintained by both participants.

Extract 7.20 Asami Dialogue 2

```
087   ASAMI but `actually the `boy is `bad boy, [90 bpm]
                                 * <- left hand point toward table
                                          * <- repeat same gesture
088         (271)
           3====
089        be┌`cause,┐
090   NS2      └`oh!   ┘
           ======+
091         (260)
092   ASAMI `he: [//]
               * <- looks away
```

In the following example from Dialogue 1, shown below, a rhythmic pattern with a tempo of 110 bpm has been established by the utterances in lines 176 and

178, with one intervening silent beat. After the low-rise boundary tone and another single silent beat, NS1 responds with *mmhm* in line 180, just as Asami resumes speaking—both approximately on the next beat. Therefore, this example shows overlap resulting from alignment of rhythm, which might have been avoided if Asami had begun the phrase earlier with anacrusis, such that *ride* occurred on that beat.

Extract 7.21 Asami Dialogue 1

```
176   ASAMI `pears `box `on [110 bpm]
              * <- gestures bicycle handle and hold through
177         `(685)
178         `bicycle,
             * <- gesture bicycle handle downbeat
179         `(760)
            5==== <- intense then multiple small nods
180   NS1   ┌mmhm,┐
181   ASAMI └`and ┘ `he `ride `on `the `bicy`cle, [150 bpm]
                          * <- RH short wave to right
                                * <- return to bicycle handles
                                         * <- gestures steering
                                              bicycle handles
            =====+
```

Keiko’s cases of overlap are similar in some ways to Asami. There are examples where she comes in on a beat following an utterance that has been syntactically and prosodically completed. In the following example, Keiko has established a strong rhythmic pattern with the long utterance in line 148. It would not be surprising to pause after the relative clause in *the man picking up pears*, and at this point the listener inserts a continuer on the following beat. However, Keiko doesn’t pause at this point and finishes the C-unit. In this example, it is impossible to put “blame” on the speaker or the listener for the overlap, since both have justification to speak at that moment. NS4 responds on the next beat after the completion with *oh okay*. Keiko’s *yeah* overlaps with the stressed syllable in *okay*. It may be that Keiko believed that NS4 was only saying *oh*, and therefore began her response to the listener response, as she does in other cases.

Extract 7.22 Keiko Dialogue 2

```
148   KEIKO `yeah and the `man uh `picking up `pears ┌he `did┐n't
`realize it.
                             * <- right hand points out to right
                          right hand palm up -> *
                    right hand rapid shake on "didn't" -> *
149   NS4                                          └mm`hm, ┘
                                                   1====+
150         (140)
151         `oh o┌`kay. ┐
             * <- NS leans back
152   KEIKO      └yeah. ┘
```

Probably the most striking case of overlap occurs later in Keiko's Dialogue 2, shown in the extract below. Keiko is either struggling somewhat to make her point or trying to convey uncertainty, as evidenced by the 445 msec silence and lengthened filled pause in lines 157-158, and the fact that she is looking away from the listener. She completes the phrase with a falling boundary tone in line 160 as she moves her gaze back to the listener. This is followed by a 655 msec silence over one beat of the rhythm, where the listener leans back and opens her mouth as if saying "ah". Perhaps Keiko misunderstands the meaning of this reaction, because she appends a phrase to her previous (completed) C-unit, to explain what she means in more detail. Keiko starts this utterance on the same beat as the listener's verbal continuer (lines 162-163). It appears then that the listener perceives that Keiko thinks she has misunderstood, and rushes to insert a repeated information receipt response. Overlapping with this is Keiko's rushed explanation, which is given up after she hears the information receipt response. Keiko then sums up this section at a slower tempo in lines 165-167. Although this is a case of clear overlap, and perhaps mutual rushing to clear up any misunderstanding, it can also be seen as an example of interactional achievement, where Keiko and NS4 collaborate to quickly clear up the problem, so that Keiko can immediately continue the story.

Extract 7.23 Keiko Dialogue 2

```
154   KEIKO an(d)
             * <- K looks left
155         (235)
156         i thought at firs(t)
157         (445)
158         u:h `he [120 bpm]
159         (295)
160         `tried to `help him.
                       * <- K resumes eye contact
161         `(655)
             3====+ <- opens mouth as if “ah”, pulls head back
162         ┌to `car┐ry the ┌basket but i think yeah.┐ <- rushed
163   NS4   └uh`huh.┘       └oh i see i see.        ┘ <- rushed
              2====+        * <- NS looks upwards briefly
164         (320)
165   KEIKO `just he `try to [90 bpm]
166         (320)
167         `steal it.
```

In summary, both Asami and Keiko are able to participate in smooth turn exchanges during their narratives, generally avoiding overlap that gets in the way of the narrative. This is achieved by the establishment of rhythm in speech such that listener responses can occur in the intervals between the speaker’s turns, and the appropriate use of boundary tones and control of sentence structure to signal the completion of a turn and whether or not there is intention to continue the previous C-unit. There are times when there are weaknesses with the control of prosodic and syntactic structures, however, they are able to coordinate with the listener to maintain the flow of the conversation.

7.7 What causes high AR and short pauses in dialogues?

Asami tends to speak at a faster pace (tempo) than Keiko, however, Asami speaks with a significantly lower articulation rate and length of utterances. This higher tempo is achieved by Asami stressing more words in each phrase than Keiko, i.e., speaking in more of a “syllable-timed” manner while Keiko speaks in a more “stressed-timed” manner. On the other hand, in Chapter 5, it was found that the mean length of pauses did not vary between participants. Within a given rhythmic pattern, pauses can be either (1) short enough to fit between stressed words on consecutive beats, (2) or long enough to accomodate one silent beat. Longer pauses may result in a minor disruption with a delay and restart of the rhythmic pattern, or a major disruption that loses the rhythm. They may also simply result in two missed beats without a loss of the rhythmic pattern, however these can sometimes be filled by listener responses in the dialogues. Even when the speech has a steady rhythmic pattern, the length of individual pauses can vary greatly based on the unaccented syllables following the previous beat and preceding the following beat. The lack of differentiation in the mean length of pauses between Keiko and Asami could be a factor of Keiko having longer periods between beats (due to the lower tempo), but partially filling those spaces with unstressed syllables, as in the following example where a short 205 msec pause contains two silent beats. The two silent beats occur during the combined time of the lengthened vowel on *the* in line 094, the pause in line 095, and *on the* preceding the stressed word, *bicycle*. Asami, with higher tempo, has shorter periods between beats, but has fewer unstressed syllables to fill them due to her manner of placing prominences.

Extract 7.24 Keiko Dialogue 1

```
094   KEIKO is `coming <on the:> [//]
095        ``(205)
096        on a `bicycle.
```

On the other hand, in Chapter 5 it was found that for both subjects, the dialogues had higher articulation rates and shorter average pause lengths. Mathematically, a higher pace would imply a higher articulation rate and shorter pause lengths, but the tempos did not vary much by genre. The higher articulation rates and shorter average pause lengths observed in the dialogues seem to be related to the more consistent rhythm present in the dialogues. In the monologues, the portions that do not show a rhythmic pattern tend to have lower articulation rates, with many single-word utterances and lengthened vowels. The majority of Keiko's dialogues are rhythmic and show alignment between the listener and the speaker. During these periods, the listener's responses result in many stretches in the narrative with no silent beats, and other stetches with single silent beats. Multiple silent beats are rare. The listener responses do not appear to have a direct effect on the length of the pauses in the speaker's narrative, but they seem to help to maintain the established rhythm through potentially less fluent segments of the speaker's narrative.

For Asami, the mean articulation rates were also higher and mean pause lengths shorter in the dialogues than the monologues. As noted above, Asami tends to speak in more of a syllable-timed manner, which results in generally higher tempos when rhythm is established compared to Keiko. However, Asami's most fluent narrative, Dialogue 2, which features longer utterances than the other retellings, also has lower tempos. The second half of Dialogue 2 is a long stretch of continuous rhythm at 90 bpm, with strong alignment between the listener and speaker. In this segment, Asami shows a tendency to stress fewer words as in *but actually the boy is bad boy* (line 087), with an articulation rate of 4.44 syllables per second. Compare this to an earlier utterance in this same retelling, at a point in the story that was difficult for Asami in the previous three retellings, where she stresses almost every word, *the bicycle is as big as the boy* (line 072) at a articulation rate of 3.20 syllables per second. Therefore, it seems that with both Asami and Keiko, stressing fewer words in the utterances tends to be associated with higher utterance fluency, i.e., a higher articulation rate due to more rapid unstressed syllables, while at the same time resulting in a slower tempo. This tempo, however, is based on being able to produce longer utterances that can establish and maintain a rhythmic gestalt. Keiko, as a higher-proficiency speaker, is able to produce longer utterances than Asami, and is more consistent across retellings.

Therefore, to summarize the temporal effects of pace on other temporal variables of fluency, all other factors (such as length of utterance) being equal, a higher tempo should result in a higher articulation rate and shorter pauses, and thus a higher speech rate. However, this also depends highly on the manner in which rhythm is created, i.e., the frequency of stressed words in speech. For example, a four-syllable utterance spoken at a given articulation rate will have a lower tempo if only two syllables of the four are stressed than if all four syllables are stressed. The learners in this study were able to shift between speech where more words, including function words, are stressed, and speech where fewer words are stressed. This appears to depend on the proficiency of the speaker, such that Keiko tends to stress fewer words and Asami tends to stress more words. However, it varied within each learner's speech, and even within a short segment of conversation.

There is also likely a social factor responsible for the higher articulation rates in the dialogues. In all four of Keiko's retellings, this segment begins with the same four words: *and then, a boy.* In Dialogue 2, these utterances (lines 100-102) take a total of 1443 msec to utter, while in Monologue 2, lines 087-089 take a total of 2057 msec—a much slower rate of speech. Keiko's Dialogue 1 and Monologue 1 are slightly faster, both cases taking 1400 msec to utter these four words separated by one pause. With the same words uttered in all four retellings, and the fastest being the first retellings, there is no evidence for practice being a factor in the speed of these utterances. It appears that the speech rate is a choice she makes, rather than merely the result of the limits of her cognitive fluency. Comparing the utterances in Keiko's two monologues, Monologue 2, which begins this segment more slowly, has fewer restarts and self-corrections, and uses more subordinate clauses. Therefore, it may be that Keiko's slower start reflects a trade of lower speed for more accuracy and syntactic complexity.

7.8 The effect of listener feedback

Previous research has shown that the listener can have a powerful effect on the fluency of the speaker (Bavelas et al., 2000). In Asami's Dialogue 1, the listener differed from the other dialogues in that he rarely used verbal responses. He also showed some signs of disengagement, such as yawning. It is impossible to generalize with the limited data in this study, however, it could be that the listener's lack of verbal responses in Dialogue 1 puts much more of the onus entirely on the speaker to create and establish rhythmic patterns. Further research is needed to determine if this kind of listener would have effects on higher-proficiency speakers. In Dialogue 2 with a more verbally responsive interlocutor, the feedback Asami gets from the listener allows her to avoid some of the difficult phrases that she dealt with in the previous retellings. In particular, she avoided explaining why the fat man did not notice the boy stealing the pears, which she had difficulty with in both Dialogue 1 and Monologue 2. The following extract shows this example in Dialogue 1. Here Asami has difficulty explaining this portion, as evidenced by the predominance of single-word utterances in lines 083 to 105. She also stresses many more syllables, including vowels added to the end of words, in order to maintain the rhythmic pattern. She gets feedback from the listener after the first phrase (lines 083 to 089) in the form of head nods (lines 090 to 091). In the next phrase, she has much more difficulty, using some Japanese in line 101. Even with the gestures and eye contact, the listener does not seem to understand until line 105, when he finally nods as an information receipt response.

Extract 7.25 Asami Dialogue 1

```
083        `duɾiŋ`gu@u [: during] [130 bpm]
084        `(395)
085        the `fat `man;
                 * <- right hand up in prep for next gesture
086        (465)
087        `gæða`dʊ@u [: gathered];
            * <- resumes eye contact
            * <- right hand gestures picking pears on 2 beats
088        (265)
089        `pea`zu@u [: pears],
090        `(705)
           2====
091        `he
           ==+
092        `(620)
093        `kʊdɪn`tu@u [: couldn't];
            * <- looks away to left
094        (300)
095        <`look-> [//]
             * <- gesture pointing down and behind her
             * <- regains eye contact
096        (270)
097        `watch
            * <- same a previous gestures repeated twice
098        (520)
099        `uh
           * <- gesture with both hands pointing down and behind
           * <- looks briefly down behind her right side
100        (385)
101        °nan@s:jpn darou@s:jpn.°
102        (730)
103        back of uh
           * <- NS leans forward and turns head
                moving left ear closer
104        (450)
105        &hɪ- [//] him?
                     * <- quick single head nod
           3==+ <- NS as if to show comprehension or "ah!"
106        (265)
107   NS1   mmhm,
          2===+
108        (220)
109   ASAMI mm mm mm so,
           * <- A nods type 2
                    * <- looks away to left
                2===+
110        (720)
111        <the-> [//]
112        (260)
113        `he: [120 bpm]
114        `(670)
115        `kʊdɪn`tu@u [: couldn't];
116        (240)
117        `noticed` the `boy,
            * <- resumes eye contact
                          * <- Asami head nod
118        (260)
119   NS1   mm`hm
           1===+
```

In contrast, in Dialogue 2, this same point is much more short and concise. She avoids the most difficult part where she explains the reason, and gets an immediate information receipt response from the listener, which allows her to continue the story.

Extract 7.26 Asami Dialogue 2

```
097   ASAMI but a `fat man [90 bpm]
                    * <- gestures 2 hands down to right side
             * <- looks up to left
            ====+
098         `(443)
099         `couldn't
100         `(622)
              * <- resumes eye contact
             2==
101         `notice.
            =======
102   NS2   (99)
            ====
103        m┌`m:.  ┐((.pt))
104   ASAMI  └`mm mm┘ mm mm ┌ mm.┐ <- A rapid nods type 2
105   NS2    *              └`i  ┘ see. <- NS wince;
             ==========================
```

7.9 Conclusions

This section will bring together the results of this and the previous two chapters to answer the research questions posed in Chapter 3. To review, the main research questions for the present study were the following.

1. To what extent should co-constructed features of fluency be taken into account in definitions and frameworks of fluency?
2. What factors interact in a complex system that account for the dialogic (co-constructed) fluency of a conversation?

   Furthermore, three secondary questions were also posed.

A. Are co-constructed aspects of dialogic fluency able to create confluence and therefore, an impression of fluency, in spite of weaknesses in monologic fluency?

B. In particular, can alignment of rhythm between speakers lead to an overall higher impression of fluency in a conversation in spite of lower fluency in terms of temporal variables?

C. What are the most effective research methods for a study on fluency?

7.9.1 What factors interact in a complex system that account for the co-constructed fluency of a conversation?

Even a one-sided dialogue such as these Pear Story retellings has been shown to involve a complex collection of factors that participants employ to create a smooth conversation. One important point where these factors come together is at turn boundaries. In a narrative-type conversation such as these, turn boundaries result in smoothly continuing the narrative while maintaining the floor with the storyteller. This is achieved through prosody and syntax, with low-rise boundary tones generally used at points where the phrase is syntactically complete but the clausal C-unit is not, and falling boundary tones used at points where the clausal C-unit is complete, as well as using a plateau boundary tone for pauses that are internal to phrases. The listener uses these boundary tones and the accompanying syntax to predict when pauses in the narrative will occur, so that the listener can respond in a way that signals comprehension to the speaker, so that the speaker can continue the narrative smoothly. These listener responses can be verbal and non-verbal. Verbal responses tend to occur in a way that aligns with the rhythm of the speaker, and follows changes in the tempo as they occur. The rhythmically aligned listener responses also appear to help the speaker to maintain the rhythm too, inserting filled pauses or discourse markers when needed, and thereby avoiding long silences. Non-verbal listener responses often occur with the verbal responses, but may also occur in places where a verbal response would not be appropriate, such as overlapping the speaker's utterances. Eye contact also plays a role such that in these narratives, the speaker looking away from the listener generally precluded listener responses, and the speaker returning the gaze to the listener often prompted head nods.

Although in this study I initially approached alignment of rhythm as a feature that I hypothesized was important in enabling a fluent conversation, it seems that a more accurate description of this phenomenon would be to say that alignment of rhythm is the perceivable result of the participants are working together to co-construct the conversation. To use a musical analogy, conversation is a kind of duet where the speakers are working together to construct something that is smoother, or more fluent, than the individual contributions of the speakers taken alone. Rhythm is one way that this alignment is revealed. Another is through engagement listener responses, and facial expressions. Another is through mutual eye contact, which the speakers align to each other in these case studies, but could be mutually aligned to something else (for example, materials on the table in a task of some sort).

Particularly for the storyteller in these case studies, cognitive fluency is also an important factor such that the learner needs to have some degree of language competence to be able to construct the utterances to build the narrative. Without a certain degree of cognitive fluency and language competence, the learner would not be able to contribute enough to establish a rhythmic conversation, and contribute enough to allow the listener to help.

Although eye contact, head nods, gestures, and facial expressions clearly contribute to the conversation, the fact that telephone conversation is possible shows that these non-verbal aspects are not necessary for conversation to happen. On the other hand, most people with second-language experience would probably agree that talking on the telephone is more difficult than a face-to-face conversation in a second language. Therefore, as above, it may be more appropriate to think of alignment and the co-construction of conversation as something that speakers strive to achieve when there is affiliation or rapport between speakers, and this may be what makes a conversation feel smooth or confluent. This is supported by evidence from the monologue retellings of the Pear Stories, where the subjects were not able or did not find it necessary to maintain rhythm throughout the retelling, or to avoid long pauses.

7.9.2 Can alignment of rhythm lead to an overall higher confluence in spite of weaker monologic fluency?

As discussed in the previous section, it appears that rhythmic alignment is not a requirement of conversation, but a preferred condition, and is one of many ways that speakers align to each other. As the rhythm can be lost and restarted, and the tempo can be varied as the conversation progresses—reflecting such things as difficulties and pressure to catch up after hesitations—the participants must be able to adjust to each other to co-construct the rhythm. It does not appear that rhythmic alignment itself is responsible for differences in temporal measures of fluency of the dialogues, such as articulation rate and length of pauses, although they are certainly related such that with other factors being equal, a higher articulation rate would result in a higher tempo and shorter pauses. Rather, it appears that in dialogue, the learners strive more to align with the listeners, resulting in fewer long pauses and fewer long hesitations (lengthened words or filled pauses), as well as more consistent rhythm. While rhythm gestalts were sometimes established in the monologues, without the need for alignment with another participant, these were not maintained as consistently as they were in the dialogues. Therefore, it is not just that the listener is aligning to the speaker, but that both speakers are aligning to each other, with the onus on the storyteller to establish the rhythm.

7.9.3 Are co-constructed aspects of fluency able to create an impression of fluency in spite of weaknesses in monologic fluency?

As hypothesized from observations in the pilot study, Asami, the lower-proficiency learner in this parallel case study, was indeed able to participate in a conversation, involving the retelling of a short movie, with strong signs of alignment between the speaker and listener. Although her utterances tended to be short, and her articulation rate and speech rate much lower than nativelike speeds, she was able to establish rhythmic patterning in her speech, which could be aligned to by a native-speaker listener. Maintaining the rhythmic alignment appeared to result in fewer long hesitations in the middle of her story. In this way, she was able to avoid the awkward silences that often occur when a typical lower-proficiency learner speaks.

Comparing the higher-proficiency learner, Keiko, with the lower-proficiency learner, Asami, the most obvious differences are in aspects of monologic fluency, or more specifically cognitive fluency. Keiko is much more consistent in the choice of words across the retellings, sometimes using exactly the same phrase in multiple retellings. This implies that much of her language may be formulaic in the sense of being prefabricated (as opposed to being frequently recurring in a native speaker corpus). This resulted in generally longer utterances than Asami, spoken at a higher rate of articulation. Asami, on the other hand, showed much less consistency in the choice of words, and appeared to be constructing each utterance anew in each retelling, resulting in lower temporal measures of fluency.

In terms of rhythm, Keiko generally stresses fewer words than Asami, i.e., with more unstressed words occuring between stressed words, so that her sentence stress tends toward being more nativelike. This results in Keiko generally having lower tempos (or pace) in spite of the higher articulation rates. Both speakers also were able to vary their sentence stress—stressing more words per phrase—when necessary, so that a certain rhythm could be maintained with a lower articulation rate. For Asami, this meant sometimes putting stress on two syllables of longer words, or adding vowels to short words and then stressing those vowels.

Furthermore, Keiko appears able to more effectively involve the listener in the construction of the narrative. She was able to appeal to the listener for help when she was uncertain of what she was saying, which seemed to result in feedback from the listener sooner, enabling the narrative to continue efficiently. She also used discourse markers more effectively to engage with the listener, for example, using them to follow up listeners' reactions. While Asami used discourse markers, they were more limited in variety, and sometimes appeared to be more self-directed.

While Keiko is more consistent across the retellings in terms of utterance fluency and language use, Asami varies in several ways. She appears to be improving in utterance length over the course of the retellings. Some of this is likely due to a practice effect, but the main effect appears to be that in her second session, she slowed her articulation rate in order to achieve longer runs. She may be focusing on accuracy or complexity, rather than speed. This shows that Asami's fluency is not necessarily a constant feature of her speech. She is able to adapt her speech within her limitations to strive toward a desired effect—in this case, slowing down to achieve longer utterances.

As mentioned above, she is also able to vary her placement of stress in phrases to maintain rhythm while varying articulation rate. Particularly in her Dialogue 1, she appends vowels to the end of many words, and then stresses those vowels, resulting in what would appear to most native speakers of English as very Japanese-like pronunciation. In the majority of her speech, she avoids this, so this would appear to be a choice rather than merely non-nativelike pronunciation. In this case, she made the choice to use Japanese-like pronunciation to maintain rhythm with a much lower speech rate.

Asami is also able to employ speech management devices to compensate for her weaknesses in cognitive fluency, which may be due to a lack of formulaic language. These include using filled pauses, repetition, and discourse markers to maintain the conversation and maintaining alignment with the speaker.

To summarize, Asami has a variety of skills to employ in conversation, with limitations that come with her level of proficiency, but she is also able to trade these skills against each other to participate effectively in conversations, and appears to prioritize the maintenance of rhythmic alignment. Again, if alignment between speakers is what makes a conversation feel smooth to the participants, then Asami appears to be able to bring what skills she has to achieve that alignment, which may be what allows her to be perceived as more fluent in conversation than would be expected for her level of proficiency.

7.9.4 To what extent should co-constructed aspects of fluency be taken into account in definitions and frameworks of fluency?

