

ACOUSTIC AND PERCEPTUAL CORRELATES OF L2 FLUENCY:
THE ROLE OF PROLONGATIONS

By
Jessica Liane Rohr

DISSERTATION
Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics at
The University of Texas at Arlington
August, 2016
Arlington, TX

Supervising Committee:
Cynthia Kilpatrick, Supervising Professor
Laurel Stvan
Naoko Witzel

ABSTRACT

Acoustic and Perceptual Correlates of L2 Fluency:

The Role of Prolongations

Jessica L. Rohr, Ph.D.

The University of Texas at Arlington, 2016

Supervising Professor: Cynthia Kilpatrick

This dissertation examines the contribution of lengthened segments/syllables or *prolongations* to the utterance and perceived fluency of native and non-native speech via three related studies. First, utterance fluency was measured using speech samples collected from a narrative task and along nine temporal variables of fluency (mean length of run, speech rate, mean length of syllable, mean length of silent pause, and rates of silent pause, filled pause, self-corrections, repetitions, and prolongations) (Skehan, 2007; Bosker, 2014). Results showed native speakers spoke significantly faster and with fewer disfluencies than their non-native counterparts.

Prolongations were further investigated for frequency, duration, syntactic and phonological distribution in both native and non-native speech. This study showed non-native speech contained more prolonged syllables per speech sample (calculated as a ratio to total syllables) than non-native speech. Native speakers were more likely to prolong syllable-final segments of function words at clausal boundaries. Non-native speakers, by contrast, were less likely to prolong at clausal boundaries, and were more likely to prolong content and function

words equally, factors which point to the learner's need for additional planning time and which likely contribute to impressionistic differences in native and non-native prolongation use.

Finally, a perceptual experiment utilized speech samples manipulated for clause position (initial, non-initial) and frequency (high, low) showed that naïve native raters were able to judge the relative fluency of both native and non-native speech (contra Riggenbach, 1991; Davies, 2003). Results also showed non-native speakers were rated less fluent regardless of prolongation placement or frequency, but raters were sensitive to the clause position condition in native speech. These findings suggest native-like placement is more salient than frequency in determining overall fluency. Moreover, they show raters are sensitive to the misuse of prolongations even when all other disfluencies are removed, indicating that prolongations deserve more individual attention in fluency research.

Copyright by
Jessica Liane Rohr
2016

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, committee chair, colleague, co-researcher, mentor, and friend, Dr. Cindy Kilpatrick for her unfailing support and leadership throughout my time at the University of Texas at Arlington. She has been my cheerleader and advocate and has constantly gone above and beyond to make sure this particular document made its way into the world. Suffice it to say that none of what has been accomplished during these years would have been possible without her.

I would like to thank my committee members Dr. Laurel Stvan and Dr. Naoko Witzel for their support and patience through different drafts of this dissertation. Their insightful comments have made the finished product so much richer than it would have been without the benefit of their expertise.

The financial support I received from the Department of Linguistics and TESOL and the College of Liberal Arts in the form of EGTA-ships, Summer Dissertation Fellowships, research and travel grants have ultimately made this dissertation possible. There are no words to describe my appreciation for their understanding in difficult circumstances, and for fostering an environment in which student success is the communal aim.

To all those colleagues in the field who have provided valuable feedback on this project at its various stages, and whose insights often directed me toward new paths of inquiry: Dr. Ann Wennerstrom, Dr. Sylviane Granger, and countless others at various conferences throughout the years.

And for Lori McLain Pierce and Iya Khelm Price, for their encouragement and support (and occasional commiseration) on matters linguistic and non-linguistic alike. Without their friendship these past five years would not have been the rich, rewarding time they have been.

DEDICATION

This is for my parents, who never ever not even once stopped believing this was possible—even when I told them in no uncertain terms it wasn't! For my siblings and nephew, who kept me sane by filling my life with laughter and noise and fun. And for my grandmother, to whom this dissertation belongs almost as much as it belongs to me, for putting in nearly as many hours of prayer as I put in hours of research and writing.

Finally, this is also for Carrie, who has shown me what a godly friendship looks like by living it every single day.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	v
DEDICATION.....	vii
TABLE OF CONTENTS.....	viii
LIST OF TABLES.....	x
LIST OF FIGURES.....	xi
CHAPTER ONE: INTRODUCTION.....	2
The Research Gap.....	6
Reading Guide.....	8
CHAPTER TWO: LITERATURE REVIEW.....	10
Defining key terms.....	11
What does it mean to be <i>fluent</i> ?.....	13
How is fluency measured?.....	16
A Fluency Framework.....	22
Relating Utterance Fluency to Perceived Fluency.....	23
Summary and motivation.....	28
Conclusion.....	29
CHAPTER THREE: CORPUS COLLECTION AND ANALYSIS.....	31
Research Question.....	32
Learner corpus typology: An overview of the field.....	34
Pilot Corpus.....	43
Full corpus collection and analysis.....	57
Discussion.....	68
CHAPTER FOUR: SURVEY OF PROLONGATIONS.....	70
Research Questions.....	70
Literature Review.....	71
Methods.....	78
Results.....	81
Discussion.....	91
Conclusion.....	93

CHAPTER FIVE: PROLONGATIONS AND THE PERCEPTION OF L1 and L2 FLUENCY	94
Background	94
Research Questions	97
Predictions	98
Methods.....	100
Analysis	107
Results.....	108
Discussion.....	111
CHAPTER SIX: CONCLUSIONS AND APPLICATIONS	116
Summary of Results	116
Applications.....	119
Directions for future research	125
Conclusion.....	126
REFERENCES.....	128

LIST OF TABLES

Table 1	Acoustic measures dimension of fluency.....	21
Table 2	Sample of common symbols in Jefferson Transcription System.....	36
Table 3	Comparison of NS and NNS disfluency measures in narrative task.....	51
Table 4	Comparison of NS and NNS disfluency measures in free-speech task.....	52
Table 5	Comparison of NS and NNS disfluency measures in read-aloud task.....	54
Table 6	Number of non-native participants by L1 Background.....	59
Table 7	Global means and effects of trial on fluency measures.....	65
Table 8	Means and effects of trial on temporal variables by language background.....	65
Table 9	Global means and effects of tell vs. re-tell on fluency measures.....	66
Table 10	Means and effects of tell vs. re-tell on fluency measures by language background...	66
Table 11	Means and effects of language background on fluency measures.....	68
Table 12	Data overview for prolongation survey.....	80
Table 13	Percentage of prolongations within NS/NNS word tokens.....	82
Table 14	Confidence intervals for prolongation duration.....	83
Table 15	Prolongations by position within syllable.....	86
Table 16	Prolongations by segment type and syllable position.....	87
Table 17	Total and percentage of frequent segments by syllable position.....	88
Table 18	Content vs. function tokens of prolongations.....	89
Table 19	Initial vs. non-initial tokens of prolongations.....	90
Table 20	Mean and standard deviation of z-scores across frequency and clause position in non-native speech samples.....	111

LIST OF FIGURES

Figure 1	Levels of learner proficiency by Common European Framework (2001).....	58
Figure 2	Screenshot of analysis in Praat.....	61
Figure 3	Spectrographic representation of average and prolonged ‘she’ (in milliseconds).....	79
Figure 4	Sample item by condition.....	103
Figure 5	Instructions for fluency rating.....	105

CHAPTER ONE

INTRODUCTION

Ask any student in a second language classroom and they will likely say that their ultimate goal in learning their new language is to become fluent. However, the question which immediately follows—*What does it mean to be fluent?*—has no straightforward answer, despite an ever-growing body of research which aims to answer this very question. More often than not, it is easier to approach the elusive functional definition of fluency indirectly; thus, the study of fluency is also the study of those disruptions in the speech stream which make a speaker less fluent.

Overview

Lennon (1990) separates fluency into *broad* fluency, which he defines as general competence in a second language and *narrow* fluency, which is a measure of the listener's impression of non-native speech and which he later built upon to encompass both linguistic and performance competence which is executed smoothly under the time constraints of online-processing (Lennon, 2000). The present study is concerned with aspects of narrow fluency.

From Lennon's definition, Segalowitz (2010) proposed that narrow fluency is a multimodal concept which includes the three interrelated notions of *cognitive*, *utterance*, and *perceived* fluency. Cognitive fluency is defined as the relative efficiency of speech planning and production and its associated processes. Utterance fluency is the acoustic realization of fluent speech and the measurements which operationalize this notion of fluency. And finally, perceived fluency is defined as the judgments which listeners make about the speaker's cognitive fluency based on their utterance fluency. As the very definition of perceived fluency makes clear, these

three categories of fluency are very closely linked. Indeed, fluency is typically studied by relating one notion of fluency to another.

A rich body of research has examined second language fluency by taking acoustic measures of utterance fluency and correlating these measures to ratings of perceived fluency (Lennon, 1990; Riggensbach, 1991; Cucchiarini, Strik & Boves, 2000, 2002; Derwing et al., 2004; Kormos & Denes, 2004; Rossiter, 2009; Kang, 2010; Kang, Rubin & Pickering, 2010; Kang & Pickering, 2012; Bosker et al., 2013). In addition, studies have compared the speech of fluent and non-fluent speakers (Ezjenberg, 2000; Riazantseva, 2001; Tavakoli, 2011) and the development of fluency over time in a series of longitudinal studies (Freed, 1995, 2000; Towell et al., 1996). These and other similar studies make use of acoustic measurements related to fluent speech, generally broken down along the dimensions of fluency proposed by Tavakoli & Skehan (2005) and operationalized by Skehan (2003, 2009); namely: *breakdown fluency* (the number of silent and filled pauses and pause duration), *repair fluency* (the number of self-corrections and repetitions), and *speed fluency* (the mean length of run, speech rate, articulation rate, etc.).

The results for studies which focused on aspects of speed fluency tell a somewhat cohesive story: speech rate and mean length of run are strongly associated with ratings of L2 fluency (Lennon, 1990; Towell et al., 1996; Kormos & Denes, 2004; Segalowitz & Freed, 2004). The case for breakdown and repair fluency, however, is not as clear. Negative correlations with fluency have been shown to exist for pause frequency and length (Bosker et al., 2013), pause length but *not* pause frequency (Cucchiarini et al., 2000), and pause frequency but *not* pause length (Kormos & Denes, 2004). Similar conflicting results are found in the study of pause distribution, in which small sample sizes and slight differences in methodology and pause measures have meant results that are frequently contradictory (Lennon, 1990; Towell et al.,

1996; Pawley & Syder, 2000, Riazantseva, 2001, among others). For example, studies in pause distribution have repeatedly shown that non-native speakers are more likely to have pauses within clauses than native speakers (Riggenbach, 1991; Freed, 1995). However, with a larger number of speakers in each group (L1=20; L2=30), Riazantseva (2001) found the reverse: neither group was statistically more likely than the other to produce pauses within clauses rather than at boundaries.

Lastly, the case for repair fluency in general and prolongation in particular is even less consistent. Despite a few brief mentions in studies of other hesitation phenomena (Collard, 2009; Vural, 2008; Khang, 2014), prolongations are conspicuously understudied in the disfluency canon, being termed by Eklund (2001) as “the dark horse of the disfluency stable”. So why have prolongations been so long overlooked?

The challenge of studying prolongations

Part of the initial challenge with studying prolongations is categorical: what exactly constitutes of *prolongation* in the disfluency sense of the term? As the name itself suggests, a prolongation is simply a lengthened (or *prolonged*) syllable, but syllables and segments can be lengthened for a number of reasons. Syllables are regularly lengthened as a result of phonological processes: stressed syllables are longer than unstressed syllables (Oller, 1973), syllables are longer at the end of an utterance (Gaitenby, 1965; Klatt, 1976, and segments may be longer in certain syllable positions as opposed to others (Klatt, 1976) (for a full review of lengthening and durational questions in phonology, see Savithri, (1984)). Additionally, segments or syllables can be intentionally lengthened for pragmatic purposes: to signal a dispreferred second in an adjacency pair (Rose, 2008), to mark emphasis (Ladd, 1996; Kohler, 2006), to indicate turn-taking intent (Du Bois et al., 1993), and to provide a signal to the listener

to suspend conversational implicature (Rose, 2008). Such prolongations—which here are termed *functional* prolongations—occur across languages to varying degrees, and must be carefully identified and set aside when examining disfluent prolongations.

The overlap of functional and disfluent also causes challenges in providing an acoustic account for duration in prolongations. Though the listener may impressionistically know when something is longer than average, acoustic accounts of prolongation must take into account not only variability between speakers (and their average length of syllable) but also the phonological weight of the syllable, the stress on it, and its position within the intonational contour. Perhaps because of these difficulties, studies of prolongations tend to shy away from making durational claims. In fact, speakers themselves seem to recognize this: disfluent prolongations appear to be significantly and distinctly longer than functional prolongations, making them easier to recognize and perhaps signaling the categorical distinction between them. Still, at this time this difference is impressionistic only and much work remains to be done to define what, if any, are the differences in acoustic realizations between functional and disfluent prolongations.