The dialogic aspects of fluency are critical to the overall fluency of the conversation, which may be a more important effect on the impression of the interlocutor than purely an impression of their fluency. Holistic definitions of fluency focus on “ease of speaking” for a speaker or “ease of comprehension” of a speaker, but there is also a kind of “easy to talk with”, which is facilitated by the interaction between the speakers, rather than simply the product of one speaker’s input. Although not addressed in the present study, many second language teachers can probably think of weaker students who have some degree of language knowledge, but who do not hold up their end of the conversation, leaving long silences and not dealing with difficulties in a way that allows the listener to help. These conversations are painful for both participants. Dialogic fluency, created by the alignment between speakers, may reflect this sense of “easiness” in a conversation, which is of course dependent on the learner’s cognitive fluency, but not wholly determined by cognitive fluency.

In fact, it may be difficult to describe the cognitive fluency of a speaker in isolation, although some recent language tests are targeting just that (Van Moere, 2012). Certainly learners can be forced to speak alone to a recording device or to an examiner that does not respond, but that only reflects some of the factors that contribute to the ability to participate in a smooth conversation. If we take conversation to be the central context or format for spoken language use, then definitions of fluency need to take into account those other aspects that lead to a smooth conversation. Cognitive fluency is a very important aspect of that, however, how speech unfolds in conversation is not just a matter of producing correct sentences.

Therefore, *cognitive fluency* remains one important part of fluency, reflecting the monologic linguistic knowledge and language production skills of the speaker, but it does not take into account dialogic factors. *Utterance fluency* reflects the measurable output of language production, but can vary due to a wide variety of factors even through a single conversation. *Perceived fluency*, although probably influenced by dialogic factors, does not really describe those factors. I suggest that the term *confluence* be reserved to describe the smooth flow of the conversation itself, and the term *dialogic fluency* be used to refer to the skills that speakers require to successfully co-construct that conversation.

7.9.5 What are the most effective methods to research fluency?

A confluent conversation is the result of the interaction and alignment of the participants, co-constructed with both verbal and non-verbal means, with many of the non-verbal aspects compensating for what would traditionally result in weaker measurements of temporal variables. Therefore, future research on spoken fluency needs to be based on multi-modal data, as has been proposed by others (Adolphs & Carter, 2013; Knight, 2011; Knight & Adolphs, 2008), including annotations for head nods, gestures, facial expressions, and gaze. This will be necessary for future research to investigate the interactions between non-verbal and verbal features, and for longitudinal studies to investigate how the interaction of verbal and non-verbal features changes over learners' development. Unfortunately, this is quite difficult for studies based on larger corpora due to the time required to transcribe these features, as well as the difficulty of categorizing them. For example, even with something as seemingly simple as a head nod, the difference between a nonchalant head nod and an intense head nod is often difficult to distinguish, and in the present study seemed to vary with the particular person, with some participants' nonchalant head nods being smaller and others being larger. Therefore, even with technology to detect these features automatically (for example, face tracking technology), individual differences in participants may need to be taken into account.

It is clear that transcriptions of conversations for fluency research need to be done with software that allows annotations to be time-aligned with the audio and video files, so that the audio and video can be referred to during analysis, and so that lengths of utterances and silences can be accurately measured. Ideally, if speakers voices are recorded on separate channels with high fidelity, silences can be automatically annotated with software such as ELAN, and categorized using the method of Kirsner et al. (2002), which also takes into account individual differences between speakers. However, this study has highlighted the importance of rhythm in conversation, and rhythm may be impossible to automatically annotate due to the fact that it is a perceptual phenomenon, distinct from *timing*, which is the actual placement of prominences in speech (Arvaniti, 2009). As a perceptual phenomenon, it requires a person to perceive it in order to annotate it, although perhaps in the future machine-learning techniques might make automatic annotation of these features possible as well. Rhythm should also affect the categorization of pauses. The present study has shown that, at least in learner speech, pauses may occur that do not result in missed beats in the rhythm, and therefore should not necessarily be considered hesitations.

In conclusion, there are many factors that need to be taken into account when looking at a learner's ability to carry on a fluent conversation. Specifically, it is important to look at not only the features of the learner's cognitive fluency (as measured by utterance fluency such as length of run, articulation rate, and a simplistic lack of hesitations), but also how the learner is able to cope with holding up her end of a conversation and contribute to creating alignment between the speakers which results in a confluent conversation. This includes the ability to speak rhythmically, using stressed words and varying these to accommodate the learner's own speech rate at that point in the conversation. It also includes the ability to use repeats, restarts, filled pauses, and discourse markers to maintain confluence while planning speech. Although a native speaker lacks many of the production difficulties that a learner has to deal with, there is no perfect fluency, and even native speakers have signs of planning in speech. The learners in this study were able to use these techniques to hold their own in the conversation while dealing with the same planning issues that any speaker would have, and the additional language issues that learners have.

8 Summary and Conclusion

In this final chapter, I will review the main findings of the present study, and what these results imply for the understanding of dialogic fluency and our understanding of fluency in general. In particular, I will discuss how complex systems theory and a sociocognitive approach to second language acquisition may be promising avenues for future research in the co-construction of fluency. I will also show how existing models of fluency, such as that of Segalowitz (2010) and Rühlemann (2007), compare with the results of the present study, and how they fit into a theory of dialogic fluency. In the following sections of the chapter, I will discuss implications for language teaching, and especially language assessment, and discuss directions for further research. Finally, I will finish by attempting to address the need for a more comprehensive definition of fluency.

8.1 Summary of main research findings

A host of previous research has shown that fluency is affected by many factors beyond mere temporal variables of speech, and that fluency may be a co-construction of the two participants in a conversation (see Chapter 3). For example, McCarthy (2005) has described some of the observable features of the co-constructed fluency of a conversation, which he refers to as *confluence*. The goal of the present study was to examine how participants in a conversation do this, through parallel case studies of two Japanese learners of English—one a higher-proficiency learner, and the other a lower-proficiency learner—as determined by Versant test scores and my experience as a former teacher of both.

The lower-proficiency learner was atypical in that she was able to create a strong impression of fluency with native speaker interlocutors in spite of her lower-proficiency. The design of this research methodology was unique in that (1) it was multimodal, including information on nonverbal features such as head nods and gaze, (2) it employed a grounded theory approach to the examination of the case study data, and (3) it used the same story-retelling task so that the same content could be compared between monologic and dialogic contexts, and between the two participants.

Both participants had higher articulation rates and shorter pause lengths—and therefore higher speech rates—in the dialogues than in the monologues. This is consistent with the results of previous research that found that speech rates were higher in dialogic than monologic tasks (Bell, 2003; Ejzenberg, 2000; Riggenbach, 1989), and that speech rates were higher in discussion tasks than in narrative tasks (Witton-Davies, 2012). These studies differ from the present study in that the content of the tasks was not the same in the monologues and dialogues.

There were significant differences between the subjects as well. Consistent with previous research (for example Towell et al., 1996), the higher-proficiency learner had overall higher speech rates than the lower-proficiency learner. Also, the mean length of uninterrupted run was longer for the higher-proficiency learner, but did not vary between monologue and dialogue for either participant, which implies that this measure is more dependent on the proficiency of the learner, and not co-constructed. Additionally, the higher proficiency learner showed much more consistency in the mean length of runs across the four retellings, often using nearly the same phrase to express the same information in the story—an indication that her longer runs were achieved by greater use of formulaic sequences, and therefore higher cognitive fluency. The lower-proficiency learner, on the other hand, did not often use the same phrases in subsequent retellings, but her mean length of run showed some improvement chronologically over the course of the four retellings. This implies that length of run may develop with practice, unless the speaker already has a solid foundation of pre-fabricated language to draw upon for the given task.

In the pilot study (see Chapter 3), involving a free conversation between the lower-proficiency learner mentioned above and a native English speaking teacher, it was found that the two participants often broke long pauses at the same time, and it appeared that this was due to rhythmic alignment between the speakers. It is known that native speakers align in speech rate in conversation (Street, 1984; Street, Street, & Van Kleek, 1983), but in this pilot study, the learner had a much lower speech rate than the native speaker. However, the speakers showed a tendency to align the pace of stressed syllables, even over multiple turn exchanges and silent pauses. The learner was able to do this with a lower speech rate by stressing a higher percentage of words in her utterances relative to the native speaker. In this way, turn exchanges tended to occur so that the rhythmic alignment was maintained, and silences were broken on the following beat after what was perceived as a problematic silence—consistent with previous research showing that more than one silent beat is marked (Schegloff et al., 2002). In the main case studies (Chapters 6 and 7), it was found that the majority of the speech in the dialogues could be associated with a rhythmic gestalt, established by the learner telling the story, although with occasional restarts of the pattern and changes in the tempo. The listeners' responses tended to be aligned with this rhythm, helping to support and maintain it during pauses in the narrative. On the other hand, in the monologues, rhythmic gestalts tended to be abandoned quickly after being established.

The observed differences between dialogue and monologue appear to be due to the overall striving toward alignment in the dialogues, while there is no need for alignment in the monologues. Striving for alignment in the dialogues results in the need to create and maintain rhythm in speech, which results in fewer long pauses where the rhythm is lost (as in the monologues). Learners have various techniques to achieve this such as relying on filled pauses and giving up on accuracy and/or complexity to maintain the required speech rate. Also, it was observed that both of the learners in this study were able to vary the stresses on words to maintain the rhythm, within their own capabilities, at the given point of the conversation. In the most extreme cases, the lower-proficiency learner adopted a very Japanese-like (*katakana*) pronunciation, with vowels appended to words ending in consonants, in order to enable multiple stressed syllables in a single word, which enabled her to speak at a very low speech rate while maintaining the rhythm at that point in the conversation. This particular learner is capable of and generally uses much more nativelike American pronunciation, implying that this is a technique to maintain rhythm rather than a pronunciation weakness. Therefore, the co-construction of fluency is *active* in that speakers must be able to adjust and adapt their speech to maintain alignment as the conversation unfolds.

It may therefore be that alignment is one of the main goals of conversation—perhaps more than the propositional and transactional goals of the words. Rhythm is one type of alignment, which Fiksdal (1990) found was required for conversation to feel fluent to the participants. In the present study, it was also observed that alignment occurred not only in terms of rhythm, but also with non-verbal features such as head nods and hand gestures. These actions tended to occur in the physical space between the speakers, which was created when there was eye contact between them, and appear to be aligned with the rhythm of the speech to some extent. In much previous research on fluency, correlations with raters' judgments are often used to determine whether factors are relevant or irrelevant to perceptions of fluency. It may be the case that many of these features contribute to the creation of alignment, which is not necessary for conversation, but is a desired state. Therefore, while head nods cannot contribute to alignment in a telephone conversation, that does not diminish their contribution in a face-to-face conversation.

In conclusion, it seems that fluency in conversation is co-constructed to a significant degree. Conversation is the default or most basic context for speech, and the ability to participate in conversations is what many students strive for. A person talking alone to a recording device cannot be the basis for determining fluency in conversation. It can be said that fluency is more the ability to participate in a confluent conversation, than the ability to produce high-speed language. Asami was able to contribute to the co-construction of a conversation even with a lower speech rate. Of course, this depends on the ability to produce language in real time, which is based on knowledge of language, and the efficiency and automaticity of the processes of retrieval and production. This is Segalowitz's *cognitive fluency*. In the present study, the higher-proficiency learner achieved articulation rates and longer runs of speech with formulaic language, showing higher levels of cognitive fluency than the lower-proficiency learner. However, to participate in a smooth conversation also requires the ability to manage natural difficulties with language production and the interaction with another person—to hold the other person's attention while you search for a word, to get help from the other person when you need it, and to use speech management devices to manage natural difficulties without leaving silences that feel awkward. Even native speakers have to deal with these difficulties—both content-related and language-related, although learners have additional difficulties that native speakers may not have. All of these skills contribute to the ability to co-construct alignment in a conversation.

8.2 Sociocognition & alignment

The tradition of research in fluency appears to be mainly from a cognitive approach, which is not surprising since this has been the dominant view in second language acquisition for several decades. The results of the present study suggest that approaches to second language acquisition that emphasize interaction, co-construction, and alignment are better suited to the study of second language spoken fluency in conversation. One of these approaches is the complexity theory or dynamic systems theory approach (Larsen-Freeman, 2011; Larsen-Freeman & Cameron, 2008a). In this approach, second language acquisition is viewed as a non-linear system composed of a complex network of factors that interact with each other, resulting in unpredictable outcomes that emerge from the system. The difficulty of quantifying the variables involved make this theory difficult to apply mathematically to dialogic fluency. However, the implication that fluency may be not a property of the learner per se, but an emergent property of a system composed of the learner, the interlocutor, and their surroundings, that is not static but is constantly developing from internal and external pressures, fits well with the results of the present study.

Another approach is the sociocognitive approach (Atkinson, 2011), which has some similarities to the dynamic systems approach. This approach can provide a theory and methodology for the study of alignment, in particular. The theoretical basis of the sociocognitive approach is that cognition is just one part of a system comprised of mind, body, and world (Atkinson, 2011). Humans are continuously and dynamically adapting to their environment, and therefore learning is adaptive—facilitating survival in complex environments. *Alignment* is the major way that learning occurs (Atkinson, 2010) and is defined as “the complex means by which human beings effect coordinated interaction, and maintain that interaction in dynamically adaptive ways” (Atkinson, Churchill, Nishino, & Okada, 2007, p. 169). Cooperative social actions, such as conversation, require the participants to be aligned or synchronized, and this appears to be natural social behavior for humans. Kinsbourne and Jordan (2009) argue that the brain has an *embodied anticipatory system* which enables *entrainment*, i.e., individuals create a context of reciprocating interaction from each other’s behavior. For example, Goodman, Isenhower, Marsh, Schmidt, and Richardson (2005) found that if two people, sitting next to each other in rocking chairs, look at each other, their rocking will become synchronized—even if weight is added to one of the chairs. If they stop looking at each other, they will fall out of synchrony. Bargh and Chartrand (1999) describe the *Chameleon Effect*, where people tended to unconsciously mimic the postures, facial expressions, and behaviors of the people they were interacting with, and found that this made the interaction go more smoothly and created a good feeling between the participants. Semin and Cacioppo (2009) argue that entrainment tends to occur either when the synchronization facilitates some higher goal, or when there is no inhibition from a higher-order goal. Applied to conversation, this would explain how alignment in terms of rhythm of speech and gestures, as well as eye contact, come together to facilitate the co-construction of this complex social action. Language, gesture, and gaze interact dynamically in conversation, but none of them alone can account for the overall behavior of the interaction (Goodwin, 2003).

The research agenda that Atkinson recommends (his *multimodal interaction analysis*) is similar to what was used in the present study, and is what I recommend for future work on fluency. Because of the complex and dynamic interactions involved in the co-construction of a fluent conversation, methods that can examine these interactions as they develop during the course of the conversation are needed. In particular, there is a need for multimodal case studies of individual conversations—whether with proficient speakers of the language, learners of the language, or combinations of both. Considering that two participants may be the most basic context for conversation (Stivers, 2015), dialogues involving a variety of conversational content would be a good place to focus initially.

8.3 Towards a dialogic model of fluency

In this section I would like to review two of the more recent models of fluency proposed by other researchers, and propose my own tentative model based on the examination of the data in this study. Segalowitz's model of fluency is shown in Figure 8.1. This model is based on a cognitive perspective, focusing on speed and stability of processing (*cognitive fluency* in the diagram). However, even though it includes factors external to the speaker which influence the speaker's development, it has as its center *utterance fluency*, which implies that the speaker's L2 production is the final goal. Even though some people might say that utterance fluency or cognitive fluency is their goal in learning a language, for a language learner, conversation is where these abilities are put to the test. A conversation can be felt to go well or go poorly for a host of reasons, but in a certain way a learner's ability to perform in conversation is where they may feel satisfied or frustrated with their language skills, which they would probably blame on their cognitive fluency. Therefore, dialogic fluency is the point where a conversation succeeds or not.

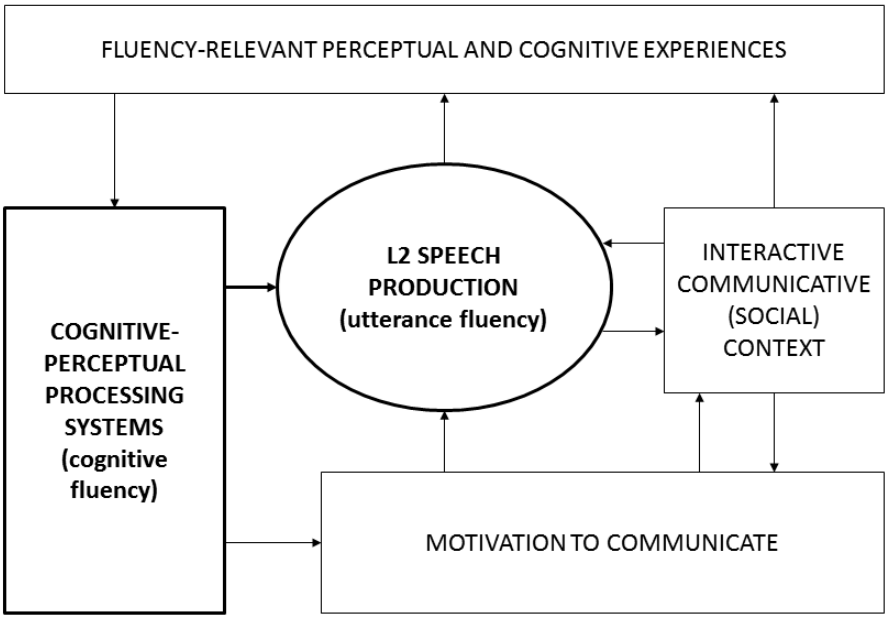

Figure 8.1 Model of fluency from Segalowitz (2010, p. 164)

Rühlemann's (2007) model of conversation is shown in Figure 8.2, and includes five factors: shared context, co-construction, real-time processing, discourse management, and relation management. This model focuses more on the discourse level. Of particular relevance to the present study is that he places *co-construction* at the center of the model. *Shared context* includes the nonverbal features, shared background knowledge, and the immediate environment. *Real-time processing* refers to the pressure that speakers have to produce language in a timely manner within the conversation; a speaker cannot reduce this pressure by simply slowing down or pausing, because another speaker could take their turn—something that happens often with learners. *Discourse management* refers to the use of discourse markers and deixis to maintain the structure of the conversation in the face of pressures of real-time processing. Finally, *relation management* refers to such factors as face and rapport.

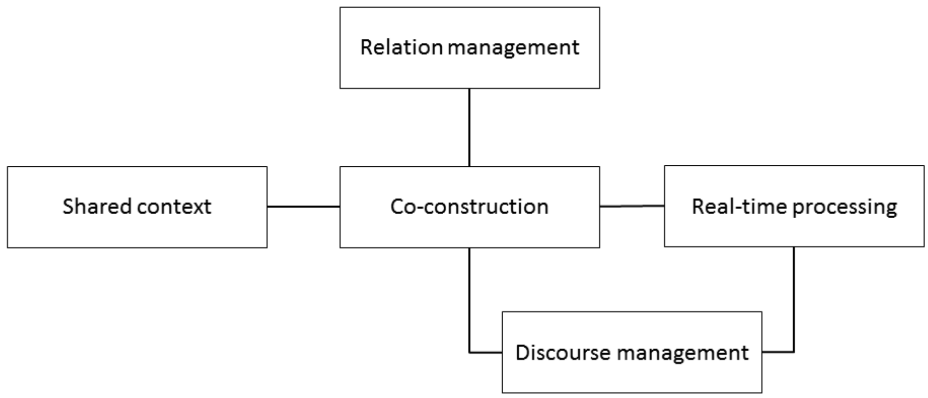

Figure 8.2 Model of fluency from Rühlemann (2007, p. 49)

Dialogic fluency brings together the cognitive systems of both people in the conversation, as well as the surrounding context, such that the speakers are seen as co-participants in the context, rather than as independent organisms. A tentative model of the co-construction of dialogic fluency in the context of retelling a story is shown in Figure 8.3. Speakers are able to adapt their speech to the environment to communicate effectively. This will be limited by the cognitive fluency of the speaker, of course, but will also be limited by the interlocutor and the situation. Gestures and gaze contribute to the communication and help to create confluence in face-to-face conversations. These factors contribute to the co-construction of alignment and dialogic fluency, although the contribution of the factors is not necessarily predictable. Telephone conversations lack the contribution of visual elements, but are still possible, presumably because the factors available in that particular context are employed by the participants to co-construct a confluent conversation within those constraints.

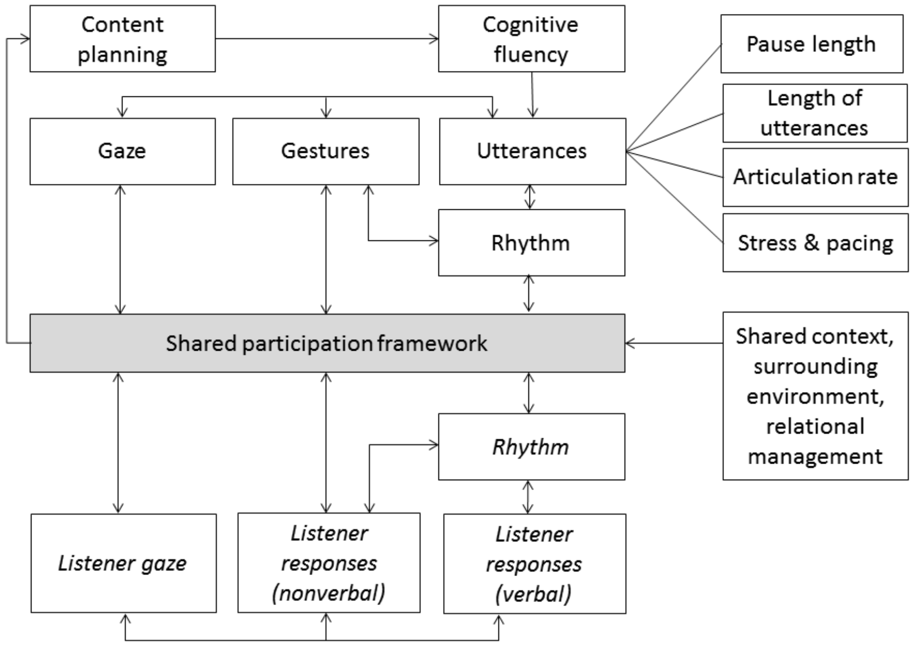

Figure 8.3 Model of dialogic fluency with alignment in the shared participation framework at the center. Assuming a narrative conversation type, the speaker is the top half and the listener is the bottom half.

Rhythm, therefore, may be a central verbal component of conversational alignment, which may help to explain many features of turn-taking. For example, there is discussion in the literature about the difficulties in delineating the difference between a backchannel, which does not take the turn away from the speaker, and a short turn (Campbell, 2007; Duncan, 1974). This kind of difficulty may be a result of the theory itself—that is, the model of conversation where turns are passed back and forth between the participants. If the participants are viewed as co-constructing the conversation, then the length of the listener turns or backchannels serves to further the co-construction of the conversation as a whole. Therefore, the turn-taking mechanism may perform more functions than merely relinquishing a potential turn at a perceived turn relevance point; they also function to create alignment between the speakers in a way that silence (which would also relinquish the turn for the silent person) does not.

In a study of listener response tokens, McCarthy (2003) criticizes the economy of turns proposed by Conversation Analysts (Schegloff, 1982), who argue that turns tend toward the minimum length necessary to achieve their function. McCarthy notes there is much additional language that addresses the relational and interactional aspects of the talk. The results of the present study suggest that one of the functions of backchannels is to create alignment in rhythm, as well as in interactional and relational aspects. In the present study, backchannels were not used when there was no eye contact between the speakers, which suggests that they operate within a larger system than merely the exchange of verbal turns.

In my own experience, I have noticed that the inability to align in rhythm can create problems in conversations. I have participated in teleconferencing meetings with people in other parts of the world, where there was a large perceptible delay due to the distance involved. In these situations, my backchannels are aligned to the delayed interlocutor’s voice that I hear. When my backchannels arrive, they are delayed even more, resulting in them being heard by the interlocutor after they have started their next utterance. This almost always is perceived as an interruption, where the speaker then stops and we have to re-negotiate who has the turn. Although this is only anecdotal, it is an interesting example of what happens when rhythmic alignment is not possible.

8.4 Implications for testing and teaching

Tests of spoken language skills vary in format. Tests such as Versant (Pearson, 2011) or the TOEIC/TOEFL speaking tests (“TOEIC speaking and writing test: For test takers: Test content,” 2016) use a recorded monologue format, while tests such as the ACTFL Oral Proficiency Interview (“Oral Proficiency Interview (OPI) - Language Testing International,” 2016) use online or in-person face-to-face interviews. In spite of these different formats, in all of these cases, the “fluency” of an individual learner is being assessed.

The results of the present study suggest that these types of exams may not be appropriate measures of fluency. First, dialogic fluency is not a property of an individual, but rather a property of the co-constructed conversation as a whole. This suggests that monologic examinations, with no interlocutor present, are poor measures of the ability of a learner to co-construct fluency in conversation. Van Moere (2012) has argued that the listen-and-repeat portion of the computer-graded, monologic Versant test is a strong measure of a learner’s cognitive language abilities. This could be an effective measure of cognitive fluency (reflecting the automaticity of the processes of language production). In the present study, both participants were able to create rhythmic alignment, for example, with Asami having a score of 40 and Keiko having a score of 54 for the *fluency* component of the Versant English test (on a scale from 20 to 80). The Versant scores support some of the differences between the subjects, but they do not measure the ability of the learner to participate in alignment and the co-construction of a conversation.

It would seem, then, that interview-style examinations are the best bet to assess dialogic fluency, however, they too have problems. Smooth conversation is the result of alignment between interlocutors, with both people contributing to the construction of the confluence of the conversation. It is doubtful whether an interviewer would be able to do this for several reasons. First, the interviewer is usually pre-occupied with making notes and judging the performance of the learner. Second, the interviewer is often making an effort to *not* help the speaker. Additionally, interviewees have been shown to receive higher scores when examiners are more supportive (Brown, 2003, 2005). Therefore, these kinds of exams are not necessarily any more “realistic” than monologic exams in terms of a learner’s performance in an actual conversation.

Fujimoto (2012) argues that oral testing of students in class should be peer-to-peer or in small groups, where the conversation is scored as a whole, with all students receiving the same grade, rather than evaluating the contributions of individual participants. She notes that silence by a participant does not necessarily mean lack of participation, as the silent participant may be listening and using nonverbals. She adds that scores can be adjusted for groups with students who do not participate at all. This is an interesting suggestion and may be a fruitful direction for future testing, although the lack of individual scores would cause problems for high-stakes testing. Also, the score for the group would be a holistic measure of conversational proficiency rather than really targeting the particular features of dialogic fluency.

Most of the other exams mentioned above use holistic ratings as well. For example, the TOEIC speaking test recordings are rated by evaluators using a rubric. Even with high inter-rater reliability, ratings are still measures of *perceived fluency*, which has been shown to be influenced by a wide range of factors (Götz, 2013), many of which (such as *perceived accent* and *grammatical accuracy*) are theoretically components of proficiency rather than fluency.

Another finding of the present study is that fluency is dynamically created through a multitude of factors, and therefore is not constant—that is, learners have the ability to vary their language to achieve certain aims, whether that be alignment of rhythm or accuracy in grammatical constructions. A definition of fluency as “the maximally effective operation of the language system so far acquired by the student” (Brumfit, 2000, p. 69) is a very narrow view of how a learner contributes to the interactions that they are involved in, and does not allow for the fluctuations due to the different factors involved in creating a fluent conversation.

McNamara and Roever (2006) have argued that language testing needs a social dimension. I agree and would add that fluency is a complex construct, with monologic and dialogic aspects. Testing needs to encompass all of these. A combination of tests is necessary to address the complex construct that is fluency. Furthermore, descriptors of fluency levels, such as those of the Common European Framework (Council of Europe, 2001), generally fail to include features of dialogic fluency, and need to be updated.

Direct, practical implications for teaching are somewhat difficult to suggest because some aspects of dialogic fluency, such as rhythmic alignment or head nods, may not be teachable. In my own teaching experience, however, I have seen many learners who cannot maintain a fluent conversation, resulting in long silences and leading to conversations that feel painful. Therefore, a consciousness-raising or language awareness approach could be beneficial for learners to know about aspects of alignment and dialogic fluency, in order to better understand and evaluate their own conversational abilities. One example of a potentially beneficial activity is for learners to record themselves in a conversation, and then listen to it or even transcribe it, in order to evaluate their own strengths and weaknesses in creating a smooth conversation. To facilitate their ability to notice these aspects of conversation, it would also be beneficial for learners to study examples of successful conversations, either between second language speakers at a similar level, or speakers of higher proficiency.

Although not a new idea, the results of the present study also re-emphasize the importance of repetition to help learners to develop automaticity. Within the context of *meaningful output*, repetition can occur with activities organized according to the ACCESS framework (Gatbonton & Segalowitz, 2005), where several tasks are organized so that they recycle language and content. Additionally, smaller scale speaking tasks such as the 4/3/2 activity (Nation & Newton, 2009), and repetitive, meaningful practice of useful grammatical constructions (Kirk, 2014) also result in repetitive practice that can help to build cognitive fluency.

Another implication for teachers, is that they should understand that both monologic and dialogic fluency are not constant, and dialogic fluency is not a property of individual learners that develops linearly. Additionally, what may initially appear to be problems in learners' language, may actually be speech management at work. For example, what may seem to be a pronunciation problem, might be a function of maintaining rhythm, or of the use of non-nativelike hesitation devices (Carroll, 2005).

8.5 Limitations and suggestions for further research

There were a number of limitations to the present study. One obvious limitation is that the results are not generalizable due to the case study methodology. The present study included only two learners, with four retellings of the same story for each, yielding only a few minutes of conversational data. However, fluency is a complex construct that may not be able to be explained in terms of linear models with predictor variables and outcome variables, as evidenced by the lack of significant results in much previous research (Riggenbach, 1989; Towell et al., 1996). Conversation may have a kind of redundancy built into the system, with various interacting factors, but without a clear cause-and-effect relationship. Speech rates may vary. Pauses may be intentional, unavoidable, frustrating, or awkward. Heads nods contribute importantly to conversation, but conversation can exist without them. A complex or dynamic systems theory approach is likely a fruitful direction for further research, where researchers must use *retrodiction* rather than *prediction*—that is, researchers seek to explain what has happened in the data, rather than trying to predict what will happen based on initial conditions (Larsen-Freeman, 2011). Van Lier (2010) has also noted the problems with attempting to extract certain variables from natural language data because of the interrelated, ecological nature of language. Case studies, involving multimodal conversational data, and using a variety of methods of analysis, can be an important method to shed light on the complex development of fluency as it unfolds over time. However, it is also important to triangulate results from studies using other methods, such as corpus linguistics.

Many factors can influence interactions, including head and body movement and orientation, tools, settings, and roles and relations (Atkinson, 2011). In the present study, these factors were determined by the conditions of the study—the location chosen for the retellings, the individual interlocutors and their relationship to the subjects, and the set-up of the desk and chairs. Further research is necessary to unravel the potentially very important influence of these effects. For example, the genders of the participants may affect how the fluency of the conversation unfolds. In the present study, there was one male interlocutor, who responded mostly non-verbally, while the three female interlocutors responded both verbally and non-verbally. With the limited data available in the present study, it is impossible to know whether the relative lack of verbal responses by the male interlocutor was typical for male English speakers, particular to the speaker himself, or particular to that single context. Also, in the present study, the native speaker interlocutors were all teachers, and older than the two learners, which likely contributed to an unequal power relationship. The fluency of the conversation as it unfolds would likely be different with an interlocutor who is younger than the speaker, or not a native speaker of English. The personality of the speakers quite likely could be an important factor in the co-construction of dialogic fluency, and both of the subjects were outgoing and talkative. The similarity of their personalities makes it impossible to discern how personality affects dialogic fluency. Further research with other learners is needed to illuminate these factors.

Further research using multimodal case study methods is needed to investigate how learners' fluency varies in different contexts. In the present study, the results show that learners' fluency may vary, not only by genre (monologue vs dialogue), but even within the same task and same genre with different interlocutors at different times. Therefore, a promising direction for future research would be to examine individual learners in different contexts, such as with different topics, interlocutors of different gender and age, and with interlocutors who are other learners. In the present study, the Pear Story retellings occurred after approximately 15 minutes of free conversation between the two participants. This free conversational data was removed from the present study in order to focus the analysis on the controlled part of the task. However, in the future, I intend to examine this data, using the same methodology employed here. This would help to confirm or modify the results of the present study with a dialogic context that is not narrative. Witton-Davies (2012), in a study of Taiwanese learners of English, found that speech rates in discussions were higher in discussion tasks than in narrative tasks. It would be interesting to see if Asami and Keiko's free conversations support these findings.

Finally, it must be pointed out that the present study looks at the two learners at only one point in their language development. Longitudinal studies, following the development of learners' monologic and dialogic fluency features, are sorely needed. I would especially recommend these types of studies for researchers living in countries where the target language is spoken, since it is much more likely that the learners' language will change over a few years' time.

Another difficulty with the present study was with the process of making annotations to the transcription for the analysis. In the present study, pauses were classified according to two categorizations: whether they were preceded by a prosodic boundary tone, and location within the syntactic structure. The goal of the syntactically-based classification was to identify pauses that occurred within phrases that would normally be expected to occur in a single utterance—that is, to identify pauses that broke up phrases. The classification was based on syntactic structure. However, I noticed while manually making those classifications that a pause between the subject and verb in *the man picking pears didn't realize it* and *he didn't realize it* seem to be different in terms of the perceived dysfluency of the interruption. This impression, which was confirmed by asking several other people, seems to imply that there needs to be a further classification in terms of *new information* and *old information*, which is the essential difference between the use of the full phrase and a pronoun. On the one hand, this suggests that further research from an emic perspective would lead to a better categorization scheme for pauses. On the other hand, the causes of pauses and the impressions they impart on other speakers may be a complex, dynamic system that defies attempts at categorization. In the present study, I opted to use the two-level categorization in order to maximize the objectivity of the classifications, while taking into consideration the problems inherent in the system during the discourse analysis in Chapter 6.

The transcriptions, annotations, and rhythm analyses in the present study were carried out entirely by myself. The disadvantage of this method is with reproducibility, that is, whether another researcher would make the same annotations and come to the same conclusions. On the other hand, it has the advantage of a high level of internal consistency, with no need for concerns about inter-rater reliability. That said, as I progressed through the data, my methods for analyzing that data developed and improved. To take this into account, I repeatedly went back to the video and audio recordings to verify the annotations I made, particularly when difficult cases came up that challenged parts of my method. The annotation of rhythm was especially a concern, since rhythm, as a perceptual gestalt (Auer et al., 1999), is not a phenomenon that can be measured objectively with phonetic data, but must be perceived subjectively (Arvaniti, 2009). In order to verify my annotations of rhythm to myself, I asked several other people to listen to short segments of the data, and describe the observable rhythm. This included asking Susan Fiksdal, whose method of rhythmic analysis served as the model for my own, to annotate one segment independently, so that I could verify my own annotation.

Up to this point, I have argued for the case study as the best methodology for further investigation of co-constructed fluency, however, corpus linguistic methods also show promise. The difficulty with investigating spoken language with corpora has, until relatively recently, been the difficulty of obtaining large amounts of transcribed spoken language. Additionally, since dialogic fluency involves such non-verbal aspects of language as gestures, gaze, and head nods, multimodal corpora are needed that incorporate this data. This adds another layer of difficulty to the already time-consuming process of transcribing spoken conversation, not to mention the ambiguity involved in annotation decisions mentioned above. However, methods for studying dialogic fluency with corpora will likely need to be modified from the methods designed to study vocabulary with corpora. Counts of co-occurrences of gaze directed away from the listener with clause-internal pauses, for example, can give some indication that the particular behavior often occurs, but is not very informative to understanding how gaze and pausing and the other relevant factors interact in each instance. Similar to the conclusions of Knight and Adolphs (2008), I would argue that corpus linguistic investigations of dialogic fluency need to be based on theory derived from case studies, and then the results of corpus research can feed back into those theories. In this way, these two methodologies can complement each other to further our understanding of conversation and fluency.

The study of rhythm in speech and rhythmic alignment is very limited in the literature, possibly due to the subjective nature of the analysis. However, the results of the present study suggest that rhythmic alignment is a very important element of conversation and the co-construction of dialogic fluency. Rhythmic alignment may help to elucidate some aspects of conversation that otherwise defy explanation. For example, in the present study, the fact that silent gaps between turns are a necessary part of maintaining rhythmic alignment was shown. Backchannels often overlap with speakers' turns (McCarthy, 2003), and Rühlemann (2007) found that 17% of the occurrences of *yeah* as a listener response, and 10% of *mm*, occurred in overlap. Rühlemann's study was based on the British National Corpus, which has no data about rhythm. It would be interesting for future research to investigate whether overlaps like those occurring in the BNC are a function of rhythm alignment maintained with faster, native speaker speech, as compared to the slower speech in the data of the present study.

Previous research of fluency tends to use ratings by native speakers as the means of deciding whether a particular speaker or particular extract from a monologue or conversation is fluent or not, and to rank speakers in terms of fluency (Götz, 2013; Riggenbach, 1989; Towell et al., 1996). One criticism of the present study could be that I did not use native speaker ratings to evaluate either overall performances of the subjects, or extracts from the conversations. At an early point I considered having native speakers rate extracts from the retellings, but I felt that a more insightful and objective method of analysis was to rely on the data only. One reason for this decision is that *perceived fluency* is an ambiguous construct. It is unclear whether a rater, evaluating dialogic fluency by listening to a recording or watching a video, would have a similar impression to the interlocutor, who is actually participating in the co-construction of that conversation. Another reason for this decision was that asking external raters to compare monologic and dialogic extracts seemed to me to be an unfair task, which at best would yield results that were difficult to interpret. Therefore, I used the computer-rated, purely monologic Versant test as a way to compare the subjects' cognitive fluency, and the analysis of the Pear Story retellings was conducted just as a comparison between the two subjects, rather than attempting to place them on an external scale of fluency, which would likely be subjective.

One final criticism of the present study might also be the lack of data from post-conversation interviews with the participants. Although I did listen to the conversations with the individual participants (learners as well as teachers) immediately after making the recordings, I was generally only able to get their comments on their overall impression of the smoothness of the conversation, and was not able to get comments on the reasons for particular points of interest in the story. One reason for this is that I found that these points of interest do not become salient to the researcher without having transcribed and annotated the data in detail first. The transcription process took months, and by the time I had formulated detailed questions, it was too late to expect the participants to have reliable memories about such things as the reason for a particular hesitation.

Furthermore, it is unclear whether speakers can make reliable comments on their own dysfluencies (Hopper & Drummond, 1990; Lennon, 1990). As a way of evaluating my methodology for myself, I did the Pear Story retellings in monologue and dialogue in Japanese. In order to focus on the main study, I did not analyze this data in detail, but I did listen to it and asked myself about the reasons for hesitations and other apparent dysfluencies. Occasionally, I was very aware of my difficulty at that point in the conversation, but for many cases, I was not. Therefore, I questioned the reliability of self-reflective comments about details in the retellings.

After having completed this study, I have developed a more streamlined process of segmentation, transcription, and annotation, as well as clearer ideas of the categorization systems involved. Therefore, in the future, I would like to repeat some aspects of the present study with the ability to get comments from the speakers themselves. As I did with this study, I also think that self-reflective studies by researchers could be very helpful in clarifying the methods involved.

8.6 Final thoughts

As Riggenbach (1989) noted, fluency is a higher-order concept, with many levels of complexity. One way of looking at fluency is by referring to the outward appearance, that is, *utterance fluency* for monologic speech and *confluence* for dialogic speech. Utterance fluency encompasses the temporal variables of utterances that reflect the speed and smoothness of speech. Confluence reflects the result of a co-constructed conversation, such as the smoothness of turn boundaries. These terms should be retained to give an umbrella term for observable features, although utterance fluency can be construed as a component of confluence. Another way of looking at fluency is through a cognitive lens as does Segalowitz (2010), with *cognitive fluency* as a term reflecting the efficiency and automaticity of the cognitive processes underlying language production. For now, this remains a useful concept, however, it may be, as some suggest (Atkinson, 2011; Garrod & Pickering, 2007), that human cognition is extended and embodied, and therefore not an isolated system that can be described independently. A last way of looking at fluency is as the skills that are needed to successfully co-construct a conversation with a participant, which I refer to as *dialogic fluency*. This encompasses *cognitive fluency* and wide range of other skills needed to co-create alignment with an interlocutor. This includes the ability to establish rhythmic speech and align this rhythm with the interlocutor's speech.

Thus, I have proposed that *fluency* be replaced by the four inter-related concepts of *utterance fluency, cognitive fluency, confluence,* and *dialogic fluency*. However, in the end, the concept of fluency may be somewhat of a distraction from the real issue for future research, that is, the understanding of how conversation, that most fundamental of contexts of language use, is created by speakers. Future research that focuses on alignment and the co-construction of fluent conversation will help to fill a great gap in a field that for a long time has been based on the study of written language and decontextualized language. I hope that the present study has at least highlighted the importance of these aspects of language for future research.

References

Adolphs, S., & Carter, R. (2013). *Spoken corpus linguistics: From monomodal to multimodal*. New York: Routledge, Taylor & Francis Group.

Adolphs, S., & Dahlmann, I. (2007). Pauses as an indicator of psycholinguistically valid multi-word expressions (MWEs)? In *Workshop on a Broader Perspective on Multiword Expressions* (pp. 49–56). Association for Computational Linguistics.

Adolphs, S., Knight, D., & Carter, R. (2015). Beyond modal spoken corpora: A dynamic approach to tracking language in context. In T. McEnery & P. Baker (Eds.), *Corpora and discourse studies: Integrating discourse and corpora*. Palgrave.

Aijmer, K. (1996). *Conversational routines in English: Convention and creativity*. New York: Addison Wesley Longman.

Aijmer, K. (2004). Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies*, *3*(1 special issue), 173–190.

Aijmer, K., & Simon-Vandenbergen, A.-M. (2004). A model and methodology for the study of pragmatic markers. *Journal of Pragmatics*, *36*(10).

Allwood, J., Nivre, J., & Ahlsén, E. (1990). Speech management—on the non-written life of speech. *Nordic Journal of Linguistics*, *13*(1), 3–48. https://doi.org/10.1017/S0332586500002092

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning*, *38*, 561–613.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1–2), 46–63. https://doi.org/10.1159/000208930

Atkinson, D. (2010). Extended, embodied cognition and second language acquisition. *Applied Linguistics*, *31*(5), 599–622. https://doi.org/10.1093/applin/amq009

Atkinson, D. (2011). A sociocognitive approach to second language acquisition: How mind, body, and world work together in learning additional languages. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (1st ed, pp. 143–166). London ; New York: Routledge.

Atkinson, D., Churchill, E., Nishino, T., & Okada, H. (2007). Alignment and interaction in a sociocognitive approach to second language acquisition. *The Modern Language Journal*, *91*(2), 169–188. https://doi.org/10.1111/j.1540-4781.2007.00539.x

Auer, P., Couper-Kuhlen, E., & Müller, F. (1999). *Language in time: The rhythm and tempo of spoken interaction*. New York: Oxford University Press.

Baddeley, A. (1992). Working memory. *Science*, *255*, 556–559.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462–479.

Bavelas, J. (2000). Nonverbal aspects of fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 91–101). Ann Arbor: University of Michigan Press.

Bavelas, J., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*(6), 941–952.

Bell, C. D. S. (2003). L2 speech rate in monologic and dialogic activities. *Linguagem & Ensino*, *6*(2), 55–79.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*(4), 1165–1188. https://doi.org/10.1214/aos/1013699998

Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing*, *27*(3), 355–377. https://doi.org/10.1177/0265532210364404

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Blakemore, D. (1987). *Semantic constraints on relevance*. Blackwell.

Blakemore, D. (1992). *Understanding utterances: An introduction to pragmatics*. Wiley.

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer (Version 5.3.16) [computer program]. Retrieved May 23, 2012, from http://www.praat.org

Bohle, U. (2007). *Das wort ergreifen - das wort übergeben. Explorative studie zur rolle redebefleitender gesten in der organisation des sprecherwechsels*. Berlin: Weidler.

Boomer, D. S., & Dittman, A. T. (1962). Hesitation pauses and juncture pauses in speech. *Language and Speech*, *5*, 215–220.

Bosch, L. ten, Oostdijk, N., & Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Communication*, *47*(1–2), 80–86. https://doi.org/10.1016/j.specom.2005.05.009

Brinton, L. (1990). The development of discourse markers in English. In J. Fisiak (Ed.), *Historical linguistics and philology* (pp. 45–71). Berlin: Walter de Gruyter.

Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, *20*(1), 1–25. https://doi.org/10.1191/0265532203lt242oa

Brown, A. (2005). *Interviewer variability in language proficiency interviews*. Frankfurt: Peter Lang.

Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge: Cambridge University Press.

Brumfit, C. (2000). Accuracy and fluency: The basic polarity. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 61–73). Ann Arbor: University of Michigan Press.

Butcher, A. (1980). Pause and syntactic structure. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 85–90). Paris: Mouton Publishers.

Campbell, N. (2007). Approaches to conversational speech rhythm: Speech activity in two-person telephone dialogues. In *Proceedings of the XVIth International Congress of the Phonetic Sciences, Saarbrucken, Germany* (pp. 343–348).

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 conference* (pp. 199–202). Aix en Provence: Laboratoire Parole et Langage.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. Richards & R. Schmidt (Eds.), *Langauge and communication* (pp. 2–27). New York: Longman.

Carroll, D. (2000). Precision timing in novice-to-novice L2 conversations. *Issues in Applied Linguistics*, *11*(1), 67–110.