Despite these inherent difficulties, a thorough investigation of prolongations is needed and has a wide range of applications for foreign language teaching and testing as well as speech perception. For instance, strategic and pragmatic competence are considered integral components of communicative competence in a second language (Canale & Swain, 1980). Understanding how and why syllables are functionally prolonged can increase language learners' pragmatic competence; being able to employ prolongations correctly to buy additional time for speech planning (without causing communication difficulties or intelligibility issues for the listener) can increase the language learner's strategic competence. However, to adequately teach

prolongations, we must first have a thorough knowledge of how they are used and perceived by native and non-native speakers.

A better understanding of prolongations is also needed for studies in speech perception, which seek to establish how disfluencies affect the listener—either causing processing delays which hinder intelligibility, or creating a helpful scaffolding to facilitate the intake of new or complex information (Watanabe et al., 2008; Bosker, 2014). An ongoing body of research has shown the effects of silent and filled pauses to speech processing, generally showing that disfluencies are generally helpful in NS-NS interactions and harmful in NS-NNS interactions. As a member of the disfluency class, prolongations should also be studied in this manner, thus providing two vital pieces of information which are currently missing from the disfluency discussion—first, whether listeners have similar expectations for the patterning of filled pauses and prolongations, disfluency types which are frequently classed together despite the lack of empirical motivation for doing so; and second, whether prolongations can also trigger attentional effects in the same way as their counterparts in the disfluency canon.

Both teaching prolongations effectively and the impact of prolongations on speech perception and processing are far-reaching questions. To answer them we must first begin by building upon our understanding of what characterizes disfluent prolongations, how they appear in the speech of native and non-native speakers, and how they are perceived by native listeners.

The Research Gap

There is much to be gained by studying how fluency can be reliably measured. However, in order to do that, there must first be a thorough account of each type of disfluency, both in distribution across native and non-native speakers and also in how deviations from native-likeness in non-native speech affect the perceived fluency of the speaker. These questions remain

unexamined in the case of prolongations. For this reason, this dissertation has as its main research aim, to address the following:

RQ 1: In what ways do measures of utterance fluency related to prolongations differ from native to non-native speakers?

RQ 2: To what extent do prolongations affect the perceived fluency of native and non-native speech?

To address these questions, this dissertation reports on three studies. The first, a corpus of learner speech, provides the basis for understanding the difference between native and non-native breakdown, repair, and speed fluency Skehan (2007). This study also broadens the scope for the acoustic representation of disfluency by including independent prolongation measures. Overall, this study confirms previous findings concerning the differences between native and non-native speech; namely, that native speakers speak faster and produce more language per utterance with fewer disfluencies than non-native speakers. The second study, a survey of the distribution of disfluent prolongations, examines the frequency, duration, phonological, and syntactic features of prolongations as realized in native and non-native speech and illustrates that non-native speakers produce more and more irregular prolongations than native speakers. . Finally, the third study analyzes fluency ratings of native and non-native speech samples which included prolongations manipulated for frequency (high, low) and clause position (initial, non-initial). This study indicates that raters in this study were sensitive to the placement of prolongations even when all other disfluencies were removed (filled pauses, hesitations, repetitions, self-corrections) or normalized (silent pauses). Moreover, this sensitivity applied only when raters were listening to native speech samples but not non-native samples.

The following section presents an overview of each chapter, including further details of the three aforementioned corollary studies of prolongations.

Reading Guide

In order to best address the main research question, this dissertation takes a multi-tiered approach.

Chapter 2 provides a detailed account of the major themes in the literature regarding fluency, including an examination of the prevailing fluency frameworks (Segalowitz, 2010; Skehan, 2007), common methods for measuring fluency, and classifications of disfluency.

Chapter 3 takes a first step in examining prolongations by first taking a look at overall measures of fluency. This chapter reports on the collection and analysis of a corpus of learner speech. First, different task types are examined in order to determine the most appropriate one to use in an examination of prolongation factors in utterance fluency. Second, this chapter addresses the question of how disfluency patterns differ in native and non-native speech by comparing native and non-native speech across nine measures of fluency (mean length of run, speech rate, mean length of silent pause, mean length of syllable, and rates of silent pauses, filled pauses, self-corrections, repetitions, and prolongations).

Chapter 4 focuses specifically on determining the characteristics and distribution of prolongations across native and non-native speech. Building on the findings in Chapter 3 regarding prolongations, Chapter 4 undertakes a detailed acoustic examination of the frequency, duration, and distribution of prolongations in native and non-native speech. Few studies have looked so extensively at prolongations; the few studies that do exist have focused primarily on patterns of prolongation in first language speech (Eklund, 2001; Den, 2003; Lee et al., 2004; Moniz, Mata & Viana, 2007). The findings in this chapter constitute a much-needed qualitative

look at how native and non-native prolongations differ in English, beyond the impressionistic differences first noted in Chapter 3.

Chapter 5 addresses the main research question of perceived fluency in order to determine whether prolongations play a role in how listeners judge fluency. This chapter describes an experiment which holds the prolongation variable constant across native and non-native speech in two conditions: clause position (initial, non-initial) and frequency (high, low). This is, to my knowledge, the first such study done exclusively with prolongations, following the model set out by Bosker (2014a, 2014b) which examined similar questions with regards to silent and filled pauses.

Chapter 6 summarizes and concludes the dissertation.

CHAPTER TWO

LITERATURE REVIEW

Natural or spontaneous speech is, by its very nature, disfluent (Bosker, 2014)— that is, filled with hesitations, silent pauses of various lengths, self-corrections, and vocalized pauses (*uh, um, er*, etc.), as well as any number of highly individualized hesitation markers (the recently popularized *like* is an excellent example of this). Most of these disruptions are easily filtered out in interactions between native speakers, and there is even evidence to suggest that those disfluencies which draw the attention may be useful indicators for upcoming low-frequency items (Watanabe et al., 2008; Bosker et al., 2013; Bosker, 2014a).

Between native and non-native speakers, disfluencies are largely thought to have negative effects on the native speaker's perception of L2 fluency (Kormos and Denes, 2004). In addition, recent work suggests that disfluencies also affect the online processing of non-native speech in ways that look quite different from native-native interactions (Watanabe et al., 2008). In order to understand disfluencies and their role in second language fluency research, it is necessary to examine the two broad categories of fluency research; namely a) the temporal nature of second language fluency, its development over time, and its perceptual effects on the hearer; and 2) the use of oral fluency as a metric for language planning and production for both the non-native speaker and the native listener.

Although findings in both sub-fields continue to broaden our understanding of how native speakers evaluate and process non-native speech, much work remains to be done to determine precisely which disfluencies trigger which effects.

This chapter examines the existing literature on disfluencies by laying out the terminology and theoretical parameters for the chapter, addressing different notions surrounding

second language fluency (including key definitions), and providing the framework for measuring fluency. The final major section lays out a fluency framework based on those given by Segalowitz (2010) and major studies as they pertain to different fluency types. This section also identifies the gaps in the literature which this dissertation addresses. The chapter is concluded with a brief section summarizing this chapter and presenting the directions for the remainder of the dissertation.

Defining key terms

At this point it is necessary to set limitations in terminology for the scope of this dissertation. First, much of the literature involved uses the terms suprasegmental and prosodic interchangeably, if not synonymously. These terms reflect two different theoretical foundations—the notion of the segmental and suprasegmental features of speech have their origins in a Structuralist framework, while prosody is a language-specific, organizational structure of speech put forth by the Autosegmental-Metrical framework (Ladd, 1996). While I recognize that which distinguishes these terms, I opt to use the terms prosody and prosodic for this dissertation, as suprasegmental phonology covers additional factors which are not necessarily dealt with here. However, if a particular study makes claims concerning suprasegmental factors and not prosodic factors, I defer to the original research for accuracy of reporting.

Due to this dissertation's focus on prosody, segmental information will be ignored and acoustic measures as used in this study will only and ever refer to acoustic measures of prosody: stress, intonation, pausing, and speech rate.

I also make a distinction between evaluative, cognitive, and acoustic measures of L2 speech. Evaluative measures are those based on the subjective perceptions of non-native speech

by native speakers, and are often measured by way of a rating scale, or a binary (good-bad) judgment. Cognitive measures vary in implementation, and can range from simple comprehension tasks, where NS listeners may be asked to transcribe a NNS utterance, but may also stretch to encompass varieties of psycholinguistic perceptual tasks. In the following overview of the literature, the cognitive tasks utilized will be explicitly detailed to avoid unnecessary confusion. Acoustic measures of L2 fluency are counts of individual disfluencies such as pauses and filled pauses as well as measures of temporal variables such as speech rate and mean length of run. Further details on acoustic measures are given in section 2.4.4.

Finally, because of the occasional lack of agreement in defining the four main measures of L2 attainment, the definitions I will be assuming throughout this dissertation are included here. These are based on an understanding of the research which forms the foundation for each respective field of study.

- *Fluency*—the ability of a non-native speaker to translate communicative intention into efficient speech in the target language with minimal disruption in the utterance
- *Comprehensibility*—the ease with which a native speaker is able to understand non-native speech
- *Accent* or *Accentedness*—the degree to which an L2 learner approximates (or fails to approximate) native pronunciation
- *Intelligibility*—the effective processing of a non-native speech signal by a native speaker of the target language

In certain instances studies which deal with measures of L2 attainment other than fluency will be used to highlight a particular point concerning the effects of prosodic variables on these

evaluative measures. As many of these notions overlap with one another, results may also be extrapolated wherever appropriate. Nevertheless, the stated goal of this dissertation remains to examine questions related to the idea of fluency, especially as it pertains to prolongations.

What does it mean to be *fluent*?

A frequently-used term in language teaching but also in the common vernacular, *fluency* often lacks a precise, technical, and widely agreed-upon definition.

In one of the generative definitions of the term, Fillmore (1979) highlighted what we might mean when we say a native speaker is *fluent*. There is a sense of fluency which is roughly equivalent to speed—someone who talks fast, or is able to fill long speech streams in a relatively short time might be described as fluent (i.e., radio personalities, sports announcers, etc.). Fluency may also mean the speech of someone who is able to speak coherently and clearly on complex topics (i.e., lecturers and politicians). A third and fourth sense Fillmore provides respectively ascribe *fluency* to those who exhibit great pragmatic (using language appropriate to the circumstances) or aesthetic (showing creativity with language through puns, metaphors, etc.) command of the language.

Broad Fluency

In second language environments, *fluency* may be used to denote general language proficiency. Lennon (1990, 2000) categorizes fluency as either *narrow* or *broad*. The broad sense of fluency encompasses what is generally meant when a non-native speaker might claim to be “fluent” in their second language, and as such may include all manner of lexical, grammatical, and pragmatic competence above and beyond simple phonological ability. In other words,

fluency in the broad sense may be thought of as a measure of global competence in a language, or overall speaking ability (Chambers, 1997).

This sense of fluency as a measure of global command of a second language creates a rough equivalency with “native-likeness” or the ability to produce native-like stretches of discourse (Pawley & Syder, 1983). This equivalency has been shown to be misleading and even problematic as, particularly for adult learners, there are various constraints to attaining full native-likeness (Krashen et al., 1979; DeKeyser, 2000, 2012) and even highly proficient L2 speakers may retain a degree of phonological accent (Granena & Long, 2013). Furthermore, there is ongoing lack of consensus among teachers and researchers alike as to what factors predict, produce, or describe second language fluency (Riggenbach, 1991; Freed, 1995).

With the growth of communicative language teaching (CLT) methods beginning in the 1980s, fluency has taken on new life in the L2 classroom as a complement to *accuracy*. Brumfit (1984) differentiates between accuracy-based activities, which focus on producing correct linguistic structures in the L2, and fluency-based activities, which target oral production of the L2 through meaningful interactions. Within the CLT classroom, the emphasis on authentic language use in a naturalistic setting helped to set *fluency* as a goal apart from native-likeness and instead as simply the “smooth, rapid, effortless” use of the L2 (Chambers 1997).

This particular definition of fluency suggests ties to the communicative competence model, and strategic competence in particular (Canale & Swain, 1980). In developing and using strategic competence, the L2 learner employs knowledge of the language beyond the strictly linguistic sphere, using everything they know about the L2 to craft the best response to a particular situation, environment, or language task. In this sense fluency encompasses not only the speed with which a second language speaker can use the language, but also their success in

achieving the communicative aim. This conception is formalized by Sajavaara (1987) as “the communicative acceptability of the speech act, or ‘communicative fit’” (pg. 62), and includes the ability to speak appropriately depending on the communicative context.

Narrow Fluency

Narrow fluency, with which this dissertation is primarily concerned, has been described in a variety of ways. Lennon (1990) conceptualized narrow fluency as a measure of a listener’s impression of non-native speech or, in other words, a performance on the part of the non-native speaker. Rehbein (1987) also defines fluency along these lines, claiming that in fluent speech the speaker can: a) plan and produce speech nearly simultaneously; and b) adapt to specific contexts and manage hearers’ expectations.

This performative approach to fluency assumes that fluency is, at root, a skill or set of procedural skills which develop over time, moving from a plodding, effortful delivery to one that is automatic and does not require a heavy cognitive load (Carlson, Sullivan, and Schneider, 1989). Such fluency depends on procedural knowledge (Faerch and Kasper, 1984): the knowledge *of* rather than knowledge *about*. An analogy could be made here that speaking fluently is like learning to drive a car: difficult at the beginning, but almost mindless at the end, if the learner successfully internalizes the rules, motions, and mechanics of the act.

Fluency in this sense exists for both productive (speaking, writing) and receptive (reading, listening) language skills categories. One might imagine a *fluent* reader as someone who can read texts of all registers with a high level of comprehension for long periods of time. By the same token, listeners might also be fluent—able to quickly and effectively process speech regardless of the speed or register with which it is delivered. In other words, *fluency* might

simply be ease of processing regardless of modality (Schmidt 1991). Most commonly, however, *fluency* is used to describe the processes through which oral communication is planned and delivered (Schmidt 1992).

Lennon (2000) provides perhaps the most comprehensive definition of fluency by stating: “[...] a working definition of fluency might be 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).”

Although it is clear that fluency on the part of the second language speaker is a skill which takes time to develop, how do we determine how or if an L2 learner has reached any degree of fluency? In order to do this someone, most likely a teacher or a native speaker must make a scalar quality judgment as to the relative fluency of the learner’s speech. Much if not most of the work in second language fluency focuses on this intersection of production and perception, and seeks to isolate and examine the features of L2 speech on which listeners rely when making such judgments.

While defining fluency as broad or narrow has helped to distinguish fluency as a mechanical function of language ability, it still does not provide a set of component skills or acoustic measures by which said fluency can be measured. For this, we must turn to a more quantitative analysis of fluency.

How is fluency measured?

To answer the question of what it means to be fluent, we must first establish a narrow definition of fluency (Lennon, 1990) as a speaker-based performance phenomenon built upon objective, measurable temporal variables (Mohle, 1984; Lennon, 1989; Schmidt, 1991).

Measurements of fluency involving acoustic data vary from study to study, but the temporal foundations of fluency are centered around pauses, which may be empty (also termed *silent*) or filled, and hesitation phenomena, a collective term for repetitions, corrections, and prolongations. This dissertation examines prolongations in depth (see chapters 4 and 5).

Silent Pauses

Pauses are the spaces between words in an utterance, which may be silent or occupied by fillers with no lexical content (such as *uh* or *er*). Some pauses are features of fluent speech such as breath breaks or natural spaces between words. These types of pauses are what Goldman-Eisler (1968) terms articulatory pauses. However, pauses have been further sub-classified either by function or by feature. One example of the first type of classification system is O'Shaughnessy (1992) who identifies a difference between grammatical and ungrammatical pauses, where the former is between clauses and the former is within a clause. Feature-based taxonomies generally appeal to pause length rather than placement in determining whether a pause is fluent or disfluent. In the seminal work on silent pauses, Goldman-Eisler (1968) differentiates between articulatory pauses, those which are breath breaks or the natural spaces between words, and hesitation pauses which may signal impediments in the planning and production of speech. The study claims 250 milliseconds to be the point at which these two pause types diverge. Though this measure has been widely used in pausological research, it is not without dispute. Brown and Yule (1983) propose a three-way taxonomy of pauses, classified by length, in which pauses of 800 milliseconds or longer denote topic boundaries, pauses between 600 and 800 milliseconds are "substantial pauses" and coincide with single contours, and pauses below 500 milliseconds, which occur in incomplete syntactic structures. More succinctly, Kirsner, Dunn & Hird (2003) point out that the majority of pauses in the range between 130 and

250 milliseconds cannot be strictly attributed to articulation due to individual variation. As a uniform approach to pause criterion has not emerged, studies have classified silent pause duration at many points along the wide spectrum from 100 milliseconds up to over a full second (O'Connell & Kowal, 1983). Other salient aspects of pausological research will be brought up throughout this dissertation.

Filled pauses

Related to pauses, non-lexical fillers or *filled pauses* (Maclay & Osgood, 1959) are studied as they are believed to provide insight into the nature of speech production. Filled pauses give indications as to the speaker's lexical access (Goldman-Eisler, 1968) and difficulties in the discourse-planning process (Chafe, 1980). As with silent pauses, there is a distinction to be made between filled pauses which serve a communicative function and those which do not. Fox Tree (1993), for instance, shows that the filler *um* may serve to help the listener discern upcoming lexical items faster. Similar results were also shown by Bosker et al., (2014) in an experiment which required speakers to make predictions concerning upcoming lexical items, and which showed that participants were more likely to predict a low-frequency lexical item when said item was preceded by a filled pause. Further discussion of the communicative effects and disfluencies in general and filled pauses in particular on the NS-NS and NS-NNS interactions is provided in Chapter 3.

Hesitation phenomena

Hesitation phenomena—repetitions, corrections, and prolongations—have traditionally been classed together and only comparatively rarely examined for their effect on fluency

judgments, typically as addenda to a larger study dealing with the more noticeable pause-related disfluencies.

Repetitions are extremely common in natural speech (Maclay & Osgood, 1959). It is believed that repetitions occur mainly when speakers opt to suspend the speech stream to give themselves additional planning time (Maclay & Osgood, 1959; Goldman-Eisler, 1968, Schegloff, Jefferson, & Sacks, 1977; Levelt, 1983, 1989;) or simply change their mind about the direction of the utterance (Clark and Wasow, 1998). Once the obstruction or planning difficulty is navigated, the last word or phrase spoken is then repeated in order to re-take the utterance.

Often wrongly conflated with repetition, self-correction makes a value judgment on the act or effect of hesitation; namely, that the intent of the speaker was to smooth out or correct a real or perceived error. Self-corrections are thought to be connected to internal language-production processes and directly affected by increases in metalinguistic awareness in the L2 (Verhoeven, 1989). In a descriptive study of self-corrections among children, Fathman (1980) noted that self-corrections could be one of five types: lexical, syntactic, semantic, morphological, or phonological. Regardless of type, it has long been suggested that self-correcting patterns are determined to a great extent by L1 and by individual habits (Lennon, 1990; Kormos, 1999).

The final category within hesitation phenomena is doubtless the least studied. Prolongations are a phenomenon of syllabic lengthening, generally occurring word-finally, which may indicate the need for added planning or recall time (e.g., *sooo* instead of *so*) (Collard,

2009)¹. Because of their use in buying planning time prolongations have often been classed with filler words (e.g., *like* or *you know*) and filled pauses (*um*, *uh*) (Munro & Derwing, 1995). Other studies have examined prolongations as a feature of intonation, since they are usually delivered in monotone (Wennerstrom, 2000). Few studies have looked specifically at the rate, placement, and delivery of prolongations in non-native speech (Eklund, 2001), and none, to my knowledge, have focused solely on the effects of prolongations upon the perception of L2 fluency, a gap which this dissertation seeks to address.

Acoustic measures of temporal variables

Taken together the use of these temporal variables helps to form a picture of the general speed and ease with which a speaker can deliver their intended message (Schmidt, 1991). The following table (Table 1) provides a useful index of the different acoustic measurements which have been used in the literature to provide a standardized, objective measure of fluency.

¹ For a more comprehensive look at the functional and disfluent aspects of prolongations and how they affect the distribution of prolongations within native and non-native speech, refer to Chapter 4.

Table 1. *Acoustic measures dimension of fluency*

Dimension	Acoustic Measure	Calculation
<i>Speed</i>	Mean length of syllable (MLS)	Total spoken time/# of syllables
	Mean length of run (MLR)	Mean number of pauses in a run (between adjacent silent or filled pauses)
	Articulation rate (ATR)	Mean # syllables per minute/total speech time (excluding pauses)
	Phonation-time ratio (PTR)	Percentage of time speaking proportional to time of total sample
	Syllables per second (SPS)	Total syllables/total speech time (including pauses)
<i>Breakdown</i>	Number of silent pauses (SP)	Number of SP/total spoken time
	Number of filled pauses (FP)	Number of FP/total spoken time
	Mean length of SP (MLSP)	Sum of SP duration/number of SP
<i>Repair</i>	Number of repetitions (NR)	Number of NR/total spoken time
	Number of corrections (NC)	Number of NC/total spoken time

Note that this measurement scheme does not include prolongations, an omission which has left an ongoing gap in the understanding of utterance fluency and all the factors which affect it. This dissertation proposes and implements an additional measure to this accepted measurement scheme, which accounts for prolongations (PR) and is modeled in the calculation for similar variables of breakdown fluency, silent and filled pauses. This would appear as below:

Prolongation rate: $\text{Number of prolongations} = \text{Number of PR} / \text{total spoken time}$

A Fluency Framework

The multidimensional nature of fluency has resulted in a number of fluency frameworks. One such framework is proposed by Segalowitz (2010). Further refining Lennon's (2000) definition, Segalowitz (2010) proposes three categories of fluency: *cognitive*, *utterance*, and *perceived* fluency.

Cognitive fluency is defined as the relative efficiency of speech planning and production and its associated processes. Problems with cognitive fluency are thought to emerge when the L2 speaker lacks a complete knowledge of the grammar, vocabulary, etc., of the target language, or inadequate skills (e.g., speed of access and articulation) in using this knowledge. Difficulties in either aspect may result in a decrease in fluency (i.e., slower and/or more disfluent speech) (De Jong et al., 2012).

Utterance fluency is the acoustic realization of fluent speech and, by natural extension, measurements of disfluencies. In this realm, Tavakoli & Skehan (2005) make a distinction between *breakdown fluency* (the extent to which filled and silent pauses interrupt the speech signal), *repair fluency* (the presence of repetitions, hesitations, etc., in the speech stream) and *speed fluency* (the rate of speech), which were further operationalized by Skehan (2003, 2009) as:

1. Breakdown fluency: the number of filled pauses, silent pauses, and silent pause duration
2. Repair fluency: the number of self-corrections and repetitions in the utterance
3. Speed fluency: the mean length of run

Finally, perceived fluency is defined by Segalowitz (2010) as the judgments which listeners make about a speaker's overall fluency based on the features of their utterance. Studies focusing on perceived fluency generally obtain fluency judgments by one of several subjective methods, the most common of which is the 7 or 9-point Likert scale. Many such studies have used Likert scale judgments to investigate the correlation between these evaluative ratings of perceived fluency and acoustic measures of utterance fluency in L2 speech (Bosker et al., 2014). The following section gives an overview of key studies in this category.

Relating Utterance Fluency to Perceived Fluency

It has long been assumed that native speakers are fluent by default (Raupach, 1983; Riegenbach, 1991); though as the definition by Fillmore (1979) given in section 2.3.1 has shown this may be true to differing degrees depending on what is meant by saying someone is *fluent*. Regardless, perceived fluency is based on the impression of the listener—generally a native speaker—of the speech performance, or utterance, of the non-native speaker. By comparing measures of utterance fluency to judgments of perceived fluency, we seek to determine which aspects of utterance fluency most strongly impact the perceived fluency of the speaker.

Each of the studies detailed below asks raters to judge the fluency of non-native speech (*perceived fluency*) which has also been analyzed for the most commonly measured objective markers of fluency: speech rate, pauses (filled and unfilled), hesitation phenomena (*utterance fluency*). What remains, then, are the many possible variations on this simple experimental design. For instance, does the type of task the non-native speakers are performing influence their ability to speak fluently? Does the L1 background of the speaker influence fluency ratings? And, finally, does the background of the rater (e.g., untrained rater vs. ESL instructor) impact fluency

ratings? This section addresses these questions by examining key studies relating utterance and perceived fluency.

Lennon (1990) examined speech by 4 female native speakers of German each of whom were advanced English learners and who had been determined by their institution to have improved in proficiency during a 6-month period in which they were residents of Britain. Tests were conducted at the beginning and end of the 6 months and it was determined that all 4 had had statistically significant improvements in the mean number of syllables per length of speech sample (MLR) and overall speech rate, as well as a reduction of hesitation markers such as repetitions, corrections, and filled pauses. These improvements correlated with improvements in fluency ratings by ten native-speaking instructors. Further, Lennon notes that there are individual differences among the subjects indicating that the perception of fluency could be affected by different characteristics in different speakers. In order to examine each individual speech characteristic (filled and silent pauses, hesitations, etc.) and its precise effect on fluency ratings, controlled experiments which isolate said characteristics are yet needed.

Riggenbach (1991) analyzed the speech of six L1 Chinese English learners, three of whom were rated as “very fluent” and three of whom were rated as “very non-fluent” by twelve native-speaking raters. This study took into account a total of nineteen variables when examining the non-native speech samples. These variables included speech rate, hesitations/repairs, and discourse measures such as backchannels, latching, overlapping, and topic initiations (among others). Of these variables only speech rate and the number of filled pauses were found to significantly affect whether the learner’s speech was rated as fluent or non-fluent. Riggenbach also noted that each participant had individualized pausing and repair “profiles” and that even

participants with near-native fluency profiles were rated non-fluent by some raters because of judgments on ungrammatical elements in the speech sample.

Freed (1995) examined oral fluency in two groups of undergraduate students, one of which studied English at home, the other of which studied abroad. In this longitudinal study, gains in fluency for both groups were measured at the beginning and at the end of their study term based on fluency features such as speech rate and number of disfluencies. This study lent support to the notion that those who study abroad have significantly more gains in fluency over time than those who study in a foreign language setting. Furthermore, it also confirmed that these “study abroad” students spoke faster and with fewer clusters of disfluencies than those who studied at home.