Carroll, D. (2004). Restarts in novice turn beginnings: Disfluencies or interactional achievements? In R. Gardner & J. Wagner (Eds.), *Second language conversations* (pp. 201–220). London: Continuum.

Carroll, D. (2005). Vowel-marking as an interactional resource in Japanese novice ESL conversation. In K. Richards & P. Seedhouse (Eds.), *Applying conversation analysis*. Houndsmills and New York: Palgrave Macmillan.

Carter, R. (2004). *Language and creativity: The art of common talk*. London ; New York: Routledge.

Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics*, *16*(2), 141–158. https://doi.org/10.1093/applin/16.2.141

Chafe, W. (1980a). Some reasons for hesitating. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 169–180). Paris: Mouton Publishers.

Chafe, W. (Ed.). (1980b). *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Chafe, W. (1992). Intonation units and prominences in English natural discourse. In *Proceedings of the IRCS Workshop on Prosody in Natural Speech*. Philadelphia: University of Pennsylvania.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Chambers, F. (1997). What do we mean by fluency? *System*, *25*(4), 353–544.

Clark, A. (2011). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford: Oxford University Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Corley, M., & Hartsuiker, R. J. (2011). Why um helps auditory word recognition: The temporal delay hypothesis. *PLoS ONE*, *6*(5), e19792. https://doi.org/10.1371/journal.pone.0019792

Corley, M., MacGregor, L., & Donaldson, D. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*(3), 658–668.

Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of *um*. *Language and Linguistics Compass*, *2*(4), 589–602. https://doi.org/10.1111/j.1749-818X.2008.00068.x

Council of Europe. (2001). *Common European Framework of Reference for Langauges: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Couper-Kuhlen, E. (1993). *English speech rhythm: Form and function in everyday verbal interaction*. Amsterdam: Benjamins.

Couper-Kuhlen, E. (2001). Intonation and discourse: Current views from within. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 13–34). Malden, MA: Blackwell.

Creswell, J. W., Plano Clark, V. L., & Garrett, A. L. (2008). Methodological issues in conducting mixed methods research designs. In M. M. Bergman (Ed.), *Advances in Mixed Methods Research* (pp. 66–83). Sage.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparison between read and spontaneous speech. *Journal of the Acoustical Society of America*, *111*(6), 2862–2873.

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*(3), 515–535. https://doi.org/10.1353/lan.2006.0130

Dechert, H. W. (1980). Pauses and intonation as indicators of verbal planning in second-language speech productions. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 269–285). Paris: Mouton Publishers.

Deese, J. (1980). Pauses, prosody, and the demands of production in language. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 69–84). Paris: Mouton.

Degand, L., & Simon, A. C. (2009). On identifying basic discourse units in speech: Theoretical and empirical issues. *Discours*, (4). Retrieved from http://discours.revues.org/index5852.html

Del Saz Rubio, M. M. (2007). *English discourse markers of reformulation*. Bern ; New York: Peter Lang.

Derwing, T., Rossiter, M. J., Munro, M., & Thomson, R. (2004). L2 fluency: Judgments and different tasks. *Language Learning*, *54*(4), 655–679.

Deschamps, A. (1980). The syntactical distribution of pauses in English spoken as a second language by French students. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 255–262). Paris: Mouton.

Dewaele, J. M. (2002). Individual differences in L2 fluency: The effect of neurobiological correlates. In V. Cook (Ed.), *Portraits of the L2 user* (pp. 219–250). Clevedon: Multilingual Matters.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford ; New York, N.Y: Oxford University Press.

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.

Dörnyei, Z. (2011). Researching complex dynamic systems: “Retrodictive qualitative modelling” in the language classroom. *Language Teaching*, *FirstView*, 1–12. https://doi.org/10.1017/S0261444811000516

Duncan, S. (1974). On the structure of speaker–auditor interaction during speaking turns. *Language in Society*, *3*(2), 161–180. https://doi.org/10.1017/S0047404500004322

Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on Fluency* (pp. 287–313). Ann Arbor: University of Michigan Press.

Ekman, P., & Friesen, W. V. (1978). *The facial action coding system*. Palo Alto, California: Consulting Psychologists Press.

ELAN — language archiving technology. (2012). (Version 4.3.2). Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. Retrieved from http://www.lat-mpi.eu/tools/elan/

Erman, B. (2007). Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics*, *12*(1), 25–53.

Esser, U. (1996). *Oral language testing: The concept of fluency revisited*. Lancaster University,.

Evison, J., & McCarthy, M. (2013). Social talk. In A. Barron & K. Schneider (Eds.), *Pragmatics of discourse*. Berlin: Mouton de Gruyter.

Færch, C., & Kasper, G. (1984). Pragmatic knowledge: Rules and procedures. *Applied Linguistics*, *5*, 214–225.

Fiksdal, S. (1990). *The right time and pace: A microanalysis of cross-cultural gatekeeping interviews*. Norwood, N.J.: Ablex.

Fiksdal, S. (2000). Fluency as a function of time and rapport. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 128–140). Ann Arbor: University of Michigan Press.

Fillmore, C. (1979). On fluency. In C. Fillmore, D. Kempler, & W.-S. Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85–101). New York: Academic Press.

Fillmore, C. (2000). On fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 43–60). Ann Arbor: University of Michigan Press.

Finlayson, S., Forrest, V., Lickley, R., & MacKenzie Beck, J. (2003). Effects of the restriction of hand gestures on disfluency. In R. Eklund (Ed.), *Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop, 5-8 September 2003, Göteborg University, Sweden* (pp. 21–24). Gothenburg: Gothenburg Papers in Theoretical Linguistics.

fluent. (2008, November 17). Retrieved from http://www.merriam-webster.com/dictionary/fluent

Fox Tree, J. E. (2001). Listeners' uses of *um* and *uh* in speech comprehension. *Memory & Cognition*, *29*(2), 320–326.

Fox Tree, J. E. (2010). Discourse markers across speakers and settings. *Language and Linguistics Compass*, *4*(5), 269–281. https://doi.org/10.1111/j.1749-818X.2010.00195.x

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, *31*(7), 931–952. https://doi.org/10.1016/S0378-2166(98)00101-5

Freed, B. (1995). What makes us think that students who study abroad become fluent? In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Amsterdam: John Benjamins Publishing Company.

Fujimoto, D. (2012). Challenging mainstream thinking of fluency: Its definition and assessment. In K. Bradford-Watts, R. Chartrand, & E. Skier (Eds.), *The 2011 Pan-SIG Conference proceedings* (pp. 108–115). Matsumoto: JALT.

Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing*, *13*(2), 208–238.

Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics*, *28*(3), 410–439.

Garrod, S., & Pickering, M. (2004). Why is conversation so easy? *Trends in Cognitive Science*, *8*(1), 8–11.

Garrod, S., & Pickering, M. J. (2007). Automaticity of language production in monologue and dialogue. In A. S. Meyer, L. R. Wheeldon, & A. Krott (Eds.), *Automaticity and control in language processing* (pp. 1–20). New York: Psychology Press.

Gasser, M. (1990). Connectionism and universals of second language acquisition. *Studies in Second Language Acquisition*, *12*, 179–99.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review*, *61*(3), 325–353.

Goldman-Eisler, F. (1967). Sequential temporal patterns and cognitive processes in speech. *Language and Speech*, *10*, 122–32.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Goodman, J. R. L., Isenhower, R. W., Marsh, K. L., Schmidt, R. C., & Richardson, M. J. (2005). The interpersonal phase entrainment of rocking chair movements. In H. Heft & K. L. Marsh (Eds.), *Studies in perception and action VIII: Thirteenth International Conference on Perception and Action* (pp. 49–53).

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 97–121). New York: Irvington.

Goodwin, C. (2003). The body in action. In J. Coupland & R. Gwin (Eds.), *Discourse, the body, and identity* (pp. 19–42). New York: Palgrave Macmillan.

Götz, S. (2007). Performanzphänomene in gesprochenem Lernerenglisch: Eine korpusbasierte Pilotstudie. *Zeitschrift Für Fremdsprachenforschung*, *18*(1), 67–84.

Götz, S. (2013). *Fluency in native and nonnative English speech*. Amsterdam: John Benjamins.

Gregersen, T., Olivares-Cuhat, G., & Storm, J. (2009). An examination of L1 and L2 gesture use: What role does proficiency play? *The Modern Language Journal*, *93*(2), 195–208. https://doi.org/10.1111/j.1540-4781.2009.00856.x

Griffiths, R. (1991). Pausological research in an L2 context: A rationale, and review of selected studies. *Applied Linguistics*, *12*(4), 345–364. https://doi.org/10.1093/applin/12.4.345

Guillot, M.-N. (1999). *Fluency and its teaching*. Clevedon: Multilingual Matters Ltd.

Gullberg, M. (1998). *Gesture as a communication strategy in second language acquisition*. Lund: Lund University Press.

Gullberg, M. (2008). Gestures and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition*. New York: Routledge.

Gumperz, J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.

Hall, J. K., & Pekarek Doehler, S. (2011). L2 interactional competence and development. In J. K. Hall, J. Hellermann, & S. Pekarek Doehler (Eds.), *L2* *interactional competence and development* (pp. 1–15). Bristol & Buffalo: Multilingual Matters.

Hansson, P. (1998). Pausing in spontaneous speech. In P. Branderud & H. Traunmüller (Eds.), *Proceedings FONETIK 98* (pp. 158–161). Stockholm: Department of Linguistics, Stockholm University.

Hashemi, M. R. (2012). Reflections on mixing methods in applied linguistics research. *Applied Linguistics*, *33*(2), 206–212. https://doi.org/10.1093/applin/ams008

Hasselgreen, A. (2005). *Testing the spoken English of young Norwegians: A study of test validity and the role of "smallwords" in contributing to pupils' fluency*. New York: Cambridge University Press.

Hedge, T. (1993). Key concepts in ELT (fluency). *ELT Journal*, *47*(3), 275–276.

Holliday, A. (2002). *Doing and writing qualitative research*. London ; Thousand Oaks, Calif: SAGE.

Hopper, R., & Drummond, K. (1990). Emergent goals at a relational turning point: The case of Gordon and Denise. In K. Tracy & N. Coupland (Eds.), *Multiple goals in discourse* (pp. 39–66). Clevedon: Multilingual Matters.

Kachru, Y. (1994). Monolingual bias in SLA research. *TESOL Quarterly*, *28*(4), 795–800. https://doi.org/10.2307/3587564

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, *94*(4), 554–566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge and New York: Cambridge University Press.

King, J. E. (2011). *Silence in the second language classroom* (Ph.D. (Doctor of Philosophy)). University of Nottingham, Nottingham, UK.

Kinsbourne, M., & Jordan, J. S. (2009). Embodied anticipation: A neurodevelopmental interpretation. *Discourse Processes: A Multidisciplinary Journal*, *46*, 103–126.

Kirk, S. (2014). Addressing spoken fluency in the classroom. In T. Muller, J. Adamson, S. Herder, & P. Brown (Eds.), *Exploring EFL fluency in Asia* (pp. 101–119). Palgrave Macmillan.

Kirsner, K., Dunn, J., Hird, K., Parkin, T., & Clark, C. (2002). Time for a pause... In *Proceedings of the 9th Australian International Conference on Speech Science & Technology, Melbourne, December 2 to 4, 2002* (pp. 52–57). Australian Speech Science and Technology Association Inc.

Kjellmer, G. (2003). Hesitation: In defence of er and erm. *English Studies*, *84*(2), 170–198. https://doi.org/10.1076/enst.84.2.170.14903

Knight, D. (2011). *Multimodality and active listenership: A corpus approach*. London: Continuum.

Knight, D., & Adolphs, S. (2008). Multi-modal corpus pragmatics: The case of active listenership. In J. Romero Trillo (Ed.), *Pragmatics and corpus linguistics: A mutualistic entente* (pp. 175–190). Berlin: Mouton de Gruyter.

Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 5–24). Ann Arbor: University of Michigan Press.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*, 145–164.

Kousidis, S., & Dorran, D. (2009). Monitoring convergence of temporal features in spontaneous dialogue speech. Dublin Institute of Technology: Digital Media Centre Conference Papers. Retrieved from http://arrow.dit.ie/cgi/viewcontent.cgi?article=1003&context=dmccon.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *35*, 523–529.

Kraut, R. E., Lewis, S. H., & Swezey, L. W. (1982). Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology*, *43*, 718–731.

Labov, W., & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts: Proceedings of the 1966 annual spring meeting of the American Ethnological Society* (pp. 12–44). Seattle: University of Washington Press.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford ; New York: Oxford University Press.

Lapadat, J. C. (2000). Problematizing transcription: Purpose, paradigm and quality. *International Journal of Social Research Methodology*, *3*(3), 203–219.

Larsen-Freeman, D. (2011). A complexity theory approach to second language development/acquisition. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (1st ed, pp. 48–72). London and New York: Routledge.

Larsen-Freeman, D., & Cameron, L. (2008a). *Complex systems and applied linguistics*. Oxford: Oxford University Press.

Larsen-Freeman, D., & Cameron, L. (2008b). Research methodology on language development from a complex systems perspective. *The Modern Language Journal*, *92*(2), 200–213. https://doi.org/10.1111/j.1540-4781.2008.00714.x

Lehtonen, J. (1978). On the problems of measuring fluency. In M. Leiwo & A. Rasanen (Eds.), *AFinLA yearbook 1978* (pp. 53–68). Jyvaskyla: AFinLA.

Lehtonen, J., Sajavaara, K., & May, A. (1977). *Spoken English: The perception and production of English on a Finnish-English contrastive basis*. Jyväskylä: Gummerus.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*, 387–417.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). Ann Arbor: University of Michigan Press.

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *BioScience*, *51*(5), 341–352. https://doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2

Loban, W. (1963). *The language of elementary school children: A study of the use and control of language and the relations among speaking, reading, writing, and listening*. Champaign, Ill.: National Council of Teachers of English.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy tradeoff in skilled behavior. *Psychological Review*, *89*, 483–506.

MacWhinney, B. (2013). The CHILDES project: Tools for analyzing talk - Electronic edition. Part 2: The CLAN programs. Retrieved from http://childes.psy.cmu.edu/

MacWhinney, B. (2014). The CHILDES project: Tools for analyzing talk - Electronic edition. Part 1: The CHAT transcription format. Retrieved from http://childes.psy.cmu.edu/

Markee, N. (2000). *Conversation analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behaviour. *Psychological Review*, *79*, 487–509.

McCafferty, S. G. (2006). Gesture and the materialization of second language prosody. *IRAL - International Review of Applied Linguistics in Language Teaching*, *44*(2), 197–209.

McCarthy, M. (2003). Talking back: “Small” interactional response tokens in everyday conversation. *Research on Language and Social Interaction*, *36*(1), 33–63. https://doi.org/10.1207/S15327973RLSI3601_3

McCarthy, M. (2005). Fluency and confluence: What fluent speakers do. *The Language Teacher*, *29*(6), 26–28.

McCarthy, M. (2008). Spoken fluency revisited. Presented at the 34th JALT International Conference on Language Teaching, Tokyo, Japan.

McCarthy, M. (2009). Rethinking spoken fluency. *Estudios de Lingüística Inglesa Aplicada*, *9*, 11–29.

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Oxford: Blackwell Publishing.

McVeigh, B. J. (1997). *Life in a Japanese women’s college: Learning to be ladylike*. London: Routledge.

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, *32*(2), 129–148. https://doi.org/10.1093/applin/amq035

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Morales-Lopez, E. (2000). Fluency levels and the organization of conversation in nonnative Spanish speakers’ speech. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 266–286). Ann Arbor: University of Michigan Press.

Mota, M. B. (1997). Integrating applied linguistics and cognitive science: A study on human memory and L2 oral fluency. *Revista Intercâmbio*, *6*.

Mota, M. B. (2003). Working memory capacity and fluency, accuracy, complexity, and lexical density in L2 speech production. *Fragmentos: Revista de Língua E Literatura Estrangeiras*, *24*, 69–104.

Müller, F. (1996). Affiliating and disaffiliating with continuers: Prosodic aspects of recipiency. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation* (pp. 131–176). New York: Cambridge University Press.

Mulligan, C. (2005). No English educational reforms will be effective unless Japanese English teacher can and will speak English in the classroom. *The Language Teacher*, *29*(5), 33–35.

Munro, M., & Derwing, T. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, *23*(4), 451–468.

Nakane, I. (2007). *Silence in intercultural communication*. Amsterdam: John Benjamins.

Nambiar, M. K., & Goon, C. (1993). Assessment of oral skills: A comparison of scores obtained through audio recordings to those obtained through face-to-face evaluation. *RELC Journal*, *24*(1), 15–31.

Nation, I. S. P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. New York: Routledge.

Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

North, B. (1993). *Scales of language proficiency: A survey of some existing systems*. Strasbourg: Council of Europe.

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*(4), 557–581. https://doi.org/10.1017/S027226310707043X

Oehmen, R., Kirsner, K., & Fay, N. (2010). Reliability of the manual segmentation of pauses in natural speech. In H. Loftsson, E. Rögnvaldsson, & S. Helgadóttir (Eds.), *Advances in natural language processing* (pp. 263–268). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-14770-8_30

O'Keefe, A., & Adolphs, S. (2008). Response tokens in British and Irish discourse: Corpus, context and variational pragmatics. In A. Barron & K. Schneider (Eds.), *Variational pragmatics* (pp. 69–98). Amsterdam: John Benjamins.

O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.

Olynyk, M., Sankoff, D., & d'Anglejan, A. (1983). Second language fluency and the subjective evaluation of officer cadets in a military college. *Studies in Second Language Acquisition*, *5*(2), 213–49.

Oppenheim, N. (2000). The importance of recurrent sequences for nonnative speaker fluency and cognition. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 220–240). Ann Arbor: University of Michigan Press.

Oral Proficiency Interview (OPI) - Language Testing International. (2016). Retrieved June 17, 2016, from http://www.languagetesting.com/oral-proficiency-interview-opi

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. C. Ellis (Ed.), *Implicit and explicit* *learning of languages* (pp. 393–419). London and San Diego: Academic Press.

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191–226). New York: Longman.

Pawley, A., & Syder, F. (2000). The one-clause-at-a-time hypothesis. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 163–199). Ann Arbor: University of Michigan Press.

Pearson. (2011). *Versant English test: Test description and validation summary*. Palo Alto, California: Pearson.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, *27*(2), 169-190-226.

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In M. J. Atkinson & J. Heritage (Eds.), *Structures of social interaction: Studies in conversation analysis* (pp. 57–101). Cambridge: Cambridge University Press.

Puccinelli, N. (2010). Nonverbal communicative competence. In G. Rickheit & H. Strohner (Eds.), *Handbook of communication competence* (pp. 257–275). Berlin: Walter de Gruyter.

R Core Team. (2015). R: A language and environment for statistical computing (Version 3.2.3). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

R Development Core Team. (2012). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Raupach, M. (1980). Temporal variables in first and second language speech production. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 263–270). Paris: Mouton Publishers.

Raupach, M. (1984). Formulae in second language speech production. In H. W. Dechert, D. Mohle, & M. Raupach (Eds.), *Second language productions* (pp. 114–137). Tubingen: Gunter Narr.

Richards, K. (2003). *Qualitative inquiry in TESOL*. New York: Palgrave Macmillan.

Richards, K., Ross, S., & Seedhouse, P. (2012). *Research methods for applied language studies*. Milton Park, Abingdon, Oxon and New York, NY: Routledge.

Riggenbach, H. (1989). *Nonnative fluency in dialogue versus monologue speech: A microanalytic approach* (Ph.D.). University of California, Los Angeles, United States -- California. Retrieved from http://search.proquest.com/pqdt/docview/303681528/D1C1D9A956F84E0CPQ/1?accountid=8018

Riggenbach, H. (1991). Towards an understanding of fluency: A microanalysis of nonnative speaker conversation. *Discourse Processes*, *14*, 423–441.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes*, *65*(3), 395–412. https://doi.org/10.3138/cmlr.65.3.395

Rühlemann, C. (2006). Coming to terms with conversational grammar: Dislocation and dysfluency. *International Journal of Corpus Linguistics*, *11*(4), 385–409. https://doi.org/10.1075/ijcl.11.4.03ruh

Rühlemann, C. (2007). *Conversation in context: A corpus-driven approach*. Continuum.

Rühlemann, C., Bagoutdinov, A., & O'Donnell, M. B. (2011). Windows on the mind: Pauses in conversational narrative. *International Journal of Corpus Linguistics*, *16*(2), 198–230. https://doi.org/10.1075/ijcl.16.2.03ruh

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest schematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735.

Sajavaara, K. (1987). Second language speech production: Factors affecting fluency. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 45–65). Norwood, N.J.: Ablex Publishing Company.

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 71–93). Washington, D.C.: Georgetown University Press.

Schegloff, E. A. (1984). On some gestures' relation to talk. In M. J. Atkinson & J. Heritage (Eds.), *Structures of social interaction: Studies in conversation analysis* (pp. 247–265). Cambridge: Cambridge University Press.

Schegloff, E. A. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organisation. In G. Button & J. R. E. Lee (Eds.), *Talk and social organisation* (pp. 70–85). Clevedon, England: Multilingual Matters.

Schegloff, E. A. (1996). Turn organization: One intersection of grammar and interaction. In E. Ochs, E. A. Schegloff, & S. Thompson (Eds.), *Interaction and grammar* (pp. 52–133). Cambridge: Cambridge University Press.

Schegloff, E. A., Koshik, I., Jacoby, S., & Olsher, D. (2002). Conversation analysis and applied linguistics. *Annual Review of Applied Linguistics*, *22*, 3–31. https://doi.org/10.1017/S0267190502000016

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, *14*, 357–85.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 1–22). Amsterdam: John Benjamins Publishing Company.

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning Memory and Cognition*, *12*, 419–425.

Schweitzer, A., & Lewandowski, N. (2013). Convergence of articulation rate in spontaneous speech. In *Proceedings of InterSpeech 2013*. Retrieved from http://pub.uni-bielefeld.de/publication/2580929

Segalowitz, N. (2007). Access fluidity, attention control, and the acquisition of fluency in a second language. *TESOL Quarterly*, *41*(1), 181–6.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Selting, M. (2013). Verbal, vocal, and visual practices in conversational interaction. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body-language-communication: An international handbook on multimodality in human interaction* (Vol. 1, pp. 589–609). Berlin & Boston: De Gruyter.

Semin, G. R., & Cacioppo, J. T. (2009). From embodied representation to co-regulation. In J. A. Pineda (Ed.), *Mirror neuron systems* (pp. 107–120). New York, NY, US: Humana Press.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Sinclair, J. (2002). Review of The Longman Grammar of Spoken and Written English. *International Journal of Corpus Linguistics*, *6*(2), 339–359. https://doi.org/10.1075/ijcl.6.2.09sin

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

Sperber, D., & Wilson, D. (1995). *Relevance*. Oxford: Blackwell.

Sridhar, S. (1994). A reality check for SLA theories. *TESOL Quarterly*, *28*, 800–805.