While both Lennon (1990) and Riggensbach (1991) were seminal in establishing the correlations between utterance and perceived fluency, both studies were drawn from relatively small sample populations. Following up on an earlier study examining L2 Dutch read speech (Cucchiaroni, Strik & Boves, 2000), a larger study conducted by Cucchiaroni, Strik & Boves (2002) examined spontaneous speech from 30 beginner and 30 intermediate level learners of Dutch as a second language. These participants were asked to speak for 15 seconds (beginners) or 30 seconds (intermediate) for each of 8 speaking tasks. These speech samples were then rated for fluency by ten teachers of Dutch as a second language. An analysis which related the acoustic measures of fluency to fluency ratings found that the results varied by proficiency level: speech rate was predictive of beginners’ fluency ratings, while mean length of run (MLR) was the most significantly correlated predictor for intermediate speakers.

Derwing et al. (2004) collected and examined speech samples from 20 L1 Mandarin beginner English learners. Participants were asked to engage in narrative, monologue and

dialogue tasks. The resultant speech samples were analyzed for MLR, speech, rate, and numbers of filled and unfilled pauses exceeding 400 milliseconds (following Riggensbach's earlier work). These samples were then rated by 28 native-English-speaking novice raters. Results showed that pausing and the speech rate measure *pruned syllables per second* (that is, the total number of syllables excluding self-corrections, false-starts, hesitations, and asides) accounted for 69% of the variance in fluency ratings.

Similarly, in a study of 16 L1 Hungarian speakers of English as a second language, Kormos and Denes (2004) found that for both native and non-native raters, speech rate, mean length of utterance, phonation-time ratio, and the number of words stressed per minute of speech were the best predictors of fluency ratings. This study differed from others, however, in demonstrating that raters varied in terms of the relative importance they placed on the length of pauses in speech samples, as well as the linguistic elements (e.g., lexical diversity, grammatical complexity, and accuracy) of a speech sample.

As studies up to this point had used a variety of different types of individuals for rating purposes, Rossiter (2009) intentionally set out to measure how different types of raters influenced fluency ratings. For this, she used groups of experts—here classified as linguistics students and English language teachers—non-experts, and advanced non-native speakers. These raters were asked to judge speech samples from 24 English learners of various L1 backgrounds. Rossiter found that regardless of the raters' profiles, speech rate and the number of pauses per second of speech were correlated with higher fluency ratings. This finding confirms that utterance and perceived fluency are strongly correlated, and that rater type does not seem to play a significant role in how fluency is perceived.

More recent studies have begun the laborious process of establishing a hierarchy or ranking of individual suprasegmental features of non-native speech and their effects on evaluative ratings. A series of related studies (Kang, 2008,2010; Kang, Rubin & Pickering, 2010; Kang, 2012; Kang & Pickering, 2012) examines how the prosodic variables of speech rate, pausing, stress, and intonation affect measures of comprehensibility and accentedness. Speech samples were collected from lectures given by 11 international teaching assistants (ITAs) and were analyzed for various suprasegmental measures including those typically associated with fluency studies: speech rate and pausing behavior. Samples were then rated along a series of 7-point Likert scales for comprehensibility and accentedness by 58 American undergraduates. This body of work collectively indicates that individual suprasegmental features contributed to raters' perceptual judgments of both comprehensibility and accentedness in slightly different ways. Speaking rate measures (mean length of run, articulation rate, and phonation-time ratio) accounted for 35% of variance in comprehensibility ratings such that speech rate was directly predictive of comprehensibility measures; in other words, the faster an ITA spoke, the more comprehensible he or she was rated as being. This measure did not carry over to judgments of accentedness, however, where a higher mean length of silent pause correlated with lower accentedness ratings. This was likely a result of ITAs having pauses that were longer and more numerous than the average and which were occasionally inappropriately-placed (e.g., not at topic boundaries). Taken together these findings show that both comprehensibility and accentedness as constructs share similarities with fluency in the general salience of both speech rate and pauses in determining the outcome of evaluative ratings.

Summary and motivation

The multidimensional nature of fluency as a construct requires a correspondingly nuanced framework to account for it. Referring to fluency as either cognitive, utterance, or perceived fluency provides a straightforward way to classify and disentangle the many competing notions of fluency, but it is also clear that these notions are never wholly independent of one another and, in fact, can best be studied in relation to one another.

Studies correlating utterance to perceived fluency are many and provide a somewhat unified picture of the effects of speech rate, pausing and, to some extent, individual disfluencies on fluency ratings. What this picture fails to show, however, is how individual disfluency types such as hesitation phenomena—and prolongations in particular—may affect fluency ratings differently depending on variables such as where they appear in the utterance, how frequently they are used, or what words and segments they affect.

Little experimental work has been done to show how changes in the disfluencies themselves may affect fluency ratings. For example, in a study done by Hahn (2004) placement of primary stress was manipulated on speech samples collected from Korean ITAs, with listeners hearing primary stress either misplaced, felicitously placed, or missing entirely. Results indicated that listeners rated speakers more favorably and recalled more content when primary stress was favorably placed. The methodology used in this study demonstrates that careful manipulation of acoustic variables related to fluent/disfluent speech can supply a more detailed picture than correlational studies had previously shown.

A series of studies on the impact of disfluencies on fluency ratings and attention conducted by Bosker & DeJong (2014) manipulated the placement and duration of silent pauses and revealed that both increases in number and increases in duration resulted in lower fluency

ratings regardless of whether the manipulated speech was from a native or non-native speaker. These results clearly indicate that listeners place a high value on disfluency patterns. Moreover, by manipulating individual disfluencies we may gain a better understanding of what aspect of the disfluency (e.g., placement, duration, frequency, etc.) is more impactful on fluency ratings.

Although not as common as correlational studies, perceptual research helps to determine the boundaries of acceptable disfluency usage, which has important implications not only for ESL/EFL pedagogy, but also for potentially providing insight into the underlying processes of speech perception for native listeners engaging with non-native speech.

In Chapter 5, this dissertation furthers the inroads into perceptual research of disfluencies by way of a perceptual experiment in which prolongations are manipulated for two conditions: frequency (high, low) and clause position (initial, non-initial), the first of its kind with prolongations and a necessary first step towards increasing the scope of what is known about this critically understudied hesitation phenomenon.

Conclusion

Thus far, the examination of the field of second language fluency has shown that an ever-growing body of research has failed to reach a consensus as to what factor or combination of factors makes an L2 speaker fluent. Studies have consistently shown that measures of speed fluency (mean length of run, articulation rate, and syllable ratios) most frequently correlate with fluency ratings (Cucchiarini, Strik & Boves, 2000; Derwing et al., 2004). In some cases, number and placement of silent and filled pauses have also been shown to affect the perception of fluency (Riggenbach, 1991). These effects have been found to hold true across groups of non-native speakers with different L1 backgrounds and in a variety of task types. Results for speech rate have been consistent for both expert and naïve raters, regardless of the types of instructions

given in a fluency rating task. By comparison, relatively little work has been done on hesitation phenomena and their individual effects on fluency ratings.

Kormos and Denes (2004) point out that empirical studies in fluency have used three general approaches in studying the topic. First, the development of fluency is examined over time (Freed, 1995; Lennon, 1990); second, fluency patterns in the form of temporal measures of fluency are compared in native and non-native, fluent and non-fluent speech samples (Riggenbach, 1991; Eijzenberg, 2000); and third, fluency ratings from native listeners are correlated with temporal measures of fluency (Kormos & Denes, 2004; Bosker, 2014; among others).

This dissertation will engage simultaneously with these three facets of fluency research to determine the patterns of L2 repair fluency in general, with a particular focus toward prolongations, and how said patterns affect the perception of overall fluency.

CHAPTER THREE

CORPUS COLLECTION AND ANALYSIS

Learner corpus research concerns itself with the collection and analysis of written and oral learner language (Granger, 2002). As a necessary bridge between the broader aims of corpus linguistics and those of foreign or second language research, learner corpora have been used with frequency since the 1980s to test hypotheses concerning theories of the acquisition of second language (Stubbs, 1996). Learner corpora are used widely for studies in contrastive interlanguage analysis, which compare learner and native language patterns, and also for error-tagging, which examines error patterns in learner speech to determine levels and patterns of difficulty in second language acquisition. As Granger (1994) points out, these studies help to inform how we view the many interwoven processes of second language acquisition and by extension can also be applied to how second or foreign languages are taught.

Learner corpora both large and small have been used to examine questions surrounding second language fluency (Freed, 1995; Cucchiarini, Strik & Boves, 2000; Kormos & Denes, 2004; Gilquin, 2008; Götz, 2013). However, the scarcity of adequately-annotated prototypical corpora tagged for phonological features and the lack of access to raw data for purposes of acoustic measures means that an independent corpus collection is often the best recourse when studying individual fluency characteristics, particularly within a selected population. For these reasons, I have chosen to collect and examine original data from a local population of second language learners at the intermediate stage of learning English as a second language.

This chapter presents a multi-phased corpus collection and analysis. First, research questions and aims of the corpus are laid out, followed by an overview of the growing field of learner corpus research which presents general research aims of the field, as well as its strengths

and shortcomings. The remainder of the chapter lays the two phases of corpus collection. First, the compilation and analysis of a small pilot corpus used to determine the effectiveness of different task types in eliciting disfluencies in a controlled setting. Second, the collection and analysis of a broader multilingual corpus of native and non-native speech, and its analysis for various measures of disfluency. Major findings concerning the differences in disfluency patterns between native and non-native speakers across and within tasks are summarized in the final section of this chapter, along with steps for further investigations into prolongations.

Research Question: Disfluency Patterns

The stated goal of this dissertation is to observe and document the distribution of prolongations in native and non-native speech and to apply those findings to a study which investigates the effects of prolongations on the perceived fluency of the same. However, the focused study of prolongations as disfluencies in second language speech is a relatively novel one. A few studies have included prolongations (Kormos & Denes, 2004; Gilquin, 2008), but more often than not their findings pertaining to this disfluency are only briefly mentioned or conflated with other disfluency types (Kormos & Denes, 2004). In order to study prolongations directly, it is necessary to first collect data from which prolongation tokens are extracted.

However, it is simply not clear where and to what extent prolongations are produced. Gilquin (2008) reports that non-native speakers produce more prolongations than native speakers; however, this data was from a corpus of L2 French learners and also did not make a meaningful distinction between functional and disfluent prolongations. Vural (2008) reported *no* prolongations in the L2 English of L1 Turkish speakers. Given these and other contradictory reports, it is important to first be certain that the data from which the prolongations are being

extracted is representative of native and non-native speech and to eliminate to whatever degree possible the possibility that the corpus data is in some way atypical.

One way to do this is to test the disfluency behavior of native and non-native speakers to see if they exhibit patterns comparable to those reported in other peripheral learner corpora (e.g., MLR, SR, filled pauses, etc.). If so, then it is reasonable to assume that the prolongation behavior is representative of native and non-native speech by extension.

Therefore, by compiling and analyzing a corpus of native and non-native speech taken from participants engaged in identical tasks, this chapter addresses the following research question:

Main RQ: What are the differences in disfluency patterns between native and non-native speakers when performing identical tasks?

It is predicted that native and non-native speech will differ significantly across all fluency measures, a result which would corroborate similar findings in previous research on speech rate (Lennon, 1990; Freed, 1995; Gut, 2012), pausing behavior (Trofimovich & Isaacs, 2012; Gut, 2012), and hesitations (Gilquin, 2008). Nevertheless, a lack of consensus in previous literature on the nature, function, and distribution of prolongations makes their patterning difficult to predict.

The main research question is approached by first compiling a pilot corpus which aims to establish an appropriate methodology for data elicitation and measurement, and from these constructing a larger corpus comprised of more native and non-native speakers, and non-native speakers from different L1 backgrounds.

Research questions for each phase of the corpus are included in their respective sections.

Learner corpus typology: An overview of the field

Granger (2002) points out that learner corpora make use of four key dichotomies in their conceptualization and construction: 1) general vs. technical language; 2) synchronic vs. diachronic collection; 3) monolingual vs. bilingual data; and 4) written vs. spoken language samples.

With few exceptions, learner corpora collect general, non-technical or non-specialist language from either classroom tasks, test materials, written output (assignments, essays, dissertations, letters), or free speech. Additionally, learner corpora are in the majority synchronic, or what in general linguistics is spoken of as *sample corpora*. These are snapshots of learner language at a particular moment in time or level of development. Diachronic corpora, which try to get a sense of learner development, are difficult to compile as they require tracking and recording a large number of learners over time. Few such corpora exist and those that do focus on young learners, whose progress can be more easily measured at intervals throughout the years of standard education (The Corpus of Young Learner Interlanguage—CYLI). Because of the inherent difficulty in compiling longitudinal data from adult learners, researchers have instead taken a cross-sectional approach in which data from groups of learners at different proficiency levels is compiled to determine relative progress across time.

Written vs. spoken

Although learner corpora are now established as valuable tools for second language research and indispensable for questions of frequency (McEnery & Wilson, 2001), there has remained until recently an imbalance in the proportion of written and oral corpora. In part this stems from what Linell (2005) described as the “written bias” in descriptive linguistics which regarded textual language as correct or grammatical and demoted spoken language to secondary

status (Gilquin & DeCock, 2011). This was likely a holdover from pre-technological eras when written texts were most readily available and in which grammars were based around written work.

Standing in the historical gap between the dominance of written corpora and the relatively modern emergence of learner-focused annotated spoken corpora, the field of Conversation Analysis (CA) provided a needed shift away from the written bias.

Arising from the social sciences, CA experts recognized the complex social functions carried out in everyday speech acts and worked toward establishing the institutional nature of conversation—that is, the subconscious, unexpressed, but universal norms which exist to govern how one person interacts socially with another through language. This version of CA which Heritage (2005) terms *basic CA*, investigates these unspoken rules through the lenses of repair, turn-taking, and other systems by which humans organize interaction. Theoretical underpinnings aside, the outworking of CA theory required large amounts of data in the form of recordings (audio or video) of people engaged in natural conversation, made interpretable by an extensive transcription and annotation scheme that helped the researcher make sense of the complex interactions happening within even the simplest of exchanges.

The following (Table 2) is an example of some common symbols in the Jefferson Transcription System, an accepted CA annotation guide (Jefferson, 2004), which on inspection would not be out of place in the laboratories compiling and annotating large quantities of spoken data today:

Table 2. *Sample of common symbols in Jefferson Transcription System*

Interlocutor	Transcription	Key
	(.)	Just noticeable pause
	(.3), (2.6)	Examples of timed pauses
	↑word,↓word	Onset of noticeable pitch rise or fall (<i>can be difficult to use reliably</i>)
A:	word [word	Square brackets aligned across adjacent lines denote the start of overlapping talk. Some transcribers also use "]" brackets to show where the overlap stops
B:	[word	
	.hh, hh	in-breath (note the preceding fullstop) and out-breath respectively.
	wo(h)rd	(h) is a try at showing that the word has "laughter" bubbling within it
	wor-	A dash shows a sharp cut-off
	wo:rd	Colons show that the speaker has stretched the preceding sound.

In addition, CA provides large-scale transcription and measurements for elements of prosody. For example, CA transcriptions differentiate between brief pauses and longer pauses which are measured for duration. Other prosodic features accounted for in CA transcription include intensity, speech rate, intonation, and the lengthening of segments. (Findings from CA specific to syllable lengthening will be discussed in more depth in Chapter 4).

As large-scale spoken corpora have been compiled and deployed for study over the past two decades, studies which have focused on spoken rather than written language have proliferated (Altenberg, 2002; Biber & Conrad, 2003; Cheng, 2007; Norrick, 2008, among others) and from them has emerged an understanding that spoken language is an entity which follows its own sets of rules and deserves independent study as an equal to its written counterpart (Hirst, 2001; Gilquin & DeCock, 2011). Also, where once the difficulty of compiling and storing recordings for spoken corpora limited their size relative to written corpora, now with the surplus of digital storage and comparative ease of access, spoken corpora are growing ever-larger. For an extensive overview of learner corpus typology and learner variables therein, see Gilquin (2015).

Using learner corpora to examine disfluency

As has been shown, the number of spoken learner English corpora has grown steadily in recent years and had now come to include corpora representing a variety of L1 backgrounds, subjects, and tasks (The Cambridge Learner Corpus, The College Learner's Spoken English Corpus (COLSEC), The English Corpus, ICLE, and HKUST Corpus of Learner English, among others). While these spoken corpora are valuable for studying questions inherent to spoken language data (i.e., lexical variation across tasks, morphosyntax, etc.) as none of these corpora are tagged for phonetics or prosody they fall short of providing a full picture of learner phonology and phonological development.

In the following section, I examine the reasons why these smaller corpora are both necessary to redress certain deficits of prototypical corpora, and also valuable for the examination of specific variables of learner speech.

Difficulties in compiling a learner corpus

At issue in this discussion is the value of a small, specialized corpus of learner speech which makes use of data elicited for a particular research purpose, sometimes termed "peripheral corpora" (Gilquin, 2015; Gilquin & DeCock, 2011). While not prototypical in that the data is not compiled from fully naturalistic environments (i.e., a classroom), peripheral corpora are actually very common in studies of second language disfluency (to cite just a few: Lennon, 1990; Riggenbach, 1991; Freed, 1995; Cucchiari, Strik & Boves, 2000; Kormos & Denes, 2004; Derwing et al., 2006).

Indeed it is difficult to see the value of distinguishing prototypical and peripheral corpora at least with regards to disfluency studies, for reasons which are enumerated in the following sections.

The “problem” of authenticity

It is commonly accepted that a prototypical corpus must be compiled from “authentic” materials, setting up a contrast between data gathered from people going about their daily lives and data gathered from experimental events (Sinclair, 1996; Gilquin, 2015). Granger (2002) goes so far as to claim use of experimental data precludes a learner corpus from being a true corpus in the corpus linguistics sense of the term, though she opens an exception by adding that authentic classroom data fulfills the authenticity criteria.

Even with this addendum, the traditional notion of authenticity in corpus linguistics poses a number of challenges for the field of learner corpus research. First, learners may not be able to produce fully authentic data. Lower levels in particular may not have the capacity to engage in prolonged classroom interactions in the target language. Classroom language is often a result of highly structured activities with set or prescribed outcomes (e.g., a grammar drill or a read-aloud task). Such outputs would necessarily fail to meet authenticity standards. Finally, even learners at higher proficiency levels may engage in learning tasks which are controlled to a greater or lesser degree. Pedagogical tasks thus exist along a continuum of naturalness, from those which allow the learner total control (i.e., spontaneous free speech) to those with moderate to high degrees of control (i.e., essay writing). If the majority of learner data compiled is of language at the lower end of the spectrum, or of tasks in which learners are not free to choose their own output (i.e., read-aloud) then Gilquin (2015) prescribes avoiding the term *corpus* in favor of a more neutral term like *database*.

For the niche between a prototypical corpus and a database, Nesselhauf (2004) introduces the idea of *peripheral learner corpora*, comprised of data from more controlled or elicited environments that still allow the learner some measure of communicative freedom. For example, Lennon (1990) uses a corpus of speech elicited using picture-based story-telling tasks in his study of advanced L1 German English learners, a practice common to studies investigating learner disfluency patterns which has been replicated in countless studies since. Other peripheral corpora have compiled data from story-telling tasks, description tasks, information gap activities, and others.

Many, if not most, peripheral corpora are compiled with the goal of studying a particular speech phenomenon (Lennon, 1990; Fox Tree, 1995; Cutting & Bock, 1997; Kormos, 2000). Perhaps the most common, studies in second language disfluency frequently make use of elicited speech, at times collecting quite extensive learner corpora representing various proficiency levels and task types, as has been shown in Chapter 2. By using elicited data, researchers are able to systematically control relevant variables to a far greater extent than in the naturally-occurring data of a traditional corpus, which is of paramount importance as different topics or tasks may be more cognitively demanding and thus impact the number of disfluencies (Kormos, 2000; Gilquin, 2015).

Although there are inherent limitations in using elicited data (lack of generalizability from small data sets, perceptual bias in transcribers), the potential benefits far outweigh these limitations. Furthermore, it should be noted that large-scale prototypical spoken corpora, in particular those which annotate for prosody (LeaP, LLC), use similar types of elicited data. This overlap seems to suggest that the question of prototypical vs. peripheral corpora is to some extent an artificial dichotomy.

Shortcomings of current options in prototypical spoken corpora

Another issue in the study of disfluency via a prototypical corpus is the shortcomings of the corpora themselves, whether in the transcription, annotation, or distribution phases.

Transcription and annotation are highly valuable processes for turning a prototypical corpus into a usable, searchable database but, as Gilquin (2015) points out, transcription is ultimately only a theoretical construct (Ochs, 1979) and subject to transcriber bias. In the case of learner corpora, these challenges are exacerbated because of heightened quantities and varieties of errors (Ringbom, 1987).

Some of these challenges are mitigated when studying disfluencies simply because prosodic annotation requires access to, at the very least, the spectrographic representation of the speech sample. Still, there remains a danger of subjectivity or bias in transcriptions. In the case of prosodically-annotated corpora this is problematic particularly if features such as prolongations or long pauses are measured impressionistically rather than acoustically (O'Connell & Kowal, 1990). In terms of studying these features, prototypical corpora do not provide exact measurements for the lengthened segment nor the ratio or deviation by which they are lengthened. The following section examines two of the larger examples of prototypical spoken corpora and their shortcomings as a tool for the analysis of disfluencies in specific populations.

LINDSEI Corpus

A learner corpus frequently used in studies of second language disfluencies, the Louvain International Database of Spoken Interlanguage (LINDSEI) is the spoken counterpart of the International Corpus of Learner English (ICLE). Though it originally began as a corpus of

interviews with L1 French learners of English it has since expanded to include data from other participant backgrounds, such as L1 Dutch, Chinese, Spanish, Swedish, Turkish, Italian and others, with a number of additional corpora (Brazilian Portuguese, Finnish, Czech, etc.) still in progress.

Since its creation, LINDSEI has been used in several renowned studies on L2 disfluency. Data from the LINDSEI corpus was used by Götz (2007) in comparison with the Louvain Corpus of Native English Conversation (LOCNEC) to conclude that L1 German learners of English use higher numbers of filled and unfilled pauses in comparison with native speakers, and also tend to use fewer contractions and repetitions of contractions than native speakers, which contributes to the perceived “non-nativeness” of learner speech. In addition, the placement of disfluencies within the sentence/utterance differs considerably from native to learner speech. Gilquin (2011) similarly found that learners tend to use hesitations (pauses, smallwords, syllabic lengthening, and truncation) more than native speakers and from this conclusion prescribed teaching learners how to use hesitations in a more native-like fashion as part of a proactive FLT curriculum.

The LINDSEI corpus annotation system accounts for features of learner speech which can and have been previously categorized as disfluencies (i.e., empty and filled pauses and syllable lengthening). It does not, however, provide any account of repetitions or self-corrections and therefore paints an incomplete picture of disfluency patterns. Other issues with LINDSEI include the fact that most of the data is compiled from oral interviews with learners, with only a small portion representing a thematically-controlled picture-description task. This means that comparing disfluency patterns across a controlled task is difficult if not impossible. Furthermore, LINDSEI has a recognized challenge in giving an account of learner level (Jendryczka-

Wierszycka, 2009, with regards to the Polish sub-corpus). Learner data was compiled from learners attending universities with different learner-leveling systems. Nor was years-of-study a reliable indicator of level, as English instruction for Polish students is notably “uneven” (Jendrycka-Wierszycka, 2009). Overall, the difficulty in accounting for level within the corpus means that using this corpus for a cross-sectional analysis of learners at a specific point in language development is difficult if not impossible; ergo, it is concluded that using the LINDSEI corpus to examine disfluency patterns of intermediate learners provides more challenges than it does answers.

LeaP Corpus

The first of its kind, the Learning Prosody in a Foreign Language (LeaP) Corpus was created by a team led by Ulrike Gut at the University of Bielefeld, Germany. Between 2001 and 2003 this project compiled four types of speech: nonsense word lists, reading of story text and extemporaneous re-telling of the story, and free speech interviews. Data was gathered from non-native speakers of German and English and represented a range of ages, sex, proficiency levels, and first languages. This corpus was the first to be fully annotated for all features of prosody including intonational phrases, filled and unfilled pauses, hesitation phenomena and elongated phonemes. In all, an average of 1000 events were annotated for each minute of recording, resulting in a corpus that continue to be, to my knowledge, the most extensive example of a prosodically-annotated corpus of learner English.

Although certainly ground-breaking in its reach and detail, the LeaP corpus still suffers from two major disadvantages for use in disfluency research. First, while pauses and hesitation phenomena are marked, they are not measured, nor are they measured proportionally. In the case of prolongations, this means that a phoneme marked as “elongated” is often marked as such only

impressionistically, without regard to its length in relation to the mean length of syllable. Nor are exact measurements readily obtainable as, just as with the LINDSEI corpus, the raw sound files cannot be accessed directly without also accessing the corpus's time-aligned stamps. Another lesser criticism is simply the utter breadth of learner type and associated metadata which makes it difficult to focus on a particular population.

Conclusion

I cannot go so far as to claim that the differences between prototypical and peripheral corpora are negligible. Prototypical corpora are by nature more extensive, more streamlined in distribution and accessibility, and more varied, while peripheral corpora tend to focus on a particular type of learner or speech. Nevertheless, the similarities in elicitation devices and the shortcomings of currently available prototypical corpora lead me to conclude that a specialized learner corpus is the optimal solution for a study of disfluency patterns in a cross-linguistic sampling of intermediate L2 English learners.

Furthermore, while I recognize the methodological questions and limitations implicit in the term, for clarity and ease data compiled in the following sections is referred to as a learner corpus in the broad sense, as data which has been carefully designed, systematically compiled, annotated, and analyzed.

Pilot Corpus

Motivation

The direct comparison of native and non-native prolongation patterns is a novel one. Having established in the previous section that existing prototypical corpora are not optimal for addressing questions requiring careful measurement and manipulation of acoustic representations

of disfluency, instead a peripheral corpus was designed and compiled specifically for the study of prolongations in native and non-native speech.