Stam, G. (2007). Second language acquisition from a McNeillian perspective. In S. D. Duncan, J. Cassell, & E. Levy (Eds.), *Gesture and the dynamic dimension of language: Essays in honor of David McNeill* (pp. 117–124). Amsterdam: John Benjamins.

Stivers, T. (2015). Is conversation built for two? Presented at the 14th International Pragmatics Conference, Antwerp, Belgium.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *PNAS (Proceedings of the National Academy of Sciences)*, *106*(26), 10587–10592.

Strayer, D. L., & Kramer, A. F. (1990). An analysis of memory-based theories of automaticity. *Journal of Experimental Psychology: Learning Memory and Cognition*, *16*, 291–304.

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, *11*(2), 139–169. https://doi.org/10.1111/j.1468-2958.1984.tb00043.x

Street, R. L., Street, N. J., & Van Kleek, A. (1983). Speech convergence among talkative and reticent three year-olds. *Language Sciences*, *5*(1), 79–96.

Tannen, D. (1989). *Talking voices*.

Tao, H. (2003). Turn initiators in spoken english: A corpus-based approach to interaction and grammar. *Language and Computers*, *46*, 187–207.

Tao, H., & McCarthy, M. J. (2001). Understanding non-restrictive which-clauses in spoken English, which is not an easy thing. *Language Sciences*, *23*(6), 651–677. https://doi.org/10.1016/S0388-0001(00)00026-7

Tauroza, S., & Allison, D. (1990). Speech rates in british english. *Applied Linguistics*, *11*(1), 90–105. https://doi.org/10.1093/applin/11.1.90

ten Have, P. (2007). *Doing conversation analysis* (2nd ed). Los Angeles: Sage.

TOEIC speaking and writing test: For test takers: Test content. (2016). Retrieved June 17, 2016, from https://www.ets.org/toeic/test_takers/speaking_writing/about/content/

Tottie, G. (2011). Uh and um as sociolinguistic markers in British English. *International Journal of Corpus Linguistics*, *16*(2), 173–197. https://doi.org/10.1075/ijcl.16.2.02tot

Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *International Review of Applied Linguistics*, *40*(2), 115–150.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *7*(1), 84–119.

van Compernolle, R. A. (2014). *Sociocultural theory and L2 instructional pragmatics*.

van Gelderen, A. (1994). Prediction of global ratings of fluency and delivery in narrative discourse by linguistic and phonetic measures- oral performances of students aged 11-12 years. *Language Testing*, *11*(3), 291–319. https://doi.org/10.1177/026553229401100304 van Lier, L. (2010). The ecology of language learning: Practice to theory, theory to practice. *Procedia - Social and Behavioral Sciences*, *3*, 2–6. https://doi.org/10.1016/j.sbspro.2010.07.005

Van Moere, A. (2012). A psycholinguistic approach to oral language assessment. *Language Testing*, *29*(3), 325–344. https://doi.org/10.1177/0265532211424478

Vanderplank, R. (1993). “Pacing” and “spacing” as predictors of difficulty in speaking and understanding English. *ELT Journal*, *47*(2), 117–125. https://doi.org/10.1093/elt/47.2.117

Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication*, *50*(2), 81–94. https://doi.org/10.1016/j.specom.2007.06.002

Wennerstrom, A. (2000). The role of intonation in second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 102–127). Ann Arbor: University of Michigan Press.

Wennerstrom, A. (2001). *The music of everyday speech*. New York: Oxford University Press.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer.

Wingate, M. (1987). Fluency and disfluency: Illusion and identification. *Journal of Fluency Disorders*, *12*(2), 79–101.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *Proceedings of Language Resources and Evaluation Conference (LREC)*.

Witton-Davies, G. (2012). The variability of fluency in dialogue and monologue. In *Selected Papers from the Twenty-first International Symposium on English Teaching* (pp. 224–237). Taipei: English Teacher's Association.

Wolf, J. P. (2008). The effects of backchannels on fluency in L2 oral task production. *System*, *36*(2), 279–294.

Wood, D. (2001). In search of fluency: What is it and how can we teach it? *Canadian Modern Language Review*, *57*(4), 573–89.

Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *The Canadian Modern Language Review*, *63*(1), 13–33.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed). Thousand Oaks, Calif: Sage Publications.

Young, R. (2011). Interactional competence in language learning, teaching, and testing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 426–443). London & New York: Routledge.

Glossary of terms

This glossary includes terms that I have defined or used in a particular way throughout the thesis, as well as some selected other terms that were deemed to be potentially confusing or unknown to readers.

alignment	defined by Atkinson (2007, p. 169) as "the complex means by which human beings effect coordinated interaction, and maintain that interaction is dynamically adaptive ways"; for example, two speakers are aligned in rhythm when the rhythm of their individual utterances occur such that the following speaker appears to be a continuation of the rhythm of the previous speaker
articulation rate	rate of syllables per second over a continous utterance
beat	one pulse of a rhythmic pattern
boundary tones	an intonation contour that occurs prior to a silent pause, that is, at the end of an utterance
C-unit	a unit of speech composed of an independent clause and any related dependent clauses (a clausal C-unit), or other utterances that are not clauses (a non-clausal C-unit)
cognitive fluency	fluency of a speaker as conceptualized as the efficiency and automatization of the underlying processes of language production
dialogue	two speakers in face-to-face conversation
dialogic fluency	fluency construed as that of a conversation as a whole, co-constructed by the participants, and the features of conversation and underlying skills of the participants that facilite it
falling boundary tone	a boundary tone where the pitch falls from level of the previous sylllable to the lower range of the speaker's normal speech
false start	a type of hesitation phenomenon involving an interrupted phrase that is abandonned, i.e., not repeated or corrected.

filled pause	verbal, non-lexical hesitation devices, such as *uh* and *um*
formulaic sequence	defined by Wray (2002, p. 9) as “a sequence, continuous or discontinuous, or words or other elements, which is, or appears to be, prefabricated: that is, stored or retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar”
hesitation phenomena	an umbrella term referring to phenomena in speech that have been traditionally associated with planning of speech including repetition, false starts, reformulation, self-correction, vowel lengthening, filled pauses, and unfilled pauses.
high-rising boundary tone	a boundary tone with pitch that starts at the base level of the preceding speech and rises, as in a typical yes/no question in English
listener	the participant in a conversation that does not hold the floor, but who may contribute to the conversation with responses to the speaker
listener responses	short responses by the listener: categorized as (1) engagement, (2) information receipt, (3) continuers, and (4) convergence tokens
low-rising boundary tone	a boundary tone with pitch that starts below the level of previous speech and then rises—typically associated with pausing at points where a comma would be in written language
mean length of run	average length of utterances in words
monologue	one speaker speaking in isolation, that is, without an interlocutor
monologic fluency	fluency as traditionally conceptualized as dependent on the cognitive fluency and utterance fluency of an individual speaker
pace	the rate of stressed syllables per unit of time—this term is used to refer to the speed of the rhythm more generally, while *tempo* is used to refer to the actual measured rate
partially falling boundary tone	a boundary tone where the pitch falls from the level of the previous speech, but not as much as in a falling boundary tone
perceived fluency	a judgment by a listener of the fluency of a speaker
plateau boundary tone	a boundary tone where the pitch is maintained constant at the end of the utterance, usually associated with a cut-off phrase

pruned speech rate	speech rate after removing filled pauses, repeats, restarts, and other apparent dysfluencies
recurrent sequence	a sequence of words that occurs multiple times in the given collection of texts
reformulation	a type of hesitation phenomenon involving repetition of the previous phrase with different words, but without necessarily being a correction of an apparent error
repetition	a type of hesitation phenomenon involving a word or multiple words repeated verbatim.
self-correction	a type of hesitation phenomenon involving the restatement of a word or phrase with at least one morpheme changed, but without changing the basic meaning
smallwords	defined by Hasselgreen (2005, p. 163) as "small words and phrases, occurring with high frequency in the spoken language, that help to keep our speech flowing, yet do not contribute essentially to the message itself"
spacing	the ratio of stressed words to total words
speaker	the participant in a conversation who holds the floor
speech rate	rate of words per minute spoken over the course that the speaker has the floor
tempo	a measurement of the rate of beats per minute
transition relevance place	a point in a conversation where a transition from the current speaker to another speaker is possible—accomplished through a combination of syntax and prosody
turn construction unit	a unit of speech (from Conversation Analysis) that ends at a transition relevance place, generally determined by a combination of syntax and prosody
turn boundary	a point where the current speaker's utterance ends and another speaker's utterance begins, involving a gap, latching, or overlap (note that in the present study minimal verbal listener responses are considered to be turns)
unfilled pause	silent pauses in speech that are not the result of articulation of plosive consonants
utterance	a continuous run of speech between two silent pauses
utterance fluency	fluency as observed in measurable temporal variables such as speech rate and length of uninterrupted utterances

vowel lengthening	a type of hesitation phenomenon where a vowel within a word is held for a perceptibly longer than normal length of time

Appendix A Transcript of pilot study conversation

In the following transcript, *A* is the Japanese learner of English, and *B* is a native speaker of English. See Chapter 3 for details.

```
B   ah so u::h
(520) what do you do in u::h rowleses class.
A   (492) u:h it's so fun.
B   (462) what kind of things do you do.
A   (815) yea:::h
(536) first,  we write do::wn
(260) seven days diary
B   (714) mm hm
A   a:nd the:n we talked about
(264) diary.
B   (420) a[:h ]
A         [u:h]
(411) students each other
B   (216) oh that sou[nds ] fun
A                    [yeah]
yes
B   so what did you do during your seven days.
A   (1032) my seven days?
B   mm hmm in your [seven day diary]
A                 [o:h          ] yea:h
(852) uh:: yest-
(390) yesterday,
(246) oh yeah.
(258) my seven days is almost all days-
(414) i worked.
B   (990) a:[h you have a part] time job?
A          [a part time j-   ]
ye[s. ]
B    [wha]t is your part time job.
A   uh-
(246) i: se:ll-
(252) clothes some clothes.
B   (396) o:h
A   yeah
B   (552) where do you sell clothes.
```

36 A (1038) o::h odaiba.
37 B (498) odaiba.
38 A yeah.
39 B i like that place.
40 A (132) yeah me too.
41 B (126) so are- do you u:h work in a big mall?
42 A (426) [yes.]
43 B [in odaiba?]
44 A yes do you know decks tokyo beach?
45 B (132) yea:h [yeah i've been there.]
46 A [yea:h yeah]
47 A (606) [yeah]
48 B [so] u:h how often do you work.
49 (384) at decks.
50 A (684) decks,
51 (540) uh
52 (168) this-
53 (636) this seven days,
54 B (324) [mm]
55 A [i] work[ed,] five days.
56 B [mm hm]
57 o:h [sounds pretty busy.]
58 A [yes so-] yea:h.
59 B (342) so do you have time to study?
60 A (1170) yeah(hn hn)
61 B ye[s?]
62 A [ye](h)ah
63 B (234) do you [study] on the train?
64 A [yeah]
65 (270) yes.
66 B (234) oh.
67 A and i couldn't sleep,
68 (324) enough.
69 B (294) o:[:h] that's not good.
70 A [yes.]
71 (258) yeah i know but,
72 B (huh huh)
73 A (570) i:
74 (210) i have to:,
75 (756) i have-
76 (252) to:: work more- because
77 (330) i
78 B (1008) mm.
79 A (156) [yeah.]
80 B [you need] money? so are you saving your money for something special?
81 A (733) no.
82 (872) yeah actually,
83 (468) i- i use credit card,
84 B (510) oh really?
85 A (huh huh) yes.
86 (~h)
87 B do you like [to shop?]

88 A   [and-   ]
89 (246) yes.
90 (huh huh ~h)
91 B   o:h
92 A   yes.
93 B   so you use most of your money to buy clothes?
94 A   yes.
95 (huh huh huh ~h huh huh)
96 B   a:h okay.
97 A   yes: [so last] mon[th   ], i: use a lot of money, [to buy.]
98 B   [so y(ou)-]   [uh huh?]
[o:]::h.
99 A   (642) clothes
100 B   (180) m:[:m.]
101 A   [so] this month,
102 (252) so hard.
103 B   (324)   a::[h.]
104 A   (huh huh ~h) [ye]ah.
105 B   so do you have some credit card debt?
106 A   (588) debt?
107 B   (126)   debt? like uh [did you bo]rrow money with your
credit card?
108 A   [what's-   ]
109 (330)   [no no no no.   ]
110 B   [do you have to pay] money back?
111 A   no no no no no. yeah.
112 B   (144) that'[s] good.
113 A   [mm.]
114 yeah.
115 B   (huh huh huh)
116 A   (216)   of course.
117 (~h huh huh huh)
118 B   (hah hah hah hah)
119 A   (408) yeah if i do so,
120 (426) my parents so angry.
121 B   m::m [that]s] good.
122 A   (huh huh)
123 [yeah.]
124 (huh huh)
125 B   (huh huh huh ~h)
126 A   no no no.
127 n[o-]
128 B   [s]o you shouldn't do that,
129 A   yeah.
130 B   (246) but uh sounds like uh so you're working hard,
131 (540) yeah.
132 B   and buying clothes, what kind of uh- where do you buy your
clothes.
133 A   (558) ah u::::m
134 (648) ikebukuro.
135 B   (582) ikebukuro.
136 A   yeah.
137 (846) [i:-]

138 B   [u:]:h-
139 (960) like at uh parco [o::r-]
140 A   [yes.] ye[s.   ]
141 B   [oh rea]lly?
142 A   i like parco.
143 B   (252) uh huh?
144 A   yeah-
145 (744) [yeah.]
146 B   [so   ] uh what kind of fashion do you like.
147 A   (336) yeah dress style.
148 B   (522) dress style like u:[:h-   ]
149 A   [one pie]ce.
150 B   (258) oh o[kay.]
151 A   [yeah.] yeah.
152 B   (162) that sounds fun.
153 A   yeah.
154 (666) hh hh hh
155 B   and uh how often do you go shopping.
156 A   (264) (~hhh) yea:::h
157 (228) if-
158 (438) i'm free,
159 B (822) mm hmm
160 A   u::h i want to go- everyday. but,
161 B   really?
162 A   (huh huh)
163 but,
164 (~h huh huh huh ~h)
165 i'm not free,
166 (270) [everyday so::]
167 B   [a:]:h, s[::]
168 A   [i]::
169 (420) go:: i: go::
170 (642) u::m
171 (534) at twice a week. i go.
172 B   oh [that's a] lot. [m]m so that's why you have to work a lot huh?
173 A   [yeah   ]   [yeah.]
174 (huh huh hah hah)
175 B   (heh heh)
176 A   but, just watching.
177 (360) [d* yeah]
178 B   [ah just] to look at clothe[s?]
179 A   [yes.] yes.
180 B   mm.
181 A   (270) yeah.
182 B   (678) yeah. you know we're very different because i hate shopping.
183 A   (426) why?
184 B   (684) [uh it's] t- it's boring.
185 A   [why?   ]
186 (510) no:: [not] boring.
187 B   [the-]
188 (222) the only kind of uh shopping i like is for books.

A   (216) books?
(huh huh)
oh go[od.]
B   [nn] yeah like chothes shopping,
A   yeah?
B   (312) it uh i hate clothes shopping.
A   (120) so: where do you-
(618) often
(768) go to buy sh-
(366) clothes?
B   (972) nn: i almost never go to buy clothes.
A   wow.
(hoh hoh hoh ~h huh huh ~h huh huh ~h)
B   (heh heh ~h huh heh heh)
A   wow.
B   but nn::
(1380) yeah [sometimes-]
A   [your   ] wife?
B   (264) sometimes my wife will buy me a shirt.
A   oh good.
B   (312) sometimes.
A   ye[ah.]
B   [but] uh-
(528) yeah i don't buy very many new clothes.
A   (hoh hoh hoh)
B   (huh huh)
A   (654) yeah very opposite.
B   (216) ye[ah.]
A   [yeah] (huh)
(750) ye[ah.]
B   [u:]:h
(282) so uh-
(246) u::h where are you from?
A   (984) uh i'm from tokyo.
B   (486) where in tokyo?
A   (630) do you know- monzennakacho?
B   (366) monzennakacho. just the name. i haven't been there.
A   (210) oh yeah.
B   how is what is uh monzennakacho like.
A   (492) u:[:h-   ]
B   [is it a ni]ce place?
A   (132) yes.
(204) shitamachi.
B   (174) ah [ok.   ]
A   [can you un]derstand? yeah.
B   mm hm.
A   shitamachi.
B   so does that mean you're a edoko?
A   (283) yes.
(hah hah hah ~h)
B   yes?
A   yeah, teiandei.
B   (heh heh heh heh)

242 A   (hah hah hah ~h) yeah.
243 (246) yeah.
244 B   so your family is also?
245 A   (162) yes.
246 B   from there?
247 A   (180) yeah.
248 B   (282) so how long has your family been in: tokyo.
249 A   (~hhhh)
250 B   ed[o.]
251 A   [o:]:h very long.
252 (384) very very long.
253 B   (186) that means you don't know?
254 A   (253) ah yeah i don't kno(h)w
255 B   (heh heh heh heh ~h heh heh heh heh)
256 A   (hah hah hah hah)
257 sorry (heh heh) yeah i'm sorry.
258 B   (210) m:[:m   ]
259 A   [so where] are you from?
260 (366) i don't kn[ow.]
261 B   [uh] i'm from the united states?
262 A   oh.
263 (306) good.
264 (336) so [sa]me-
265 B   [l*-]
266 A   (210) to-
267 B   same as [u:h] steve?
268 A   [kirk,]
269 (174) th- yeah.
270 B   (577) and u::h-
271 (540) but he's from seattle,
272 A   (198) yeah.
273 (246) and i'm from alabama.
274 B   (396) [originally.] so alabama is near florida.
275 A   [alabama?   ]
276 (198) m:[:m   ]
277 B   [in the s]outh.
278 A   (234) good place?
279 B   (247) ah it's very rural.
280 A   (450) rural. [oh.]
281 B   [yeah] it's a kind of country place.
282 A   mm good.
283 B   so a::h,
284 (768) there's a lot of nature,
285 (331) that's good.
286 A   goo:d, ye[ah.]
287 B   [and] so a lot of gree:n [gra]:ss and tree:s.
288 A   [mm.]
289 (272) oh so [good,   ]
290 B   [but uh proba]bly is a little boring for someone from: tokyo.
291 A   (huh huh huh huh)
292 (517) [yeah- mm   ]
293 B   [i think uh shopping] is not very good.

294 A   m::m yea:h-
295 B   (hh huh huh huh)
296 (216) [fashion is] better is tokyo i think.
297 A   [and?   ] (hh hah hah hah)
298 (690) yeah. but i want to go to-
299 (240) alabama?
300 B   (334) oh really? [so,] do you want to go to a-
301 A   [yeah.]
302 B   (368) have you been to a foreign country before?
303 A   (522) yes. i: went to canada.
304 B   (672) o:h.
305 A   (216) and guam.
306 B   (642) canada and guam.
307 A   yes, only two.
308 B   (162) a[h,] how long were you in guam.
309 A   [yeah.]
310 (786) u:h about-
311 (510) u::h two weeks.
312 B   (228) mm hm.
313 A   mm half a month.
314 B   (540) and u:h did you meet any people who live in guam?
315 A   yes.
316 (456) i:: i did home stay?
317 B   (402) oh really.
318 A   yes.
319 B   (456) and uh how was your home stay.
320 A   (312) good. canadian,
321 (360) people is so kind,
322 (936) and uh,
323 B   (120) oh but i thought we were talking about guam.
324 A   (402) ah guam?
325 (150) ah guam. sorry.
326 (306) yeah guam-
327 B   (huh huh huh huh)
328 A   (huh hah hah hah hah hah)
329 (330) only three days,

Appendix B Asami Pear Story

Monologue 1

001 ASAMI whole the movie,
002 (605)
003 a man,
004 (420)
005 is
006 (215)
007 taking
008 (389)
009 uh
010 (375)
011 a lot of pears?
012 (541)
013 and <he:> [/]
014 (1300)
015 he: [/-]
016 (878)
017 yeah.
018 (883)
019 um
020 (890)
021 first,
022 (334)
023 &=laughs
024 i don't know why but
025 (568)
026 <a man> [//]
027 (762)
028 um-
029 (300)
030 during,
031 (391)
032 um
033 (180)
034 a man taking
035 (370)
036 uh
037 (818)
038 some pears,
039 (869)
040 a man with
041 (110)
042 goat,
043 (833)
044 uh
045 (1168)

<&kə-> [//]
(130)
came,
(730)
to uh walk [//]
(1264)
walked
(985)
here.
(1320)
a:nd
(618)
just
(425)
<through the> [/]
(1078)
u:m
(390)
through the
(1705)
man <who taki(ng)-> [//] who are taking
(824)
uh
(891)
&=laughs
some pears,
(375)
yeah. i don't know why,
(435)
<and> [///]
(285)
and then,
(919)
a boy, [///]
(110)
a little boy,
(170)
maybe he is
(670)
eight or ni:ne,
(1273)
a:nd <he:> [/]
(1145)
he
(220)
ride,
(375)
a bicycle.
(680)
actually,
(275)
the bicycle
(155)
is bigger than
(575)
the boy?
(570)
yeah.
(426)
an(d)
(185)

he found
(225)
the pears: bag,
(466)
yeah.
(293)
actually
(278)
the man who
(477)
are taking
(403)
pears during the movie
(433)
<is
(770)
u:m
(165)
three bags> [//]
(466)
uh-
(598)
the man has
(337)
three bags
(307)
to:
(1249)
yeah. put in
(240)
&θə- [///]
(360)
many pears,
(367)
yeah. so there a:re three bags.
(514)
and
(210)
the-
(488)
boy, [///] a little boy,
(350)
fin(d)- [//]
(618)
found,
(1112)
the:
(649)
bags,
(1388)
and,
(794)
he:
(190)
robbed.
(634)
the one of
(513)
the bags.
(644)
but,

(606)
yeah.
(500)
the bicycle is bigger than
(130)
the boy,
(330)
and [=! laughing]
(385)
the bag of pears
(940)
is so heavy,
(825)
so
(414)
<it's:
(306)
looks hard to> [/]
(774)
um
(931)
it's looks hard to:
(1197)
um carry.
(1317)
yeah.
(218)
but
(395)
he robbed it,
(640)
and
(190)
then
(669)
<the &mæ-> [//]
(230)
the boy,
(1073)
&=laughs
ɾaidu@u [: ride]
(970)
a bicycle,
(1303)
then,
(630)
u:m
(604)
a: cute girl,
(2386)
oh.
(105)
yeah.
(214)
the man &θæ- [//]
(468)
the boy,
(411)
come across,
(231)
with

(423)
the cute girl,
(445)
yeah.
(232)
she also ride a bicycle,
(702)
then
(892)
the two
(536)
kids.
(901)
u:m
(630)
come across,
(596)
he: watched
(445)
her,
(1140)
him yeah.
(320)
<he:> [/]
(1691)
u:m
(630)
yeah he:
(582)
looked her,
(670)
um
(872)
so long,
(153)
um not so long but
(1149)
yeah.
(226)
he &stɪ- [//] stare.
(390)
s-
(660)
&don- um he [/-]
(805)
yeah.
(180)
anyway he look her.
(1211)
um
(285)
deeply,
(986)
and,
(327)
he
(375)
couldn't find
(361)
the big rock,
(312)

on the street, [/]
(232)
on [/]
(155)
yeah. on the street,
(890)
then,
(923)
eh he: [/]
(599)
drop out? [/] he:
(553)
yeah. drop out,
(1130)
the bicycle,
(384)
and the bag of pears
(340)
uh
(420)
also
(710)
fall down,
(1387)
then
(577)
u:m
(1463)
three boys,
(797)
came,
(290)
to him- [//] the boy,
(1487)
so
(440)
the three boys
(240)
are so
(95)
kind?
(680)
the:y helped to him,
(735)
u:m
(586)
two of the boy
(641)
um picked up pear,
(927)
um which
(447)
was
(406)
fell down,
(446)
and one of the boy,
(215)
yeah he [/]
(1025)
he: [/]

(1942)
u:m he:
(780)
went to: [//]
(225)
went near to:
(413)
the man who
(241)
robbed the bag,
(812)
yeah.
(910)
i think
(206)
the three boys
(482)
are so kind,
(630)
yeah.
(576)
and [//]
(977)
and then,
(522)
u:m
(2055)
um three boys and
(145)
the boy
(1360)
went
(355)
the opposite way,
(517)
but,
(247)
the three boy- [//] one of the three boys,
(876)
uh find
(297)
the
(180)
hat
(412)
which [/]
(760)
u:h which
(225)
fell down,
(592)
yeah actually
(502)
the hat is
(220)
the boy's hat,
(1485)
so
(525)
the one of the
(400)

three boys
(437)
um
(915)
go [//]
(324)
went,
(1604)
to
(500)
give the hat,
(2155)
then,
(427)
the
(1151)
boy, [/]
(1315)
um
(706)
the boy,
(855)
um
(850)
thought
(568)
thank you and he:
(316)
gave,
(170)
him,
(468)
to
(715)
one [//]
(598)
um three pears,
(809)
yeah.
(1547)
mm.
(1330)
and,
(2255)
uh: again,
(967)
these boys,
(365)
uh went opposite way,
(1354)
but,
(637)
yeah.
(864)
um
(468)
the man,
(310)
who are taking pears whole the movie,
(705)
um

(1285)
didn't notice,
(595)
the
(308)
boy.
(713)
yes.
(363)
the boy,
(95)
is
(556)
uh only bad boy.
(730)
yeah. three boys are
(150)
not bad boy,
(418)
but,
(1112)
yeah they- [/]
(1080)
the:y
(590)
went opposite way,
(299)
so
(320)
a bad boy
(501)
um go: [//]
(1001)
went far.
(614)
i mean
(634)
yeah not [/]
(1734)
um not way to
(906)
the:
(250)
man,
(901)
i mean man is
(491)
the
(433)
um
(665)
taking pears man,
(1236)
um but three boys went,
(257)
um
(585)
the way,
(563)
u:h ne:ar the:
(267)

534 man,
535 (878)
536 so
537 (1555)
538 mm.
539 (1182)
540 yeah.
541 (320)
542 when,
543 (640)
544 the man who are teikiŋgu@u [: taking] pears
545 (995)
546 hm:
547 (3721)
548 um
549 (1066)
550 went downstairs,
551 (1651)
552 an(d) the three boys, [/]
553 (1393)
554 um:
555 (1219)
556 uh three boys are eating
557 (790)
558 pear,
559 (1166)
560 so
561 (894)
562 u:h
563 (135)
564 he saw this,
565 (512)
566 and there is
567 (638)
568 no:
569 (650)
570 u:h one bag of pears?
571 (580)
572 so
573 (1592)
574 the man, thought,
575 (260)
576 the three boys
577 (301)
578 are bad guys,
579 (1248)
580 yeah.
581 (323)
582 the movie
583 (787)
584 was
585 (620)
586 not fun.

Appendix C Asami Pear Story

Monologue 2

001 ASAMI a fat man is gathering pears,
002 (620)
003 and <he:> [/]
004 (973)
005 uh
006 (473)
007 yes.
008 (245)
009 he (i)s gathering pears;
010 (775)
011 an(d)
012 (150)
013 next
014 (570)
015 a small boy,
016 (700)
017 showed up, with
018 (588)
019 riding bicycle,
020 (540)
021 the bicycle is very
022 (255)
023 big.
024 (495)
025 as big as the boy.
026 (910)
027 a:nd
028 (1794)
029 <the-> [//] a fat man
030 (780)
031 is
032 (235)
033 gathering pears on the ladder,
034 (296)
035 so <he canno^(570)t
036 (718)
037 watch> [//]
038 (706)
039 he cannot look
040 (379)
041 back.
042 (600)
043 of him,
044 (792)
045 a:nd uh

(988)
a small boy,
(1363)
u:h &fain^(741)dʊ: [: find]
(314)
a:
(348)
some box of <pears,> [//]
(394)
gathered pears,
(569)
and he:
(185)
robbed
(649)
a: box
(285)
of
(991)
pears.
(381)
but a fat man,
(550)
didn't notice that.
(155)
because he
(430)
cannot
(791)
look.
(749)
and,
(520)
u:h
(348)
a small boy,
(782)
left
(393)
there.
(1060)
and next,
(390)
a: cute girl,
(140)
showed up
(420)
with ɾɑidiŋgʊ@u [: riding] bicycle?
(728)
a:nd u:h
(857)
then
(236)
a small boy
(420)
find
(325)
a cute girl,
(360)
he: stared her
(1492)

for long time?
(260)
hm.
(1379)
so: he:
(549)
cannot find a big s:tone,
(798)
in front of him,
(1111)
and he
(215)
slipped up.
(411)
with
(550)
the box of paired [: pears] [*].
(760)
yeah.
(709)
and next,
(881)
three:
(347)
boys,
(257)
shows up,
(916)
and u:h
(271)
they helped to pick up
(476)
pears;
(1091)
a:nd u:h
(1446)
a small boy,
(468)
<geivu:@u [: gave]
(655)
them> [/]
(951)
u:h
(455)
gave them:
(917)
pears,
(753)
<each &bo-> [//] to each boys,
(845)
and u:h left there.
(1060)
a:n(d)
(1284)
then,
(779)
a:
(321)
fat man, noticed that
(633)
someone robbed

168 (997)
169 his
170 (606)
171 box of <paired [: pears] [*].> [//]
172 (445)
173 gathered paired [: pears] [*],
174 (1552)
175 and u:h
176 (933)
177 u:h
178 (365)
179 <three
180 (295)
181 boys,> [//] three kind boy:s
182 (810)
183 <showed up
184 (957)
185 &bə-> [//] uh showed up
186 (185)
187 on? hɪmu@u [: him],
188 (1305)
189 so:
190 (880)
191 a fat man,
192 (175)
193 thought
194 (928)
195 oh. <&dɪ-> [//] these
196 (733)
197 boys robbed my
198 (1859)
199 pears.
200 (1455)
201 finish.

Appendix D Asami Pear Story

Dialogue 1

```
001   ASAMI uh the next part
002   NS1   right yeah you mentioned something about a movie?
003         (100)
004   ASAMI movie:┌yes.┐
005   NS1   └yeah┘ yeah so you're going to ┌tell me about a movie?┐
006   ASAMI                                └yeah the mo┘vie is so boring.
007         (427)
008   NS1   &=laughs
009   ASAMI &=laughs
010         (395)
011   NS1   i'm sorry to hear that. ok.
012   ASAMI &=laughs
013         (320)
014         yeah yeah yeah. it is so short movie,
015         (689)
016         and uh
017         (349)
018         i don't know
019         (95)
020         the:
021         (394)
022         exactly title but i think
023         (175)
024         the title of movie is
025         (445)
026         pear_story.
027         (210)
028         pear,
029         (155)
030         pear.
031         (420)
032   NS1   pear_story.
033         (100)
034   ASAMI pear_┌story.          ┐
035   NS1        └pear pear like┘
036         (135)
037         the fruit? ┌or pair like two.┐ ok. ┌pear story like a┐
fruit. ok.
038   ASAMI            └yes fruit.       ┘     └fruit.           ┘
039         (90)
040         yes.
041         (250)
042         mm mm ┌mm mm.  ┐
043   NS1         └alright,┘
044         (715)
```

045 ASAMI first,
046 (285)
047 <a-> [/]
048 (190)
049 a fat man,
050 (515)
051 gathered;
052 (280)
053 the pears,
054 (240)
055 NS1 ok,
056 (75)
057 ASAMI mm. on a læbɹ@u [: ladder],
058 (450)
059 NS1 mmhm,
060 (535)
061 ASAMI and then,
062 (345)
063 uh: <a: little> [//]
064 (400)
065 a small boy,
066 (100)
067 with bicycle,
068 (660)
069 showed up.
070 (1130)
071 ændʊ@u [: and]
072 (230)
073 the boy is
074 (595)
075 bad.
076 (530)
077 bad boy,
078 (210)
079 because,
080 (350)
081 um
082 (360)
083 duɾiŋgu@u [: during]
084 (395)
085 the fat man;
086 (465)
087 gathered;
088 (265)
089 peazu@u [: pears],
090 (705)
091 he
092 (620)
093 kʊdɪntu@u [: couldn't];
094 (300)
095 <look-> [//]
096 (270)
097 watch
098 (520)
099 uh
100 (385)
101 °nan@s:jpn darou@s:jpn.°
102 (730)
103 back of uh
104 (450)
105 &hɪ- [//] him?

(265)
NS1 mmhm,
(220)
ASAMI mm mm mm so,
(720)
<the-> [//]
(260)
he:
(670)
kʊdɪntu@u [: couldn't];
(240)
noticed the boy,
(260)
NS1 mmhm
(505)
ASAMI so
(775)
<he-> [//] <the boy,> [//]
(305)
the small boy,
(908)
°nan@s:jpn darou@s:jpn.° robbed.
(215)
the
(465)
box of gathered [=! laughing] pear,
(290)
NS1 mm⌈hm, ⌉
ASAMI ⌊&=la⌋ughs
(620)
yes.
(572)
an(d)
(350)
<the bicycle is so big-> [//]
(300)
big tte@s:jpn iu@s:jpn ka@s:jpn ano@s:jpn
(340)
the bicycle is as big æzʊ:@u [:as]
(360)
boy,
(305)
NS1 mmhm,
(170)
ASAMI so <it-> [//]
(510)
°nan@s:jpn darou@s:jpn° it ʃɪmuzʊ@u [: seems]
(175)
difficult to carry,
(565)
NS1 right ⌈yeah.⌉
ASAMI ⌊mm ⌋ mm mm.
(430)
but <she> [//]
(195)
uh <he-> [//]
(290)
<he-> [//] the boy robbed
(125)
the
(295)

```
167        big box
168        (360)
169        of pear┌s,┐
170   NS1         └mm┘hm,
171        (355)
172   ASAMI so he
173        (635)
174        bring
175        (1285)
176        pears box on
177        (685)
178        bicycle,
179        (760)
180   NS1   ┌mmhm,┐
181   ASAMI └and  ┘ he ride on the bicycle,
182        (760)
183   ASAMI and go.
184        &=laughs
185        yeah [=! laughing]
186        (275)
187        yeah ┌it's dif┐ficult to say but
188   NS1       └okay.   ┘
189        (103)
190   ASAMI mm.
191        (335)
192        and then,
193        (437)
194        u:h the cute
195        (183)
196        girl,
197        (496)
198        yeah
199        (290)
200        showed up
201        (283)
202        in front of him?
203        (235)
204   NS1   mmhm,
205        (200)
206   ASAMI and <she izu:@u [: is]> [//]
207        (855)
208        um
209        (775)
210        she is on bicycle too,
211        (160)
212        ændʊ:@u [: and]
213        (120)
214        the boy is on bicycle too,
215        (200)
216   NS1   mmhm,
217   ASAMI so they
218        (485)
219   ASAMI kou@s:jpn
220        (280)
221        nan@s:jpn darou@s:jpn. kou@s:jpn
222        (1060)
223        come across?
224        (425)
225        ┌&te- ┐ mm yes.
226   NS1   └okay?┘
227        (580)
```

ASAMI and uh
(190)
the boy
(825)
θɪnku:@u [: think]
(160)
the girl is so cute ┌so┐
NS1 └mm┘ hm,
(120)
ASAMI <he-> [//]
(205)
he steado@u [: stared].
(420)
her <so &lo->[//]
(305)
for a long time,
NS1 mmhm,
(135)
ASAMI and he
(265)
didn't uotʃɪ@u [: watch]
(190)
&=laughs
(210)
front of
(960)
ano@s:jpn &fr- uh?
NS1 where he's going?
ASAMI yeah yes ┌yes. ┐
NS1 └mmhm,┘
(265)
ASAMI so;
(540)
and there is a big stone,
(490)
&=laughs
NS1 okay,
ASAMI yeah.
(320)
and
&=laughs
(505)
he slipped up.
(475)
NS1 right ┌ye ┐ah.
ASAMI └mm.┘
(155)
mm.
(655)
yeah.
(180)
but the girl;
(245)
couldn't noticed <that> [//]
(380)
<that> [//] &=laughs that [=! laughing] happen.
NS1 &=laughs
okay, alright,
ASAMI &=laughs
so
(295)

he
(520)
slipped up, and the box of
(565)
gathered pears,
(109)
NS1 mmhm,
(430)
ASAMI are doɾopdʊ@u [: dropped]
(245)
too,
(485)
so
(185)
some pears are
(255)
kou@s:jpn;
(295)
gorogoro@s:jpn.
(685)
dropped.
(125)
NS1 mmhm
(365)
ASAMI and then,
(190)
three boys
(470)
uh &kɑɪ- [/-]
(620)
uh showed up and uh
(385)
they helped to pick
(360)
the
(700)
that pears,
(1675)
and uh
(770)
the boy <&s-> [//]
(320)
say thank you ændʊ@u [: and] gɪvu@u [: give].
(180)
samu@u [: some].
(295)
pears.
(355)
┌to ┐ those
NS1 └mmhm,┘
(480)
ASAMI boys?
(120)
NS1 mmhm
(905)
ASAMI and
(260)
<say> [//]
(350)
they say goodbye and uh
(551)

go
(830)
mm? they leave <opposite side;> [//]
(245)
opposite direction,
(240)
NS1 okay,
(95)
ASAMI but unfortunately,
(140)
the boy is okay but those boys are
(775)
goiŋgu@u [: going],
(235)
the direction;
(275)
which
(400)
the fat man,
NS1 mmhm!
(415)
ASAMI collect the pear,
(255)
NS1 ┌right yeah.┐
ASAMI └and ┘ they eat
(465)
pears,
(276)
NS1 uhoh.
(80)
ASAMI mm.
NS1 &=laughs
(89)
ASAMI &=laughs
(360)
and then,
(330)
the fat man,
(155)
noticed
(190)
wow.
(410)
<there> [/-]
(580)
uh: someone robbed
(260)
my box.
(190)
NS1 mmhm
(440)
ASAMI and then,
(425)
three boys,
(530)
&=laughs
(475)
showed up to
(280)
fat man,
(116)

NS1 mmhm,
(74)
ASAMI so fat man
(205)
<thought> [//]
(280)
um
(145)
nan@s:jpn darou@s:jpn <might thought> [//] nn?
(845)
the fat man,
(170)
may think
(660)
these
(575)
three boys are criminal.
&=laughs
NS1 &=laughs
ASAMI mm.
(403)
it's finished.
(231)
NS1 &=laughs
ASAMI &=laughs
NS1 really? ┌so, ┐ [/] so,
ASAMI └yeah┘ [=! laughing]
(590)
NS1 did the fat man say anything to the boys?
(159)
ASAMI no.
(575)
the movie;
(365)
nan@s:jpn darou@s:jpn.
(982)
<any> [//]
(1049)
anybody <doesn't> [//]
(691)
didn't say anything.
(526)
NS1 oh, i see. [/] i ┌see.┐
ASAMI └mm. ┘
(355)
so only moving,
(740)
NS1 mmhm,
ASAMI mm mm mm mm. not [/-]
(680)
so <there-> [//]
(195)
there are no serifu@s:jpn?
(785)
NS1 oh okay. [/]
ASAMI mm mm ┌mm. ┐
NS1 └okay.┘ so just [/] just [/]
(440)
just actions. ┌no talking. okay.┐ oh: man.
ASAMI └just actions. ┘
(480)

┌and sh┐ort movie.
NS1 └okay. ┘
(197)
yeah.
(180)
how long was it.
(310)
just a few minutes? or
(240)
ASAMI yeah few minutes;
(759)
maybe five? or six
(175)
NS1 oh okay.
ASAMI minutes.

Appendix E Asami Pear Story

Dialogue 2

```
001   ASAMI so i would like to
002         (270)
003         tell about short movie,
004   NS2   okay, ┌sure.┐
005   ASAMI       └i   s┘aw it on
006         (195)
007         this internet, ┌it's [/] uh┐
008   NS2                  └ah.        ┘
009         (380)
010   ASAMI it's so short ┌and  ┐
011   NS2                 └mmhm.┘
012         (390)
013   ASAMI uh
014         (300)
015         boring.
016   NS2   (215)
017         really.
018   ASAMI &=laughs
019   NS2   ┌oh no.  ┐
020   ASAMI └because,┘
021         (510)
022         <it-> [//] there is no
023         (245)
024         serifu@s:jpn?
025   NS2   (390)
026         ah. no: script.
027         (115)
028   ASAMI yes. no ┌scrip┐t.
029   NS2           └mmhm.┘
030         (245)
031         ┌i see.┐
032   ASAMI └only  ┘ sound,
033         (338)
034         ┌mm.┐
035   NS2   └ uh┘huh,
036         (354)
037   ASAMI yes but please listen?
038         &=laughs
039   NS2   okay. yeah. ┌sure.┐
040   ASAMI             └yeah.┘
041         (665)
042         first,
043   NS2   uhhuh?
044         (75)
045   ASAMI a fat man,
```

046 (155)
047 <gather> [//]
048 (515)
049 gathered pears,
050 NS2 okay.
051 ASAMI on the ladder,
052 NS2 (247)
053 mmhm,
054 ASAMI mm.
055 (240)
056 so he
057 (435)
058 &kʊ- [//] he can't see back and bottom,
059 NS2 (90)
060 mmhm,
061 (229)
062 ASAMI and then,
063 (271)
064 a small boy,
065 (315)
066 showed up,
067 NS2 mmhm,
068 ASAMI with bicycle.
069 NS2 (203)
070 okay,
071 (103)
072 ASAMI the bicycle is as big as the boy,
073 NS2 (310)
074 mm!
075 (110)
076 ASAMI so <it seems
077 (530)
078 uh to-> [//]
079 (185)
080 it seems hard to carry,
081 (110)
082 NS2 &=laughs
083 ASAMI &=laughs
084 (320)
085 yeah
086 (175)
087 but actually the boy is bad boy,
088 (271)
089 be┌cause,┐
090 NS2 └oh! ┘
091 (260)
092 ASAMI he: [//]
093 (773)
094 he robbed the box of gathered pears,
095 NS2 (250)
096 okay,
097 ASAMI but a fat man
098 (443)
099 couldn't
100 (622)
101 notice.
102 NS2 (99)
103 m┌m:. ┐((.pt))
104 ASAMI └mm mm┘ mm mm ┌mm.┐
105 NS2 └i ┘ see.
106 (325)

ASAMI and he ride bicycle and leave there,
NS2 (215)
oh ┌no. ┐
ASAMI └mm mm┘ mm.
NS2 (224)
huh.
ASAMI yeah.
(300)
then,
(242)
a cute girl,
NS2 uhhuh,
ASAMI showed up ┌in f┐ront of the boy,
NS2 └uhhuh,┘
okay,
(407)
ASAMI and he
(377)
admired
(110)
her beauty,
(299)
┌and he┐ stared her &s- [//] so long time,
NS2 └mm. ┘
mmhm.
(414)
ASAMI uh without lʊkiŋgʊ@u [: looking]
(380)
front,
NS2 (71)
mm!
(225)
ASAMI and then there is a big stone,
NS2 &=gasp ┌oh no.┐
ASAMI └so he┘ [=! laughing] slipped up,
NS2 (130)
oh!
(345)
ASAMI and
(265)
the box of pears
(793)
dropped too?
NS2 (146)
mm.
(150)
ASAMI a:nd the:n,
(265)
he picked up
(1129)
uh those pears,
(573)
uh:
(190)
three boys showed up,
NS2 mmhm,
ASAMI and they
(265)
are so kind, and they helped to pick up some
(620)
pears,

NS2 aw.
(235)
ASAMI so
(300)
a small boy,
NS2 (165)
┌mmhm,┐
ASAMI └a bad┘ boy,
&=laughs
NS2 mmhm,
(560)
ASAMI uh gave,
(868)
uh pears to
(215)
those guys,
NS2 uhhuh,
(220)
ASAMI and they start to leave,
(150)
opposite way,
NS2 (213)
okay.
(105)
ASAMI but unf:ortunately,
(495)
these: three boys,
(620)
<go-> [//] went
(495)
direction
(546)
which
(508)
a fat man,
NS2 (145)
ah
(543)
┌uhoh.┐
ASAMI └is ┘
(230)
yeah gathering ┌pears,┐
NS2 └oh no.┘ huh.
ASAMI &=laughs
(115)
and then,
(431)
um
(440)
a fat man noticed,
(343)
wow. someone
(203)
┌robbed┐ my box,
NS2 └uhhuh?┘
uhhuh,
ASAMI and
(460)
three boys,
(185)

showed up ┌with ea┐ting ┌pears,┐
NS2 └mmhm, ┘ └ mm┘hm,
ASAMI so he thought,
NS2 (90)
mm┌hm,┐
ASAMI └the┘se guys
(210)
are criminal,
NS2 mmhm. ┌i s┐ee.
ASAMI &=laughs └yeah.┘
(826)
that's all.
(195)
&=laughs
NS2 wow.
(515)
ASAMI yeah.
&=laughs
NS2 oh.
ASAMI thank you for listening.
NS2 yeah. no problem. wow. it sounds like a
ASAMI &=laughs
yeah.
NS2 (130)
kind of
(235)
sad story.
(184)
ASAMI yeah! &=laughs
NS2 for the three boys.
ASAMI yes. yes. [=! laughing]
NS2 &=laughs
(735)
hm i see.
ASAMI but i [/] i feel
(350)
wow so boring movie.
NS2 oh.
ASAMI i've never seen.
&=laughs
NS2 &=laughs
(1215)
ASAMI yeah.
NS2 huh.