Task Type

Studies in second language disfluency collect learner data using different elicitation tasks, which can vary from study to study. The most frequently used elicitation tasks are narrative tasks, read-aloud tasks, and free speech or spontaneous speech tasks.

Narrative tasks require participants to describe a picture or pictures or, alternatively, to create a brief narrative based on a sequence of pictures. Narrative tasks have a rich history in pausological research (Riazantseva, 2001), but have also been used extensively in second language research on disfluency (Derwing et al., 2004, 2009; Kormos & Denes, 2004; Trofimovich & Isaacs, 2012). Similarly, read-aloud speech has often been used, frequently in contexts where the goal of research is to contrast native and non-native pronunciation and prosody (Hahn, 2004; Trofimovich & Isaacs, 2012). Finally, free speech tasks or extemporaneous speaking can be elicited either by having a participant respond in monologue to a prompt (De Jong et al., 2012) or by asking participants to engage in dialogue either with another learner/participant or with an investigator (Kormos, 1999; Segalowitz & Freed, 2004).

The pilot study placed these three tasks side by side to determine which provided an environment both controlled yet still conducive to the production of disfluencies in both native and non-native speech. This aim is stated as RQ3A below:

RQ3A: What task type provides the best environment for examining disfluencies?

An effective task for the purposes of eliciting disfluency production is one which is controlled across speakers. Every speaker should be asked to perform tasks that are identical in topic, duration, and level of difficulty. This final criterion is often not met in studies of second

language learners, simply because every person learns a second language at a different rate and to a different level of ultimate success. In spite of this, steps should be taken to equalize the level of difficulty whenever possible—in the case of the narrative and free speech tasks, by allowing planning time before the participant is required to speak; in the read-aloud task, this is done by selecting sentences with level-appropriate vocabulary and structure so that reading difficulties will be minimized. These steps do not entirely equalize differences in the level of difficulty relative to the proficiency of the speaker, but they can mitigate the worst of these effects.

Measurement Type

Based on the psycholinguistic definition of fluency provided by Segalowitz (2010), a number of quantitative measures for fluency have been used in studies of second language fluency as speed measures (mean length of run, speech rate) are thought to reflect the ease with which speech is planned and produced (Lennon, 1990; Freed, 1995; Cucchiaroni et al. 2000, 2002; Towell, 2002; Segalowitz & Freed, 2004; Gut, 2012). Implicit within these measures are the number and duration of individual disfluencies, namely silent pauses, filled pauses, and hesitation phenomena (Lennon, 1990; Freed, 1995; Cucchiaroni et al. 2000, 2002; Gut, 2012).

There are many competing methods for measuring temporal variables related to fluency—ranging from two measures (Freed, 1995; Towell et al., 1996) up to as many as 12 (Kang, 2010). For this reason, it is important to actively test a proposed system to determine if its parameters are adequate for the data at hand (Gut, 2012), stated as RQ3B below:

RQ3B: What acoustic measures of temporal variables are best suited to examining native and non-native disfluencies?

The pilot study will utilize a measurement scheme set forth within the Derwing and Munro canon (2004, 2005, 2009, among others) and later followed by Isaacs & Trofimovich

(2012), who set 400 milliseconds as a baseline for silent and filled pauses. Once identified, silent and filled pauses are counted as raw numbers per unit of speech (within 60 seconds, for instance). Hesitation phenomena, which in this scheme include only hesitations and self-corrections, are given as a total number of events per unit of speech. The speed measures in this scheme are mean length of run (MLR) and speech rate (SR). Because the pilot study is primarily concerned with evaluating task and measurement type, prolongations will not be identified or measured within this section.

Predictions

The pilot corpus is designed to test the applicability of previous task and measurement types to a new population of language learners. The pilot corpus will test three elicitation tasks in common use among learner corpus and fluency researchers with the goal of deciding which one: a) provides a controlled environment in which all participants, native or non-native, undertake an identical task; b) provides the best environment for disfluencies to occur by requiring on-line speech planning and a moderate but manageable cognitive load; and c) results in data which patterns closely with that of previous similar peripheral corpus studies.

It is predicted that the narrative task will prove to be the most successful at meeting all three criteria. A majority of learner corpus and fluency studies use variations of picture description or narrative tasks to elicit some if not all of their data. Both prototypical and peripheral corpus studies using narrative tasks have shown that native speakers speak faster (as realized in mean length of run and speech rate measures), use more words (words-per-minute, mean length of run, and syllables-per-second), and use fewer silent and filled pauses than non-native speakers. These observations have held true over studies with various different populations of learners (Riggenbach, 1991; Kormos & Denes, 2004); therefore, it is predicted

that the data elicited from the population of intermediate learners reported in this corpus will both confirm the favorability of the narrative task in eliciting disfluencies and also show differences between native and non-native disfluency patterns corresponding to previous findings. If this does not occur, then it can be concluded that either the sample population performs atypically, in which case the population may be unsuitable for an examination which hopes to eventually generalize its findings. Similarly, if native and non-native speakers perform in unexpected ways as a result of this elicitation task, it will be passed over in favor of one which is considered more favorable in terms of alignment with previously published research.

Of the three task types, it is predicted that the narrative task will prove to be the most effective in eliciting disfluencies within a controlled environment and also corroborate findings from previous studies using similar tasks.

Predictions for measurement schemes are more fluid. It is predicted that all measures will be heavily influenced by the 400 millisecond threshold for silent and filled pauses, and also by disfluency rates being given as raw totals as this does not take into account the full range of variability in individual speech differences.

Methods

Participants

In the first phase of the study, seven non-native speakers of English were recruited from the English Language Institute of The University of Texas at Arlington. These represented different levels (Beginner=2; Intermediate=3; Advanced=2) and L1 backgrounds (Arabic=1; Spanish=3; Vietnamese=3). In addition, seven native monolingual English speakers were recruited from among undergraduate classes at The University of Texas at Arlington and recorded performing the same tasks.

Materials

Different task types were created for this experiment in order to compile a broad corpus of native and non-native speech. These are shown in detail below:

- Task 1: This task consisted of a narrative activity requiring that participants tell a story based upon three thematically-controlled sets of six progressive pictures². In Stage 1A (original telling of story) and 1B of the task (re-telling of story) (conducted non-consecutively) participants were first allowed to look over the pictures for one minute (timed by the investigator) without speaking. Then, they were told that recording would begin and that they should strive to speak for 2-3 minutes, though they would not be penalized for going over time or not reaching the time target.
- Task 2: In order to elicit examples of free speech, task two consisted of two open-ended (discussion) questions, chosen from among 8 possible prompts based on English proficiency and participant response (Appendix B). Participants were given unlimited time in which to respond. If the participant felt that the question was too difficult or did not apply to them, they were offered another question.
- Task 3: This task was composed of 5 longer texts of differing lengths and registers, including fictional and personal narratives, as well as journalistic pieces which participants were asked to read aloud (Appendix C).

Tasks 1 and 2 were conducted as an interview between the investigator and

² Picture stories in both the pilot corpus and the main corpus were selected from among the Picture Stories for Adult Health Literacy, created by Kate Singleton, available as open-source material at: http://www.cal.org/caela/esl_resources/Health/healthindex.html

participant. Task 3 was presented as part of an auto-timed PowerPoint presentation, and texts were timed for either 30 seconds or 60 seconds depending on the length of the material to be read aloud. Participants were instructed to read at a natural pace, but were informed that slides would advance automatically and could not be reversed.

Procedure

Participants were recorded in a quiet lab environment (Trimble Hall 301, Speech Sounds Lab, The University of Texas at Arlington) using a portable Zoom H2n field recorder. Prior to beginning the procedure, participants were given brief instructions on the use of the recording device and the different experimental tasks. Participants first completed Task 1 with the Investigator. Once Tasks 1 and 2 were completed, participants moved to the auto-timed portion of the procedure to complete Task 3.

Analysis

For this pilot corpus, temporal variables were measured using the standard of acoustic analysis of prosody set by the Derwing and Munro body of work (2003, 2004, 2005, 2009, among others), and followed by Isaacs & Trofimovich (2012). These measures are detailed below:

Individual disfluency measures:

1. Number of silent pauses (SP): Total of any unfilled pause between utterances.

Following Derwing et al. (2004), only pauses of 400 milliseconds or more are counted for both pause measures (silent *and* filled).

2. Number of filled pauses (FP): Non-lexical pauses such as *um, uh, er, like*, etc. As with silent pause measures, only filled pauses of 400 milliseconds or more are counted for this measure.

3. Repetitions and self-corrections (Corr): Repetitions (e.g., “*And then she she said she would go*”) and self-corrections (e.g., “*But first he goes to went to the store*”) are enumerated as events in the utterance.

Speed measures:

4. Mean length of run (MLR): An average of the number of syllables produced between adjacent pauses (filled or unfilled).
5. Speech rate (SR): The calculation of total number of syllables divided by total response time (including all pause and disfluencies) and multiplied by 60.

Speech samples were analyzed using Praat version 5.4.05 (Boersma & Weenink, 2015).

Results for comparison of native and non-native speech samples by task

Tables 1, 2, and 3 show comparisons of native and non-native prosodic variables in the narrative, free speech, and read-aloud speech tasks, respectively. Statistical significance was calculated using a paired-samples t-test.

Narrative task

Each participant contributed two narrative tasks, for a total of 14 possible samples in each condition (native, non-native). Individual speech samples from the narrative task were cropped to 60 seconds. The first 10 seconds of each sound file was first excluded to minimize inadvertent inclusion of hesitation phenomenon associated with beginnings of utterances, as well as possible investigator interference. The subsequent 60-second interval was selected for analysis. In the non-native condition, one participant was excluded for failing to adequately complete the task, and one other participant failed to reach the 70 second threshold on one of two story-sequencing tasks, leaving a total of 11 non-native samples for analysis. In the native condition, 5 speech samples were excluded for failing to meet length threshold, leaving a total of

9 samples for analysis. Results for native and non-native speakers are given in Table 3 below:

Table 3. *Comparison of NS and NNS disfluency measures in narrative task*

Measure	Reporting Unit	NS Mean	NNS Mean	<i>p</i>
	<i>Number of silent pauses over 400ms within speech</i>			
SP	<i>sample</i>	13.25	18.42	*
	<i>Number of filled pauses over 400ms within speech</i>			
FP	<i>sample</i>	1.33	8.00	*
	<i>Number (not duration) of hesitations, repetitions, and self-corrections</i>			
Corr		0.56	9.00	*
	<i>Mean number of pauses in a run (between adjacent silent or filled pauses)</i>			
MLR		15.24	5.91	**
	<i>Total number of syllables divided by length of speech sample, multiplied by 60 (seconds)</i>			
SR		196.70	114.39	**

***=significant at $p < .001$; **=significant at $p < .01$; *significant at $p < .05$

Non-native speakers produce significantly more silent (SP) and filled (FP) pauses than native speakers. It could potentially be argued that this is a result of the measure chosen—native speakers simply do not produce very many pauses longer than 400 milliseconds, and therefore this result is unreliable. However, even when the pause measure is lowered to 250 milliseconds (well within native speaker range), the difference is still significant ($p < .05$). Likewise, filled pauses, argued to be the most salient characteristic of disfluent speech (Götz, 2013), are also significantly more likely to appear in non-native speech than native speech, even at 250 milliseconds.

Speech rate (SR) also differs significantly from native to non-native speakers, being well within the established range (120-260) across individual speech characteristics and task type (Riding and Vincent, 1980; Richards, 1983). Similarly, Mean length of run (MLR) is significantly higher for native speakers, meaning that they are able to fit more syllables between pauses, likely due to proficiency in phonological changes associated with

connected speech—linking, reduction, assimilation and blending (Kahn, 1976; Kaisse, 1985). These results are to be expected given that SR and MLR are generally highly correlated to overall proficiency (Götz, 2013).

As data across all measures is significantly different for native and non-native speakers, we can assume that the narrative task provides a roughly naturalistic picture of native and non-native disfluency patterns based on previous research. This task also meets the control criteria for an ideal elicitation device laid out in section 3.4: as all participants engage with the same narrative prompt, topic and register are controlled across language background.

Free speech

As with the story-sequencing task, within the free speech task several samples were excluded for failing to meet the length criterion (a sample longer than 40 seconds). The remaining speech samples were from four native and four non-native participants. Results are represented in Table 4 below:

Table 4. *Comparison of NS and NNS disfluency measures in free-speech task*

Measure	Reporting Unit	NS	NNS	<i>p</i>
		Mean	Mean	
SP	<i>Number silent pauses over 400ms within speech sample</i>	7.333	8.167	
FP	<i>Number of filled pauses over 400ms within speech sample</i>	2.333	3.667	
Corr	<i>Number of hesitations, repetitions, self-corrections</i>	0	1.167	
MLR	<i>Mean number of pauses in a run</i>	10.642	7.93	*
SR	<i>Total number of syllables divided by length of speech sample, multiplied by 60 (seconds)</i>	166.892	134.275	**

***=significant at $p < .001$; **=significant at $p < .01$; *significant at $p < .05$

This task revealed no significant differences between native and non-native speakers on either pause measure, nor in the Corr measure. It was expected that completely unplanned speech would be easier overall for native speakers than non-native speakers; instead, what is shown is

that both native and non-native speakers employ roughly equivalent numbers of filled and unfilled pauses. The cognitive load of this particular task on both speaker groups simply levels differences—in other words, there is not enough fluent speech in the sample to set apart disfluencies between native and non-native speakers.