Appendix F Keiko Pear Story

Monologue 1

KEIKO i'm gonna [: going to] talk about a movie,
(535)
and i try to describe the movie.
(702)
u:m
(962)
u:h i don't know where it is but
(519)
it (i)s
(395)
a really country place.
(585)
and a man,
(310)
is trying to
(737)
<pick> [/]
(1070)
pick up
(215)
some pears, in a tree,
(285)
an(d)
(255)
he (i)s
(805)
uh going up,
(636)
o:n a ladder,
(430)
an(d) try
(376)
to pick some pears.
(694)
an(d)
(428)
he (i)s wearing a hat,
(582)
an(d)
(195)
red scarf,
(421)
an(d)
(615)
i don't know the name of-

046 (499)
047 uh pants?
048 (1540)
049 and then,
050 (190)
051 a boy is <coming.
052 (664)
053 on> [//] uh coming.
054 (353)
055 riding on a bicycle?
056 (559)
057 an(d)
058 (680)
059 <he:>[/]
060 (748)
061 he just stop in front of tree,
062 (550)
063 an(d)
064 (593)
065 he pick
066 (405)
067 a basket.
068 (535)
069 <full of a-> [//] uh full of pears,
070 (585)
071 an(d)
072 (839)
073 uh <he try> [//]
074 (429)
075 he tries to carry the basket.
076 (761)
077 on the bicycle.
078 (645)
079 but the man didn't realize it.
080 (605)
081 an(d)
082 (245)
083 he (i)s just riding
084 (416)
085 on the bicycle,
086 (530)
087 and then,
088 (358)
089 um
090 (440)
091 he realized
092 (484)
093 uh <the boy,> [/]
094 (230)
095 u:h the boy.
096 (675)
097 so a girl,
098 (545)
099 uh who is riding
100 (754)
101 on a bicycle and coming,
102 (488)
103 an(d)
104 (294)
105 u:h
106 (395)

when
(781)
<he pass> [//]
(375)
he passes the girl,
(1185)
uh
(415)
a wind.
(555)
stole
(350)
his hat.
(1135)
an(d)
(1255)
u:h
(435)
he:
(220)
crashed into a rock.
(813)
and then
(205)
going down,
(1210)
a:n(d)
(685)
u:h
(990)
an(d) there (a)re
(310)
three
(520)
boys,
(465)
and they tried to help him. to pick up
(612)
pears, on the ground,
(1220)
u:h
(430)
then
(1221)
u:h <a boy> [//]
(185)
the boy
(335)
jus(t)
(630)
rides
(558)
on the bicycle and gone?
(1410)
a:n(d)
(560)
u:h
(376)
the three children
(1695)
have got

(495)
some pears.
(710)
an(d)
(490)
they were walking
(440)
an(d)
(315)
leaving [=! laughing] and eating
(1200)
then,
(415)
u:m
(905)
u:h
(755)
the man who was picking up
(265)
pears in the tree,
(885)
u:h
(890)
<he:> [/]
(1560)
uh he realized that
(745)
he lost.
(255)
his
(355)
basket full of pears.
(1547)
u:h at the time,
(730)
three children
(205)
just passed.
(835)
and he stared
(660)
at three children but he didn't talk to them;
(1270)
and finish.
(2125)
that (i)s the end of the movie.

Appendix G Keiko Pear Story

Monologue 2

001 KEIKO i'm gonna [: going to] talk about a movie, in countryside.
002 (1017)
003 u:h
004 (2547)
005 it has really big field and mountain?
006 (286)
007 and trees and grass field
008 (776)
009 and there is a tree,
010 (762)
011 u:h
012 (312)
013 a man,
014 (107)
015 who is wearing a hat,
016 (646)
017 and u:h overalls,
018 (820)
019 u:h is trying to pick up
020 (560)
021 some pears on the tree.
022 (740)
023 there is a ladder
024 (184)
025 by the tree,
026 (393)
027 an(d) <he is> [//]
028 (187)
029 u:h he goes up and down and goes up down,
030 (217)
031 repeat it and he try to pick
032 (440)
033 up. some pears,
034 (313)
035 and put them
036 (97)
037 into a basket.
038 (576)
039 a:n(d) <there was-> [//] uh there are
040 (157)
041 three basket,
042 (460)
043 on the ground,
044 (1056)

u:m
(450)
a:nd
(970)
uh the <pear-> [//] uh pears look like really
(300)
tasty.
(280)
an(d)
(231)
nice green.
(1354)
and then,
(293)
a man,
(340)
u:h <drawing> [/]
(785)
drawing a goat,
(276)
with rope?
(389)
is coming and passed
(370)
the trees.
(1776)
a:n(d) just gone,
(950)
uh he looks like younger than the man picking up pears,
(174)
on the tree,
(953)
uh the man picking up pears
(244)
looks like
(193)
u:h forty about forty or fifty year old,
(613)
uh he has beard.
(1083)
a:n(d) looks like big.
(1406)
an:(d) then,
(358)
a boy,
(610)
uh riding on
(107)
a bicycle is coming.
(778)
an(d) stop in front of the tree.
(849)
the man picking up pears didn't realize it.
(737)
an(d)
(180)
the boy,
(407)
u:h just put
(407)
his bicycle on the ground,

(862)
a:n(d)
(683)
u:m
(350)
an(d)
(430)
pick
(626)
a basket full of pears,
(969)
a:nd then,
(407)
u:h <put> [/]
(510)
put it
(186)
on his bicycle.
(837)
an(d) jus(t)
(154)
he has gone,
(470)
but <the man &p-> [//] u:m the man,
(459)
on the tree,
(447)
didn't realize it.
(1467)
and then,
(234)
the boy,
(333)
u:h was
(413)
just riding <on the-> [/]
(630)
on the bicycle,
(890)
u:h
(224)
the <basket> [/]
(1057)
u:h basket
(768)
is
(496)
really big and
(288)
pears
(523)
looks like growing uh:
(394)
out of the basket,
(1073)
<a:n(d)> [/]
(1430)
a:n(d)
(382)
he realized
(650)

167 that
168 (283)
169 a girl,
170 (527)
171 uh <who:: has> [///]
172 (764)
173 uh who might have
174 (645)
175 long hair,
176 (600)
177 uh is coming on the bicycle.
178 (1267)
179 a:n(d)
180 (508)
181 when
182 (667)
183 he passed.
184 (320)
185 her?
186 (863)
187 <&h-> [//] uh <a &wo-> [//] uh
188 (1037)
189 she: created <wind,> [///]
190 (823)
191 disgust, [: a gust] [*]
192 (1263)
193 a:n(d) his
194 (614)
195 hat.
196 (417)
197 is gone.
198 (2520)
199 an(d)
200 (350)
201 <the moment-> [//] at the moment <he> [/]
202 (623)
203 uh he crashed into a rock on the ground,
204 (456)
205 the rock was really big.
206 (470)
207 an(d)
208 (551)
209 <he:
210 (497)
211 &f:> [/] he felled
212 (383)
213 on the ground.
214 (1187)
215 and <his pears> [/]
216 (806)
217 uh his pears were
218 (400)
219 scattering on the ground,
220 (2016)
221 u:h
222 (466)
223 <there was> [//] there are three boys <near the place,>
[//]
224 (517)
225 near the rock.
226 (506)

an(d)
(293)
they tried to help him to pick up pears.
(1040)
a:nd help him to stand up.
(1530)
an(d)
(214)
the boys,
(352)
u:h stealing
(509)
a basket,
(793)
u::h
(1550)
uh basket
(1520)
has gone,
(767)
an(d) three boys,
(520)
u:h
(414)
three boys <started to> [//] started walking?
(463)
and leaving the place?
(434)
an(d)
(217)
actually they (ha)ve got
(484)
three pears.
(506)
for each.
(641)
an(d)
(290)
<he:> [/]
(277)
<he was just walking> [//]
(207)
uh <he-> [//] uh they were just walking and eating
(534)
pears,
(826)
at the moment,
(550)
the guy uh picking up pears,
(594)
realized that
(347)
<he has-> [/]
(530)
he has lost his basket. one of his baskets.
(757)
and he counted his basket. one two three
(1304)
and then,
(494)
u:m he realized that

(553)
three boy-
(500)
<were come-> [//] uh are coming.
(916)
an(d)
(750)
they are eating pears.
(915)
but <he jus(t)-> [///] uh the man picking pears,
(664)
u:m <he just wondering> [//]
(790)
uh he: is just wondering
(636)
oh
(310)
they (a)re eating pears but
(517)
oh i don't know where
(620)
is my basket full of pears.
(490)
an(d)
(290)
children just <passed.> [/]
(670)
passed him,
(1163)
an(d)
(370)
the man just stared at-
(670)
the children.
(1563)
but he didn't get angry or say anything,
(820)
<jus:(t)> [/] just stared at them.
(1606)
so
(233)
that (i)s
(186)
the end of this movie.

Appendix H Keiko Pear Story

Dialogue 1

```
001   NS3   i'm so excited to hear.
002   KEIKO ┌       o┐┌:h i'm gonna [: going to] talk about a┐
movie,
003   NS3   └&=laughs┘└i'm serious.                          ┘
004        okay.
005   KEIKO &=laughs
006        (365)
007        i don't know where it is,
008        (160)
009        but it's really countryside,
010        (395)
011   NS3   m┌m. ┐
012   KEIKO  └a: ┘n(d) mountains and grass and
013        (390)
014        u:m really
015        (483)
016        u:h field, really big field.
017        (159)
018   NS3   mmhm,
019   KEIKO and uh-
020        (425)
021        there is a tree,
022        (285)
023        ┌it's ┐ pear tree.
024   NS3   └mmhm,┘
025        (485)
026   KEIKO and a man,
027        (395)
028        u:h
029        (245)
030        looks like forty or fifty year old?
031        (236)
032   NS3   okay,
033        (120)
034   KEIKO an(d) he is
035        (326)
036        u:h picking up pears.
037        (190)
038        on the tree.
039        (475)
040   NS3   okay,
041   KEIKO an(d) there is a ladder,
042        (564)
043        u:h
044        (353)
```

045 by the-
046 (400)
047 tree,
048 (240)
049 NS3 mmhm,
050 KEIKO an(d) he's <trying to> [/] uh trying to pick up pears an(d)
051 (470)
052 uh put them into basket⌈s?⌉
053 NS3 ⌊mm⌋ hm,
054 (175)
055 KEIKO and there is three basket maybe. &=laughter
056 NS3 mmhm,
057 (575)
058 KEIKO a:n(d)
059 (180)
060 i don't know jus(t)
061 (350)
062 <a man,> [/]
063 (465)
064 u:h a man
065 (315)
066 is
067 (310)
068 taking goat?
069 (340)
070 or sheep?
071 (265)
072 NS3 mmhm,
073 (260)
074 KEIKO with rope?
075 (818)
076 an(d) pass past the tree.
077 (780)
078 NS3 okay,
079 (95)
080 KEIKO &=laughs
081 NS3 wait a different man or the same man.
082 KEIKO oh different man.
083 NS3 a different ⌈man okay. ⌉
084 KEIKO ⌊yeah differ⌋ent man.
085 (485)
086 a:n(d)
087 (725)
088 and then
089 (95)
090 a boy,
091 (255)
092 NS3 mmhm,
093 (65)
094 KEIKO is coming <on the:> [//]
095 (205)
096 on a bicycle.
097 (340)
098 NS3 mmhm,
099 (163)
100 KEIKO the bicycle is really big,
101 (215)
102 NS3 &=laughs
103 KEIKO bigger than him,
104 (140)

NS3 mmhm,
KEIKO and he: just put the
(185)
bicycle on the ground,
(235)
NS3 mm┌hm┐
KEIKO └ a┘nd
(555)
maybe stole?
(220)
a basket?
(320)
full of pears?
NS3 oh!
KEIKO a basket yeah.
NS3 &=laughs
(100)
KEIKO an(d) put
(170)
the basket on the-
(365)
bicycle?
(534)
NS3 ┌ah! ┐
KEIKO └and then just┘ gone,
(723)
but the man, who was picking up pears,
NS3 uhhuh,
KEIKO he didn't realize it.
(230)
NS3 oh!
(255)
KEIKO and then,
(149)
&=laughs
(175)
the boy was
(140)
riding a bicycle?
NS3 mmhm,
(535)
KEIKO an(d) he realized that a girl,
(520)
uh on the bicycle,
(238)
also
(558)
is coming.
(275)
NS3 ┌mmhm, ┐
KEIKO └to h┘im.
(402)
and jus(t)-
(185)
passed.
(490)
NS3 mmhm,
(625)
KEIKO then
(875)
<the-> [///] at the moment

(504)
<his:> [/-] ah.
(185)
a wind stole
(271) his hat,
NS3 mmhm,
(516)
KEIKO and then
(455)
he just watched the hat oh!
(585)
it's gone.
(463)
and then he crashed.
(438)
into a rock.
(265)
┌on the grou┐nd.
NS3 └uh+oh ┘
(533)
KEIKO and then jus(t)
(846)
u:h
(297)
down,
NS3 fell ┌over?┐
KEIKO └yeah ┘ so
(485)
on the ground.
(710)
an(d)
(308)
there are <three boys,> [/]
(587)
an(d) &=laughs three boys
(877)
<near> [/]
(451)
near the place,
(215)
NS3 ┌at the- ┐ near the rock or the tree.
KEIKO └an(d) ┘
u:h near the rock.
NS3 near the rock okay.
KEIKO they tried to help him to pick up pears.
NS3 mmhm,
(542)
KEIKO a:n(d)
(434)
how can i say
(290)
he
(475)
&=laughs jus(t) <draw?> [//]
(841)
<stan(d)> [///] try ┌to help- ┐
NS3 └he tried ┘ to get up?
KEIKO yeah get up,
NS3 okay,
(571)

```
226   KEIKO and then oh thank you or something [=! laughing] ┌an(d)┐
yeah.
227   NS3                                                    └okay,┘
228         (125)
229   KEIKO they didn't talk.
230         (215)
231         anything &=laughs jus(t)
232         (815)
233         jus(t) um
234         (300)
235   NS3   so there's no words.
236   KEIKO no words ┌yeah.┐
237   NS3            └okay.┘
238         (375)
239   KEIKO an(d)
240         (565)
241         <he: um rode-> [//]
242         (500)
243         uh he was riding a bicycle and gone the boy,
244         (65)
245   NS3   mmhm,
246         (455)
247   KEIKO then
248         (305)
249         three children,
250         (825)
251         u:h have got
252         (560)
253         <pears.> [///]
254         (993)
255         ┌each pe┐ars,
256   NS3   └mmhm,  ┘
257         (135)
258         ┌mmhm,┐
259   KEIKO └and  ┘ they: was
260         (225)
261         eating and walking,
262         (1279)
263         and then,
264         (287)
265         u:h <while it,> [//]
266         (215)
267         u:m
268         (864)
269         <while the moment> [//] [=! laughing],
270         (175)
271   NS3   at the mo┌ment?┐
272   KEIKO          └uh at┘ the mom┌ent, ┐
273   NS3                           └mmhm?┘
274         (335)
275   KEIKO u:h the man who was picking
276         (503)
277         up pears,
278         (160)
279   NS3   mmhm,
280         (215)
281   KEIKO realized
282         (555)
283         u:h <he:> [/]
284         (284)
285         he has
```

286 (230)
287 lost his
288 (285)
289 basket?
290 NS3 right.
291 KEIKO yeah.
292 (180)
293 an(d) oh! where is it.
294 (105)
295 NS3 mmhm,
296 (60)
297 KEIKO hm one two three he
298 (425)
299 jus(t) ┌counted ┐ yeah.
300 NS3 └he counted┘
301 (405)
302 KEIKO and then
303 (330)
304 the-
305 (295)
306 u:h three children?
307 (220)
308 NS3 mmhm,
309 (65)
310 KEIKO was coming?
311 (555)
312 an(d) ┌they were e┐ating pears ┌right? ┐
313 NS3 └oh! ┘ └ uh┘+o┌h ┐
314 KEIKO └yeah┘ and
then he was oh
315 (565)
316 they are eating pears but
317 (650)
318 where is a basket.
319 NS3 uhhuh,
320 KEIKO and just wondering.
321 (480)
322 and then they just passed,
323 (335)
324 NS3 mmhm,
325 (816)
326 KEIKO a:n(d) the man jus(t)
327 (240)
328 stared at
329 (700)
330 boys?
331 (235)
332 ┌and o┐:h
333 NS3 └mmhm,┘
334 (540)
335 KEIKO an(d)
336 (417)
337 that's it. &=laughs
338 (150)
339 NS3 an(d) that's it?
340 (75)
341 KEIKO yeah that's the end of the movie.
342 NS3 &=laughs
343 KEIKO &=laughs
344 (910)

NS3   oh ok┌ay,┐
KEIKO   └ ye┘ah.
      (155)
NS3   well i understood,
      (350)
KEIKO really?
NS3   yeah i understood the story,

Appendix I Keiko Pear Story

Dialogue 2

001 KEIKO i'm gonna [: going to] talk about a movie,
002 (60)
003 NS4 okay.
004 KEIKO in countryside.
005 (230)
006 NS4 oh okay.
007 (175)
008 KEIKO yeah actually, i don't know <the-> [///] where it is,
009 (165)
010 NS4 ┌mmhm.┐
011 KEIKO └ but┘ just countryside an(d)
012 NS4 mmhm.
013 KEIKO there are some mountains ┌and ┐ trees grass fields,
014 NS4 └okay.┘
015 mmhm mmhm.
016 KEIKO an(d) a man,
017 (350)
018 u:h
019 (480)
020 who is wearing hat,
021 (270)
022 NS4 mm┌hm.┐
023 KEIKO └and┘ overalls,
024 NS4 mmhm mmhm,
025 (108)
026 KEIKO is trying to pick up some pears on the tree.
027 (295)
028 NS4 okay,
029 (155)
030 KEIKO an(d)
031 (300)
032 uh there is a <ladder.> [/]
033 (482)
034 <ladder?> [/]
035 (305)
036 NS4 okay.
037 (170)
038 KEIKO uh
039 (170)
040 ladder
041 (190)
042 by the
043 (313)
044 the tree?
045 (95)

046 NS4 mmhm.
047 KEIKO an(d) then
048 (465)
049 <he:> [/]
050 (730)
051 u:h he looks like f:orty or fifty year old,
052 NS4 okay.
053 (130)
054 KEIKO yeah and
055 (270)
056 yeah he just pick up
057 (155)
058 &=laughs some pears.
059 NS4 okay.
060 (218)
061 KEIKO and then,
062 (565)
063 <a man who:> [/-]
064 (375)
065 uh how can you say
066 (530)
067 <a man> [/] a man
068 (260)
069 looks like
070 (635)
071 um younger than the man,
072 (160)
073 NS4 okay.
074 (150)
075 KEIKO picking pears,
076 (113)
077 NS4 mmhm.
078 (65)
079 KEIKO uh is
080 (970)
081 drawing
082 (365)
083 a goat,
084 (310)
085 with rope?
086 (265)
087 NS4 oh okay okay.
088 (170)
089 KEIKO yeah and then
090 (240)
091 pass:es the ┌tree.┐
092 NS4 └mmhm,┘
093 (460)
094 KEIKO just passes.
095 NS4 okay.
096 KEIKO &=laughs they didn't talk ┌maybe [///] ┐
097 NS4 └just right past┘ him
mm┌hm. ┐
098 KEIKO
└yeah┘ yeah.
099 (556)
100 and then,
101 (110)
102 a boy,
103 (220)

```
104   NS4   mm┌hm,┐
105   KEIKO   └uh ┘ riding on the bicycle?
106         (205)
107         is coming.
108         (165)
109   NS4   uhhuh.
110         (95)
111   KEIKO an(d) stop in front of the tree,
112         (200)
113   NS4   mmhm.
114   KEIKO an(d) he just put his bicycle on the ground.
115   NS4   mmhm mmhm.
116         (140)
117   KEIKO an(d)
118         (570)
119         i think that he:
120         (440)
121         jus(t)
122         (295)
123         try to steal,
124         (315)
125   NS4   ┌mmhm.┐
126   KEIKO └a    ┘ basket of
127         (441)
128         the pears.
129         (70)
130   NS4   oh!
131   KEIKO yeah there are some basket in there,
132   NS4   mmhm,
133         (110)
134   KEIKO uh full of pears ┌an(d)┐
135   NS4                    └mmhm.┘
136         (415)
137   KEIKO he tried to pick up one of them.
138         (625)
139         an(d)
140         (357)
141         put
142         (245)
143         on his bicycle
144   NS4   uhhuh.
145   KEIKO an(d) jus(t) gone.
146           (65)
147   NS4   wow.
148   KEIKO yeah and the man uh picking up pears ┌he did┐n't realize
it.
149   NS4                                        └mmhm, ┘
150         (140)
151         oh o┌kay. ┐
152   KEIKO     └yeah.┘
153         (485)
154         an(d)
155         (235)
156         i thought at firs(t)
157         (445)
158         u:h he
159         (295)
160         tried to help him.
161         (655)
162         ┌to car┐ry the ┌basket but i think yeah.┐
163   NS4   └uhhuh.┘       └oh i see i see.         ┘
```

(320)
KEIKO just he try to
(320)
steal it.
(85)
NS4 oh.
KEIKO yeah.
(164)
and then,
(180)
the boy was riding was riding the bicycle?
NS4 mmhm,
(285)
KEIKO an(d)
(850)
and he realized that a girl,
(145)
NS4 ┌mmhm,┐
KEIKO └ who┘: is
(435)
i forget-
(325)
the dress or a skirt or something ┌an(d)┐
NS4 └mmhm,┘
(335)
KEIKO um <she,> [///]
(140)
a girl ┌is ┐ coming on the bicycle?
NS4 └mmhm,┘
mmhm,
(215)
KEIKO and just passed.
(210)
NS4 okay.
(895)
KEIKO at the momen(t)
(270)
um she the girl created
(295)
a wind gust,
(195)
NS4 uhhuh,
(120)
KEIKO an(d)
(315)
his hat
(365)
was gone.
(320)
NS4 oh! uhhuh huh ┌huh. ┐
KEIKO └an(d)┘ <he jus(t)> [/]
(270)
u:h how can you say.
(220)
he:
(455)
u:m jus(t)
(468)
<f::ell?> [/]
(275)
fell down on the ground,

```
225   NS4   okay.
226         (305)
227   KEIKO and then
228         (305)
229         pears
230         (260)
231         are scatter┌ing.┐
232   NS4              └  uh┘huh ┌huh  ┐ huh.
233   KEIKO                      └yeah.┘
234         (260)
235         an(d)
236         (220)
237         there are three children near the place,
238   NS4   mmhm,
239   KEIKO an(d)
240         (275)
241         they try to help him to stand up ┌an(d)┐
242   NS4                                  └   mm┘hm,
243   KEIKO to pick up pears,
244   NS4   ah uhhuh huh ┌huh. ┐
245   KEIKO              └yeah.┘
246         (135)
247         and then oh okay thank you an(d)
248         (533)
249         the-
250         (335)
251         boy just gone?
252         (100)
253   NS4   mmhm,
254         (120)
255   KEIKO an(d) three children,
256         (155)
257   NS4   mmhm,
258         (265)
259   KEIKO i think they also
260         (535)
261         steal.
262         (475)
263   NS4   ┌uhhuh.┐
264   KEIKO └the   ┘ pears ┌for each,┐
265   NS4                  └okay,    ┘
266         (280)
267   KEIKO an(d) they just walked and <eating> [//] <eat-> [/] eat
pears?
268   NS4   right.
269         (305)
270   KEIKO and then at the moment,
271         (344)
272         the guy,
273         (205)
274         u:h picking up pears on the ┌tree,┐
275   NS4                               └   mm┘hm,
276         (175)
277   KEIKO he realized that oh!
278         (355)
279         i lost
280         (325)
281         one of basket.
282         (603)
283         yeah. oh where is it.
284   NS4   yeah.
```

(168)
KEIKO and he counted.
(145)
one two &th- oh i think i had three,
(111)
NS4 ┌mmhm, ┐
KEIKO └but uh┘ two. just two he┌re.┐
NS4 └ ri┘ght.
(380)
KEIKO a:n(d)
(570)
<the:
(422)
children,> [///]
(185)
to the ┌three┐ children,
NS4 └mmhm.┘
(100)
mmhm,
KEIKO just came,
(475)
an(d)
(290)
they
(205)
were eating pe┌ars. ┐
NS4 └[=! audible inhalation┘ ┌ oh┐ no yeah.
KEIKO └yeah.┘
and the guy oh!
(850)
&=laughs pears! but
(200)
he just wondered.
(260)
he didn't talk to them,
NS4 oh ok┌ay. ┐
KEIKO └an(d)┘ just they passed.
(290)
┌him. ┐
NS4 └okay.┘
(316)
KEIKO an(d) he <just watched them.> [///]
(160)
just stared at their back,
(125)
NS4 mm┌hm,┐
KEIKO └but┘ they didn't
(500)
realize it an(d)
(265)
the man didn't talk to them ┌or ┐ get angry or ┌nothing┐ so
NS4 └mmhm,┘ └yeah. ┘
(505)
KEIKO that's the end of the movie.
(392)
NS4 wow.