Differences re-emerge in the speech rate measures, but this is not surprising given that native speakers are able to fit more syllables into a run regardless of how many breaks exists between runs. It is therefore concluded that a truly free speech task, such as the one reported above, with no previous planning time, is simply too cognitively burdensome to allow natural speech patterns to appear and may not be an appropriate elicitation task for examining disfluencies in both native and non-native speakers.

Read-aloud

The simplest text within the data set was examined in order to preclude as much interference from lack of reading fluency as possible. No participants had to be excluded from analysis. A total of 14 speech samples (7 native, 7 non-native) were analyzed. On average, participants took 45 seconds to 1 minute to read the passage, but speech samples were cropped to 30 seconds in order to eliminate both early disfluencies and long silences at the end of the recording. Results are shown in Table 5 below:

Table 5. Comparison of NS and NNS disfluency measures in read-aloud task

Measure	Reporting Unit	NS Mean	NNS Mean	p
SP	<i>Number of silent pauses over 400ms within speech sample</i>	7.4	8.4	
FP	<i>Number of filled pauses over 400ms within speech sample</i>	0	0	
Corr	<i>Number (not duration) of hesitations, repetitions, and self-corrections</i>	0.4	0.4	
MLR	<i>Mean number of pauses in a run (between adjacent silent or filled pauses)</i>	13.289	8.505	**
SR	<i>Total number of syllables divided by length of speech sample, multiplied by 60 (seconds)</i>	201.032	140.693	**

***=significant at $p < .001$; **=significant at $p < .01$; *significant at $p < .05$

Significant differences exist between native and non-native speakers in both MLR and SR. However, this task again shows no significant differences between native and non-native speakers in pauses or hesitation phenomenon. It was initially feared that reading disfluencies would negatively color the results on this task, but that did not appear to be the case, though this may be the result of the relative simplicity of the text. Interference from reading disfluency cannot be discounted in other read-aloud tasks. Further, it was impressionistically evident that not only were speakers pausing the same amount, but they were also pausing in roughly the same locations, even for those who were low-proficiency English learners. This means that the only distinction between groups is that the number of syllables an average native speaker can produce per minute is higher than what a non-native speaker is able to produce. This function of read speech effectively eliminates processing disfluencies, and clearly indicates that this type of task is not an adequate elicitation tool for disfluencies.

Discussion

The goal of the pilot corpus was to determine what elicitation tasks and measurements

were most appropriate for a small but representative corpus of intermediate learner speech. The results showed that of the three task types considered, the narrative task was most effective in providing a naturalistic but controlled environment for the study of disfluencies. This confirms a view from task-based research (Skehan, 1998) and pausological research (Goldman-Eisler, 1968) that different content places variable cognitive demands on different speakers; therefore, content must be controlled in order to be able to make a quantitative comparison of fluency across speakers and language backgrounds.

However, while this same level of content control exists in a read-aloud task, this task is not naturalistic in that it does not require on-line speech planning, relying instead on the participants' ability to recognize words and pronounce them. Thus, any disfluency emerging in a read-aloud is likely a result of difficulties with reading fluency and not cognitive fluency. This was supported by the results of this pilot corpus which showed that both native and non-native speakers paused in the same places (at clausal boundaries and at punctuation) and for roughly the same amount of time.

Finally, elicitation via a free speech task fails to provide a controlled environment for studying disfluencies (Riazantseva, 2001; Kormos & Denes, 2004). Even though the prompt is identical across speakers, individual variations in prior knowledge, listening skills, and topic-competence mean that the cognitive load for such a task is noticeably higher than for the controlled narrative task.

Given these results, the larger corpus in 3.5 will use a sequence of narrative tasks as the main elicitation device and only data from these tasks will be used to examine questions concerning native and non-native fluency.

This pilot corpus also revealed that the system of measurements used is insufficient for

a full examination of disfluency. As predicted, the 400 millisecond threshold used by Derwing and Munro (1999) and Trofimovich and Isaacs (2012) caused difficulties. Native speakers simply do not produce very many pauses longer than 400 milliseconds, and therefore this result is unreliable. Furthermore, because the number and placement of pauses is so intrinsically linked with MLR, the higher threshold may also be skewing results in speed measures. This cut-off point may also be contributing to an unintentionally inflated number of non-native pauses versus native pauses, since native speakers have shorter pauses overall and may simply have fewer pauses that reach this threshold. A small trial revealed that when the pause measure is lowered to 250 milliseconds (well within native speaker range), the difference between native and non-native speakers is still significant ($p < .05$), but this lower threshold allowed for native speakers' overall shorter silent pauses. These arguments point toward the need for an adjusted pause threshold. The large corpus (Phase Two) will use 250 milliseconds as the pause threshold in compliance with recent studies in disfluency (Towell et al., 1996; Kormos & Denes, 2004; Bosker, 2014) as well as accepted pausological research in the accepted length for articulatory vs. hesitation pauses in pausological research (Goldman-Eisler, 1968; Kirsner, Dunn & Hird, 2003).

Another issue with the pilot corpus measurements is that a raw count of silent pauses, filled pauses, and hesitation phenomena (self-corrections, repetitions, prolongations, etc.) does not allow for an adequate comparison across speakers. This measurement should instead be formulated as a ratio of total to spoken time (total time of utterance minus total pause time).

Finally, the classification system for hesitation phenomena does not provide enough differentiation between the different disfluency types. For instance, hesitation may be realized as either a repeating phoneme (e.g., "sh-she"), a full repetition (e.g., "she...she"), or a

lengthened phoneme or *prolongation* (e.g., “she:”). The current measurement system does not differentiate between these, and this means that a study of individual disfluencies is heavily limited. As this dissertation is primarily interested in hesitation phenomena in general and prolongations in particular, the larger corpus will adopt a measurement system that takes each of these variations into account. These disfluencies must also be measured in a way that makes comparison possible across participants—for this reason, the main corpus will utilize a ratio similar to that employed with pauses: total number divided by total spoken time.

In addition to the adjustments made for task type and measurement type, due to a certain cumbersomeness in the elicitation process of phase one, steps were taken to minimize direct investigator involvement in phase two and thus make participation uniform across participants. All images and instruction were consolidated into a single, timed, automatic PowerPoint presentation. It was expected that not only would this provide a more controlled experimental setting, but it would also allow for more simplicity in transcription and analysis.

Full corpus collection and analysis

This section presents the collection and analysis of a more extensive corpus of native and learner speech via a narrative elicitation technique shown to be effective in phase one of the broader corpus study. By examining disfluency patterns of native and non-native speakers in identical tasks, we will be able to determine the main differences and where further investigation is warranted.

Methods

Participants

Data was compiled from 19 non-native participants, all of whom were enrolled in the English Language Institute at The University of Texas at Arlington. All non-native participants

were classed as B1/B2 or C1/C2 according to a program designed around the Council of Europe Common European Framework of References for Languages (2001), whose guidelines state the following criteria for levels B1-C2:

Level	Characteristics
C2	Can understand with ease virtually everything heard or read. Can summarize information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
C1	Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structure, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices.
B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various topics.
B1	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 whilst traveling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes, and ambitions and briefly give reasons an explanations for opinions and plans.

Figure 1. Levels of learner proficiency by Common European Framework (2001)

Participants were asked to take a Language Background Questionnaire (Appendix A) to self-report proficiency level, academic level, and language contact information. At the time of the study, participants reported having spent between 1 month and 1 year in the United States. While most participants reported learning English during secondary education, this had no effect on their overall proficiency. Eleven participants were male and 8 were female. The number of participants by L1 background is reported in Table 6 below:

Table 6. Number of non-native participants by L1 Background

Background	Number
Korean	3
Chinese	4
Vietnamese	6
Turkish	1
Arabic	2
Spanish	3

In order to compare non-native disfluency patterns to those of native speakers, speech samples were also compiled from 10 native English speakers (7 female; 3 male) all of whom were undergraduate students at the University of Texas at Arlington.

All participants were compensated in the amount of 5 dollars per 30 minutes of participation or any portion thereof. Average participation time was 40 minutes, though each participant file contained about 30 minutes of speech, for a total of approximately 10 hours of recorded material.

Procedures

Participants engaged in two types of tasks for this portion of the study. In the first, they undertook a narrative task) frequently used in pausological and disfluency research (Riazansteva, 2001; Kormos & Denes, 2004) and shown through the pilot corpus to be effective in the study of disfluencies for this particular population.

- Task 1: In this task, participants were given three thematically-controlled sets of six progressive pictures. As with the first phase, participants were first allowed to look over the pictures for one minute (timed by the investigator) without speaking. Then, they were told that recording would begin and that they should strive to speak for 1-2 minutes. To ensure that participants would not go over time, all materials were presented in a timed

PowerPoint display which automatically advanced to the next slide when the time limit had been reached. There was no penalty for going over time or failing to reach the 1-minute target.

- Task 2: Participants were given a series of syntactically and lexically controlled sentences to read aloud. These were used as fillers separating the different instantiations of Task 1 and were not included in the analysis described below.

Analysis

Speech samples for analysis were taken from the narrative task (Task 1). Samples which failed to reach the 60-second threshold were excluded from analysis. For this reason, the final number of speech samples was 116, of which 92 were from non-native participants and 24 were from native participants.

Starting points for each sample were set at the onset of phonation of the earliest discernible word and ended with the start of a pause. This was done to ensure that pause measures (and, by extension, runs) were not artificially inflated.

First pass analysis was conducted in Praat, using a four-tiered text grid. Tier 1 contained a broad transcription of the utterance. Silent pauses were marked on Tier 2 and filled pauses were marked on Tier 3. Both filled and silent pauses over 250 milliseconds were first identified using an automated script (Mark-Pauses, Mietta-Liennes, UCLA), which automatically inserted boundaries at the beginning and ends of pauses using intensity analysis. Each pause parameter was then re-checked and reset manually if needed in order to correct for disruptions from background noise which the script could not overlook. Tier 4 contained syllable counts for each run, the relative boundaries of which were determined by using the information in Tiers 2 and 3.

Total measures of silent and filled pause time were calculated using an automated script (Split-Count, UCLA). Other types of disfluencies (hesitations, corrections, repetitions, and prolongations) were marked as parenthetical notations on the transcription tier and were calculated by total number rather than by duration. These same features were later re-analyzed independently for further details.

Figure 2 below shows a screenshot of a single speech sample for analysis.

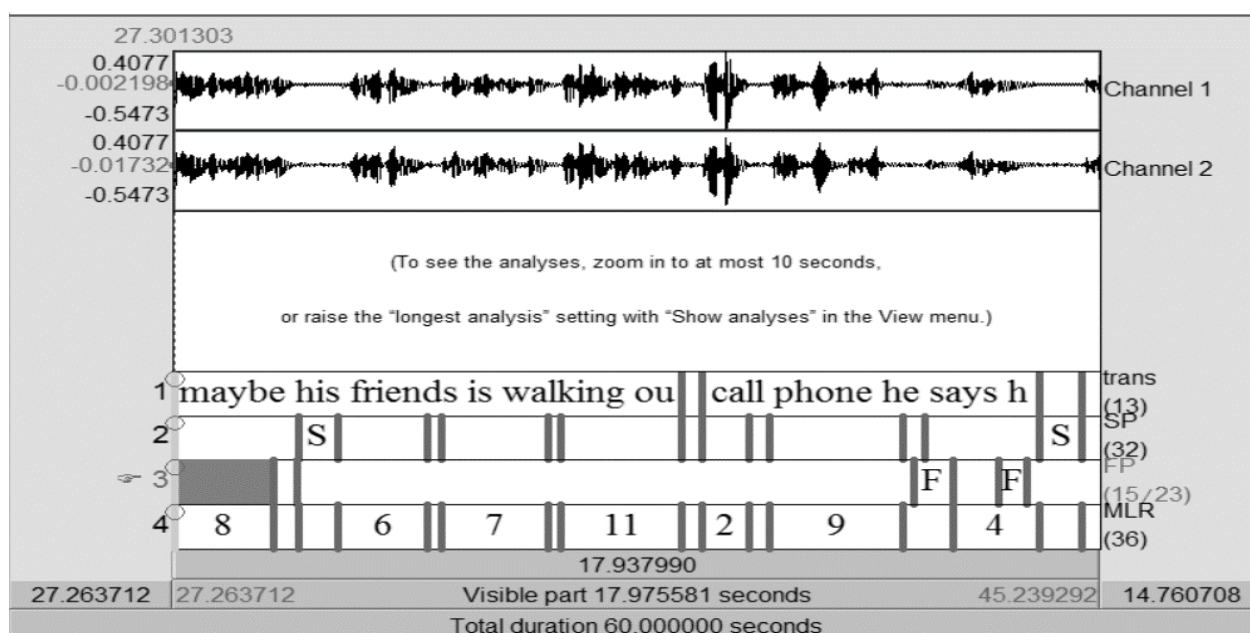


Figure 2. Screenshot of analysis in Praat

Using these base calculations, a number of temporal variables were examined. Using the taxonomy proposed by Skehan (2003; 2009) these variables were broken down according to whether they are related to speed, breakdown, or repair fluency. This scheme is similar to one used by Bosker (2014) with the exception of the prolongation rate measure, which has been created for the sake of this dissertation.