Appendix J Complication segment of Asami Pear Story Monologue 1

074 ASAMI (435)
075 <and> [///]
076 (285)
077 and `then, [108 bpm]
078 ``(919)
079 a `boy, [///]
080 (110)
081 a `little `boy,
082 (170)
083 `maybe `he `is
084 `(670)
085 `eight or `ni:ne,
086 (1273)
087 a:nd <he:> [/]
088 (1145)
089 he
090 (220)
091 ride,
092 (375)
093 a bicycle.
094 (680)
095 actually,
096 (275)
097 the bicycle
098 (155)
099 is bigger than
100 (575)
101 the boy?
102 (570)
103 yeah.
104 (426)
105 an(d)
106 (185)
107 he found
108 (225)
109 the pears: bag,
110 (466)
111 yeah.
112 (293)
113 `actually [105 bpm]
114 (278)
115 the `man `who
116 `(477)
117 are `taking
118 `(403)

`pears `during the `movie
`(433)
<`is
`(770)
u:m`
`(165)
`three `bags> [//]
(466)
uh-
(598)
the `man `has [115 bpm]
`(337)
`three `bags
`(307)
`to:
```(1249)
yeah. put in
(240)
&θə- [///]
(360)
many pears,
(367)
yeah. so there a:re three bags.
(514)
and
(210)
the-
(488)
boy, [///] a little boy,
(350)
fin(d)- [//]
(618)
found,
(1112)
the:
(649)
bags,
(1388)
and,
(794)
he:
(190)
robbed.
(634)
the one of
(513)
the bags.
(644) <- sharp inhalation
but,
(606)
`yeah. [115 bpm]
`(500)
the `bicy`cle `is `bigger `than
(130)
the `boy,
(330)
and [=! laughing]
(385)
the bag of pears
(940)
is so heavy,
```

(825)
so
(414)
<it's:
(306)
looks hard to> [/]
(774)
um
(931)
it's looks hard to:
(1197)
um carry.
(1317)
yeah.
(218)
but
(395)
he robbed it,
(640)
and
(190)
then
(669)
<the &mæ-> [//]
(230)
the boy,
(1073)
&=laughs
ɾaidu@u [: ride]
(970)
a bicycle,

Appendix K Complication segment of Asami Pear Story Monologue 2

ASAMI (775)
an(d)
(150)
next
(570)
a `small `boy, [112 bpm]
`(700)
`showed `up, `with
`(588)
`riding `bicycle,
`(540)
the `bicycle `is very
(255)
big.
(495)
`as `big `as the `boy. [102 bpm]
(910)
a:nd
(1794)
<the-> [//] a fat man
(780)
`is
(235)
`gathering `pears `on the `ladder, [102 bpm]
(296)
`so <`he can`no^(570)t
(718)
watch> [//]
(706)
he cannot look
(379)
back.
(600)
of him,
(792)
a:nd uh
(988)
a small boy,
(1363)
u:h &fain^(741)dʊ: [: find]
(314)
a:
(348)
some box of <pears,> [//]
(394)

gathered pears,
(569)
`and `he: [102 bpm]
(185)
`robbed
`(649)
`a: `box
`(285)
`of
``(991)
`pears.
(381)
but a `fat `man, [115 bpm]
`(550)
`didn't `notice `that.
(155)
be`cause `he
`(430)
can`not
(791)
look.
(749)
and,
(520)
u:h
(348)
a small boy,
(782)
left
(393)
there.

Appendix L Complication segment of Asami Pear Story Dialogue 1

```
060          (535)
061    ASAMI and then,
                  * <- A looks away
062          (345)
063          uh: <a: little> [//]
                    * <- A resumes eye contact
064          (400)
065          a `small `boy, [120 bpm]
                 * <- A looks away
                 * <- hands palms down move down
                       * <- left hand wave to left side
066          (100)
067          `with `bicycle,
                    * <- A resumes eye contact
                    * < - gestures bicycle handles toward left
068          `(660)
              1====
069          `showed `up.
               =====+
070          ``(1130)
                 2==+
071          `æn`dʊ@u [: and]
               * <- looks away
072          (230)
073          the `boy `is
074          `(595)
075          `bad.
               * <- resumes eye contact
076          `(530)
077          bad `boy,
              * <- Asami nods type 2 leaning forward
078          (210)
             1===
079          be`cause,
             ======+
080          `(350)
081          um
             * <- looks away to left
082          `(360)
083          `duɾiŋ`gu@u [: during] [130 bpm]
084          `(395)
085          the `fat `man;
                   * <- right hand up in prep for next gesture
086          (465)
```

```
087        `gæða`dʊ@u [: gathered];
            * <- resumes eye contact
            * <- right hand gestures picking pears on 2 beats
088        (265)
089        `pea`zu@u [: pears],
090        `(705)
            2====
091        `he
            ==+
092        `(620)
093        `kʊdɪn`tu@u [: couldn't];
             * <- looks away to left
094        (300)
095        <`look-> [//]
               * <- gesture pointing down and behind her
               * <- regains eye contact
096        (270)
097        `watch
             * <- same a previous gestures repeated twice
098        (520)
099        `uh
            * <- gesture with both hands pointing down and behind
            * <- looks briefly down behind her right side
100        (385)
101        °nan@s:jpn darou@s:jpn.°
102        (730)
103        back of uh
            * <- NS leans forward and turns head
                 moving left ear closer
104        (450)
105        &hɪ- [//] him?
                       * <- quick single head nod
           3==+ <- NS as if to show comprehension or “ah!”
106        (265)
107   NS1   mmhm,
           2===+
108        (220)
109   ASAMI mm mm mm so,
            * <- A nods type 2
                       * <- looks away to left
                  2===+
110        (720)
111        <the-> [//]
112        (260)
113        `he: [120 bpm]
114        `(670)
115        `kʊdɪn`tu@u [: couldn't];
116        (240)
117        `noticed` the `boy,
             * <- resumes eye contact
                             * <- Asami head nod
118        (260)
119   NS1   mm`hm
            1===+
120        `(505)
121   ASAMI so
             * <- A looks away
122        (775)
123        <he-> [//] <the boy,> [//]
124        (305)
```

```
125        the small boy,
                * <- resumes eye contact,
                * <- gestures pointing downward
126        (908)
            * <- looks away
            * <- gestures taking something
127        °nan@s:jpn darou@s:jpn.° robbed.
                                      * <- resumes eye contact
128        (215)
129        the
130        (465)
131        `box `of `gathered [=! laughing] `pear, [100 bpm]
            * <- gestures big circle as shape of box
                      * <- open left hand palm in beat down
               same gesture one more beat -> *
132        (290)
           2==== <- with smile & wince
133   NS1   mm┌hm, ┐
           =======
134   ASAMI   └&=la┘ughs
                   =====
135        (620)
            ===+
136        yes.
137        (572)
138        an(d)
            * <- looks away
139        (350)
140        <the bicycle is so big-> [//]
141        (300)
142        big tte@s:jpn iu@s:jpn ka@s:jpn ano@s:jpn
               1==========================+
143        (340)
144        the `bicy`cle `is `as `big `æ`zʊ:@u [:as] [180 bpm]
                * <- vague gesture, 2 hands in front, palms inward
                * <- NS suppressed yawn
145        `(360)
146        `boy,
             * <- resumes eye contact
147        `(305)
            2====
148   NS1   mmhm,
           =====
149        (170)
            ====+
150   ASAMI so <it-> [//]
                * <- looks away
151        (510)
152        °nan@s:jpn darou@s:jpn° `it `ʃɪmu`zʊ@u [: seems]
                                                [130 bpm]
153        (175)
154        `diffi`cult `to `carry,
             * <- resumes eye contact
                            * <- gestures bicycle handles steering
155        `(565)
156   NS1   `right ┌`yeah.┐
157   ASAMI        └`mm   ┘ mm mm.
                   * <- Asami rapid head nods
            2=====================
158        `(430)
             ====+
```

159 `but <`she> [//]
* <- looks away
* <- hands move to left side prep for taking, remain there
160 (195)
161 uh <he-> [//]
* <- resumes eye contact quickly
162 (290)
163 <he-> [//] the `boy `robbed [95 bpm]
3======+ <- NS quick head raise, preceded with lean forward
* <- gesture taking something
164 (125)
165 `the
166 (295)
167 `big box
* <- gesture 2 hands outline big box shape
168 `(360)
2====
169 `of pea┌`zu@u [: pears],┐
=+ * <- gestures 2 hands holding box
170 NS1 └ mm┘hm,
1===+
171 (355)
172 ASAMI so he
* <- looks away
* <- begins long gesture pick up box
173 (635)
174 bring
* <- resumes eye contact
* <- completes gesture put on bicycle
175 (1285)
176 `pears `box `on [110 bpm]
* <- gestures bicycle handle and hold through
177 `(685)
178 `bicycle,
* <- gesture bicycle handle downbeat
179 `(760)
5==== <- intense then multiple small nods
180 NS1 ┌mmhm,┐
181 ASAMI └`and ┘ `he `ride `on `the `bicy`cle, [150 bpm]
* <- RH short wave to right
* <- return to bicycle handles
* <- gestures steering bicycle handles
=====+
182 `(760)
* <- hands move to tabletop
183 `and `go.
* <- left index finger draws line on table
184 &=laughs
185 yeah [=! laughing]
2=== <- smile & wince
186 (275)
=====
187 yeah ┌it's dif┐ficult to say but
188 NS1 └okay. ┘
========+
189 (103)
190 ASAMI mm.
* <- looks down at table

Appendix M Complication segment of Asami Pear Story Dialogue 2

```
061          (229)
062    ASAMI and then,
              * <- looks away
063          (271)
064          a `small `boy, [116 bpm]
                   * <- resumes eye contact
                          4===
065          `(315)
              =====
066          showed `up,
             ==========
067    NS2    mm`hm,
              =====+
068    ASAMI with `bicycle.
                        * <- gestures bicycle handles
069    NS2    (203)
             2====
070          o`kay,
             =====
071          (103)
             =====
072    ASAMI the `bicycle `is as `big `as the `boy, [94 bpm]
               * <- looks away     * <- resumes eye contact
                                   * <- 2 hands beat down together
                      2 hands wave to left -> *
              ==============+                  4==
073    NS2    (310)
              =====
074          `mm!
              ===
075          (110)
             =====
076    ASAMI `so <it `seems
               * <- looks away to left
                        * <- left hand palm down downward to left
              ============+
077          `(530)
078          uh` to-> [//]
079          (185)
```

```
080        `it seems `hard to `carry,
           * <- resumes eye contact
               * <- 2 hands palms down downward beat
                              * <- gestures bicycle handles
                        5============
081        (110)
           =====
082   NS2   ┌&=laughs   ┐
            ============+ <- nod intensifies during laughter
083   ASAMI └   &=laughs┘
084        (320)
085        yeah
086        (175)
087   ASAMI but `actually the `boy is `bad boy, [90 bpm]
                                * <- left hand point toward table
                                        * <- repeat same gesture
088        (271)
           3====
089        be┌`cause,┐
090   NS2     └`oh!   ┘
           ======+
091        (260)
092   ASAMI `he: [//]
              * <- looks away to left
093        ``(773)
094        he `robbed the `box of `gathered `pears,
           * <- left hand point down towards table
               * <- resumes eye contact, gesture 2 hands
                    pull away to right side
                           * <- gestures outline of box
                      3===================+   3===
095   NS2   (250)
            =====
096        o`kay,
           =====
097   ASAMI but a `fat man [90 bpm]
                    * <- gestures 2 hands down to right side
             * <- looks up to left
            ====+
098        `(443)
099        `couldn't
100        `(622)
             * <- resumes eye contact
            2==
101        `notice.
            =======
102        (99)
           ====
103   NS2   m┌`m:.  ┐((.pt))
104   ASAMI  └`mm mm┘ mm mm ┌ mm.┐ <- Asami rapid nods type 2
105   NS2    *              └`i  ┘ see. <- NS wince
            ==========================
106        `(325)
            =====
107   ASAMI `and he `ride `bicycle `and leave `there,
             =+          * <- gesture bicycle handles
                           left hand points -> *
                            to left side
108   NS2   (215)
```

109   oh ┌`no. ┐
110   ASAMI   └`mm mm┘ mm.
* <- Asami nods type 2
111   NS2 (224)
112   huh.
113   ASAMI yeah.

Appendix N Complication segment of Keiko Pear Story Monologue 1

048 (1540)
049 KEIKO and `then,
050 (190)
051 a `boy is <`coming. [92 bpm]
052 `(664)
053 on> [//] `uh `coming.
054 (353)
055 `riding on a `bicycle?
056 `(559)
057 an(d)
058 `(680)
059 <`he:>[/]
060 `(748)
061 he just `stop in `front of `tree, [107 bpm]
062 `(550)
063 an(d)`
064 `(593)
065 he `pick
066 `(405)
067 a `basket.
068 `(535)
069 <`full of a-> [//] uh full of pears,
070 (585)
071 an(d)
072 (839)
073 uh <he try> [//]
074 (429)
075 he `tries to `carry the `basket. [90 bpm]
076 `(761)
077 on the `bicycle.
078 `(645)
079 but the `man didn't `realize it.
080 `(605)
081 `an(d)
082 `(245)
083 `he (i)s just `riding
084 `(416)
085 on the `bicycle,

Appendix O Complication segment of Keiko Pear Story Monologue 2

```
086          (1406)
087    KEIKO an:(d) then,
088          (358)
089          a boy,
090          (610)
091          uh riding on
092          (107)
093          a `bicycle is `coming. [102 bpm]
094          `(778)
095          `an(d) `stop in `front of the `tree.
096          `(849)
097          the `man `picking up `pears `didn't `realize it.
098          (737)
099          an(d)
100          (180)
101          the boy,
102          (407)
103          u:h just put
104          (407)
105          his bicycle on the ground,
106          (862)
107          a:n(d)
108          (683)
109          u:m
110          (350)
111          an(d)
112          (430)
113          pick
114          (626)
115          a `basket `full of `pears, [86 bpm]
116          `(969)
117          `a:nd `then, [90 bpm]
118          (407)
119          `u:h <`put> [/]
120          (510)
121          `put it
122          (186)
123          `on his `bicycle.
124          `(837)
125          `an(d) jus(t)
126          `(154)
127          he has `gone,
128          (470)
129          but <the man &p-> [//] u:m the `man,
130          `(459)
```

131 on the `tree,
132 `(447)
133 didn't `realize it.

Appendix P Complication segment of Keiko Pear Story Dialogue 1

```
085         (485)
086   KEIKO a:n(d)
             * <- K looks down & right
087         (725)
088         and `then
089         (95)
090         a `boy, [90 bpm]
           * <- K resumes eye contact
091         (255)
             1==
092   NS3    mm`hm,
             =====
093         (65)
             ===+
094   KEIKO is `coming <on the:> [//]
                 * <- gestures right hand move from
                           right side pointing
                      inward to meet left hand in front
                          * <- K looks away briefly
095         ``(205)
096         on a `bicycle.
                  * <- 2 hands together beat down
097         `(340)
              2== <- nod preceding by eyes up & left
098   NS3    mm`hm,
             =====
099         (163)
           =====
100   KEIKO the `bicycle is `really `big, [100 bpm]
                 * <- gesture 2 hand palms inward
                                     * <- hands spread outlining
                                          a big round shape
             ==+
101         (215)
102   NS3    &=laughs
103   KEIKO `bigger than `him,
              * <- 2 hand beats on syllables, move right on "him"
                          2=== <- nod with smile
104         (140)
           =====
105   NS3    mm`hm,
             ====+
106   KEIKO `and `he:` just `put the
                   * <- K looks up & left
            * <- hands return to front of chest together
```

```
107         (185)
108         `bicycle on the `ground,
             * <- K resumes eye contact
             * <- 2 hands palms inward beat
                            * <- 2 hands downward her right side
                    1==========+
109         (235)
110   NS3   mm┌`hm┐
111   KEIKO   └  a┘nd
           2====+
112         `(555)
113         maybe `stole? [90 bpm]
114         (220)
115         a `basket?
               * <- gestures 2 hands holding basket
116         `(320)
117         full of `pears?
                    * <- same gesture basket, beat downwards
                    * <- NS pulls head back and sidelong glance
118   NS3   oh!
119   KEIKO a `basket `yeah.
              * <- another beat with same gesture
120   NS3   &=laughs
121         (100)
122   KEIKO an(d) `put
             * <- K looks down & left, lifts basket hands up
123         (170)
124         the `basket on the-
125         `(365)
126         bicycle?
            * <- K resumes eye contact, gestures basket down
127         `(534)
128   NS3   ┌`ah!          ┐
129   KEIKO └and then just┘ gone,
                             * <- both hand flick left
130         (723)
131   KEIKO but the `man, who was` picking up` pears, [90 bpm]
                     * <- right hand points to right
                                  * <- 2 hands rolling
                                       around each
                                       other then palms up
                                     holding pears on "pears"
                          1=====+          2=====
132   NS3   uh`huh,
            =====+
133   KEIKO he didn't `realize it.
            * <- hands together in front
134         (230)
135   NS3   `oh!
```

Appendix Q Complication segment of Keiko Pear Story Dialogue 2

```
099          (556)
100   KEIKO and `then, [82 bpm]
101          (110)
102         a `boy,
               * <- right hand palm up out to right side
103          (220)
104   NS4   mm[`hm,]
105   KEIKO   [ uh ]  riding on the `bicycle?
                                    * <- 2 hands together
                                        move down palms downward
             1===+
106          (205)
             * <- NS head rise, eyebrows raised
107         is `coming.
                * <- 2 hands move from right side inward to chest
108          (165)
            3====
109   NS4    `uhhuh.
             ====+
110          (95)
111   KEIKO an(d)  `stop in `front of the `tree, [118 bpm]
                    * <- 2 hands down beat, roll around each other,
                         palms inward beat down on "tree"
112          (200)
            1====
113   NS4    `mmhm.
             ====+
114   KEIKO an(d) he: just `put his `bicycle on the `ground.
                                                  [95 bpm]
             * <- hands together      * <- 2 fists move downwards
                                        hands down -> *
                                        right side
                                                     4===
115   NS4    `mmhm mmhm.
             ========+
116          (140)

117   KEIKO `an(d)
             * <- Keiko looks to her right as if thinking
             * <- hands together
118         `(570)
119         i `think that `he:
120         `(440)
121         jus(t)
122          (295)
```

```
123        `try to `steal, [90 bpm]
                  *   <- Keiko resumes eye contact
124        (315)
          1====
125   NS4   ┌mm`hm.┐
126   KEIKO └a     ┘ `basket `of
                       * <- hands open palms up, beat down
            ====+
127        `(441)
128        the pears.
            * <- hands roll and palms up on "pears"
129        (70)
130   NS4   `oh!
             * <- NS looks up & left

131   KEIKO yeah `there are some `basket in `there, [100 BPM]
                  * <- hands together, eyes downward
                                 * <- hands open palms down,
                                      resume eye contact
                                           1=====
132   NS4   `mmhm,
             ===+
133        (110)
134   KEIKO uh `full of `pears ┌ an(d)┐ <- hands roll &
                                          palms up on "pears"
135   NS4      *              └`mmhm.┘
                                1===+   * <- K looks up & right,
                                             hands together
136        (415)
137   KEIKO `he `tried to `pick up `one of them.
                         * <- K resumes eye contact
               hands palms in grab -> *
               a basket and lift up
138        `(625)
            1==+
139        `an(d)
140        (357)
141        `put [118 bpm]
            * <- gesture continues beat in front of face
142        (245)
143        `on his `bicycle
                   * <- gesture continues down & left
                      3== <- raises head
144   NS4   `uhhuh.
             ====+
145   KEIKO `an(d) jus(t) `gone.
                         * <- 2 fists push out to left
146         (65)
147   NS4   `wow.
148   KEIKO `yeah and the `man uh `picking up `pears ┌he `did┐n't
`realize it.
                         * <- right hand points out to right
                      right hand palm up -> *
                right hand rapid shake on "didn't" -> *
149   NS4                                      └mm`hm, ┘
                                                1====+
150        (140)
151        `oh o┌`kay. ┐
             * <- NS leans back
152   KEIKO     └yeah. ┘
153        (485)
```

```
154   KEIKO an(d)
             * <- K looks left
155         (235)
156         i thought at firs(t)
157         (445)
158         u:h `he [120 bpm]
159         (295)
160         `tried to `help him.
                       * <- K resumes eye contact
161         `(655)
             3====+ <- opens mouth as if "ah", pulls head back
162         ┌to `car┐ry the ┌basket but i think yeah.┐ <- rushed
163   NS4   └uh`huh.┘       └oh i see i see.        ┘ <- rushed
             2====+         * <- NS looks upwards briefly
164         (320)
165   KEIKO `just he `try to [90 bpm]
166         (320)
167         `steal it.
168         (85)
169   NS4   oh.
170   KEIKO `yeah.
             2===
171         (164)
             ====+
```