Speed:

1. Speech rate (SR)

Speech rate was calculated by determining the total amount of syllables and then dividing this total by the amount of time in seconds used to produce the speech sample (including pauses), following the method set by Riggensbach (1991). This figure was then multiplied by 60 to standardize the figure across speech sample of slightly varying lengths.

2. Mean length of run (MLR)

The mean length of run was calculated as the total number of syllables divided by the total number of runs, producing the average of the number of syllables produced between pauses or filled pauses over 250 milliseconds (Goldman-Eisler, 1968; Kirsner, Dunn & Hird, 2003).

Breakdown:

3. Silent pause (SP)

Silent pauses were calculated by taking the number of pauses over 250 milliseconds divided by the total time spent speaking. Due to small differences in the exact length of speech samples, rather than providing a raw number of silent pauses (as was the case in section 3.4) pause numbers were provided as a ratio of total number to spoken time (total time minus total pause time), following a pattern first set out in research by Bosker (1987), Riggensbach (1991), and Kormos & Denes (2004). This measure and others below (FP, PR, NR, and NC) share the same polarity: the higher the number, the less fluent the speech sample.

4. Filled pause (FP)

Numbers of filled pauses were calculated by dividing the number of filled pauses by the total spoken time.

5. Mean length of silent pause (MLSP)

The mean length of silent pause was calculated by sum of silent pause duration divided by the number of silent pauses.

6. Prolongation (PR)

This number was calculated by dividing the total number of lengthened syllables or *prolongations* by the total spoken time.

Repair:

7. Number of repetitions (NR)

The number of repetitions was calculated by dividing the total number of repetitions by the total spoken time. A repetition in this case is seen as a single event. Therefore a sentence like: “*And then she...she went to the store*”, even if divided by a silent or filled pause is considered a single repetition.

8. Number of hesitations or corrections (NC)

The number of hesitations or corrections was calculated by dividing the total number of hesitations or corrections by the total spoken time. Hesitations and corrections were classed together because in many instances it is difficult to ascertain the underlying production difficulty. For example, a sentence like “*When John went to the store she...uh, he bought a ball*” clearly contains a necessary shift in agreement from female to male pronoun. However, a sentence like “*Jane sh-s-said she would not go*” contains what on the surface appears to be both a phonological hesitation effect (*s-said*) and what

may also be a correction (*sh...said*). With this in mind, this dissertation will consider hesitations and corrections together—not because they are identical phenomena but because without access to the intent of the speaker it is difficult to disentangle these two concepts.

In the following section these calculations will be used to address RQ3C and RQ3D, restated below for clarity:

RQ3C: To what extent, if any, do disfluency patterns vary significantly across instantiations of the same task (T1, T2, T3)?

RQ3D: To what extent, if any, do disfluency patterns differ from tell to re-tell across native and non-native participants?

Or, in other words, do native and non-native speakers display fatigue effects from beginning to end of participation time (T1-T3), and do participants become more disfluent, less disfluent, or remain consistent from tell to re-tell?

Results and discussion

A series of t-tests were conducted on the aggregated data for native and non-native speakers to determine whether there was a main effect between three different narrative tasks (T1, T2, T3) which would indicate a fatigue effect over total participation time. Results of a one-way ANOVA showed that there was no significant difference between disfluency rates in T1, T2, and T3 in 8 of the 9 tested measures: MLR $F(2,113) = .26, p = .77$; SR $F(2,113) = .16, p = .85$; MLS $F(2,113) = .06, p = .94$; SP $F(2,113) = .63, p = .53$; FP $F(2,113) = .97, p = .38$; MLSP $F(2,113) = .188, p = .16$; NR $F(2,113) = .87, p = .42$; or NC $F(2,113) = .108, p = .34$). These results and means are tabulated in Table 7 below:

Table 7. *Global means and effects of trial on fluency measures*

	MLR	SR	MLS	SP	FP	MLSP	PR	NR	NC
Task 1	7.055	139.07	.352	.467	.105	.687	.054	.056	.043
Task 2	6.402	131.91	.359	.520	.105	.720	.064	.057	.045
Task 3	6.542	134.24	.356	.512	.130	.773	.098	.074	.033
Mean									
p	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.

***=significant at $p < .001$; **=significant at $p < .01$; *significant at $p < .05$

There was also no difference when Tasks 1, 2, and 3 were compared in native and non-native speakers separately, as Table 8 shows:

Table 8. *Means and effects of trial on temporal variables by language background*

	MLR	SR	SP	FP		MLR	SR	SP	FP
NS 1	11.89	199.41	.310	.072	NNS 1	5.674	121.83	.512	.114
NS 2	11.76	201.59	.350	.065	NNS 2	5.016	113.93	.564	.116
NS 3	12.77	214.83	.330	.062	NNS 3	5.032	114.71	.555	.147
Results	n.s.	n.s.	n.s.	n.s.	Results	n.s.	n.s.	n.s.	n.s.

As seen in Table 8, a one-way ANOVA revealed no significant differences between the speech rate and mean length of run for native and non-native speakers. While these differences were not significant, it is interesting to note that, overall, native speakers tend to get incrementally faster from T1 to T3, while non-native speakers become incrementally slower. This may be an early indication that fatigue plays a greater role for non-native speakers, while familiarity tends to benefit native speakers in terms of their speed. Nevertheless, as these differences are not statistically significant, it is safe to conclude that results that there is no significant fatigue effect and data from all three trials can be used for purposes of fluency analysis.

The second question under consideration was whether there was a difference in disfluency patterns from the first narrative (Tell) and when participants were asked to recount their original narrative from memory without access to the image prompt (Re-Tell). A series of two-sample t-tests revealed that there was no main effect between the two trials (Tell, Re-tell) for any of the nine variables: MLR $t(113.42)=.06, p=.95$; SR $t(113.72)=.38, p=.71$; MLS $t(113.56)=-.37, p=.72$; SP $t(113.50)=-.02, p=.51$; FP $t(110.79)=.15, p=.88$; MLSP $t(113.93)=.61, p=.54$; PR $t(112.72)=.54, p=.59$; NR $t(112.76)=2.586, p=.56$; or NC $t(113.82)=1.089, p=.28$. These results are laid out in Table 9 below:

Table 9. *Global means and effects of tell vs. re-tell on fluency measures*

	MLR	SR	MLS	SP	FP	MLSP	PR	NR	NC
Tell	6.633	133.39	.359	.501	.113	.717	.069	.066	.036
Re-Tell	6.676	136.52	.353	.500	.115	.740	.077	.059	.044
<i>p</i>	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

This result held true also when native and non-native speakers were considered independently (Table 10):

Table 10. *Means and effects of tell vs. re-tell on fluency measures by language background*

	MLR	SR	SP	FP
Native Tell	12.430	202.474	.322	.059
Native Re-tell	11.859	208.073	.338	.073
Results	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Non-Native Tell	5.121	115.369	.547	.127
Non-Native Re-tell	5.323	117.857	.542	.126
Results	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>

The lack of significant difference between tell and re-tell may be due to a number of factors. First, a first-pass tell has a certain cognitive load to which story sequencing, lexical retrieval, and even creation of a narrative voice may play a part. While these factors are largely resolved in a re-tell event, it is possible that the cognitive load of recall mimics that of the “tell” effects very closely and thus there is no noticeable difference in disfluency patterns. It is also possible that the disfluency effects *are* different, but they simply do not show up numerically; in other words, it may be that individual participants behave differently from tell to re-tell but these small effects are washed out in the composite data. Finally, it could be the case that the differences in disfluency patterns from tell to re-tell are not necessarily in the quantity of disfluencies but in their placement. However, these are considerations for further study—in the absence of significant effects for trial, data for both tell and re-tell will be utilized in the final analysis.

Thus far, results have shown that there was a) no significant fatigue effect from T1 to T3 and b) no significant difference in disfluency patterns from tell to re-tell. With these parameters in place, it is possible to approach the main research question of the overall differences in disfluency patterns between native and non-native speakers performing identical tasks. Table 11 below gives an overview of the results of a series of two-sample t-tests conducted to measure the effect of language background (NS, NNS) on different measures of fluency:

Table 11. Means and effects of language background on fluency measures

Measure	Reporting Unit	NS Mean	NNS Mean	<i>p-value</i>
Speed				
MLR	<i>Mean number of pauses in a run (between adjacent silent or filled pauses)</i>	12.14	5.222	***
Speech Rate	<i>Total number of syllables divided by length of speech sample, multiplied by 60 seconds</i>	205.27	116.61	***
Breakdown				
SP	<i>Number of silent pauses/total spoken time</i>	.330	.545	***
FP	<i>Number of filled pauses/total spoken time</i>	.067	.126	***
MLSP	<i>Sum of silent pause duration/number of silent pauses</i>	.664	.745	*
PR	<i>Number of prolongations/total spoken time</i>	.051	.079	*
Repair				
NC	<i>Number of corrections/total spoken time</i>	.011	.048	***
NR	<i>Number of repetitions/total spoken time</i>	.018	.074	***

***=significant at $p < .001$; **=significant at $p < .01$; *=significant at $p < .05$

As expected, native and non-native participants demonstrated significantly different patterns of disfluencies within a narrative task. As with phase one, native speakers spoke significantly faster than their non-native counterparts, ($t(30.09)=12.24, p < .0001$). This was also true for the secondary speed measure, MLR ($t(27.64)=10.357, p < .0001$). In addition, it is clear that non-native speakers pause more frequently ($t(95.85)=6.98, p < .0001$), and for longer periods of time ($t(73.99)=-2.50, p < .05$). Similarly, non-native speakers use filled pauses more frequently than native speakers ($t(92.75)=-4.10, p < .0001$). They also have higher rates of hesitation phenomena whether self-corrections, ($t(105.98)=-7.00, p < .0001$); repetitions, ($t(113.61)=-6.54, p < .0001$); or prolongations, ($t(72.86)=-2.41, p < .05$).

Discussion

For the purposes of this dissertation, the corpus corroborated significant differences between native and non-native differences across all measures, as reported in previous corpus

and fluency studies. Therefore, it can be concluded that the data reported in this chapter is within range of acceptability for the disfluency production and there is no reason to assume the prolongation data is atypical.

In terms of prolongations, data from this corpus showed that of the hesitation phenomena (hesitations/repetitions, self-corrections, and prolongations) prolongations were the most frequent for both native and non-native speakers, an observation which corresponds to similar findings in Eklund (2002), a seminal study on the features, distribution, and cross-linguistic variation of prolongations. In fact, prolongations are so ubiquitous that they are outnumbered only by silent and filled pauses (Eklund 1999, 2001, 2002) and yet they remain the least studied of the hesitation phenomenon.

Despite clear impressionistic differences between how native and non-native speakers employ prolongations, the study of prolongations is often conflated with other hesitation phenomena, such as work by Kormos & Denes (2004) which classes them with filled pauses, and Wennerstrom (2000) which addresses them only as an intonational feature. A few recent studies have dealt with prolongations independently in Hungarian (Deme & Marko, 2013), Mandarin Chinese (Lee et al., 2004), and Turkish (Kuruoğlu & Altıparmak, 2015); however, prolongations in English as a second language remains a largely unexamined field, with the exception of studies which correlate overall number to fluency ratings.

For these reasons, the remainder of this dissertation will address the issue of prolongations in detail, focusing on how they are used by native and non-native speakers (Chapter 4), and how they may impact the perception of fluency by native speaking listeners (Chapter 5).

CHAPTER FOUR

SURVEY OF PROLONGATIONS

Prolongations are segments which are markedly longer than those found in fluent speech (Bell et al., 2000) and are, along with filled pauses, the most common hesitation phenomena in spontaneous speech (Eklund, 2001). Despite this fact, they remain the “dark horse of the disfluency stable” (Eklund, 2001) and are often overlooked in disfluency research.

Research Questions

This chapter presents the analysis of prolongation distribution within the corpus of non-native speech compiled in Chapter 2, with the goal of addressing the following question:

Main RQ: What are the differences in prolongation distribution in native and non-native speakers of English?

In order to address this question fully, prolongations will be examined with regards to their duration, position, phonology (syllables and segments), and position within the utterance.

These aims have been formalized as the following research questions:

RQ4A: What are the durational differences between native and non-native prolongations?

RQ4B: What are the syllabic and segmental distribution differences between native and non-native prolongations?

RQ4C: What are the differences in syntactic distribution between native and non-native prolongations?

Literature Review

Speakers have various ways of signaling hesitation. These include extending a silent pause for longer than average, filling the space with filled pauses (um, uh) or filler words (you know, like). In addition to these, the speakers may choose to prolong a sound or syllable for longer than the same sound or syllable would typically be in fluent speech. Though this type of hesitation phenomena is also referred to as *lengthening*, this dissertation instead adopts the umbrella term *prolongation* which absorbs the related terms *syllable lengthening* (Gilquin, 2008), *vowel lengthening* (Deme, 2013) and *drawl* (Fox Tree & Clark, 1997), which have been used to describe prolongations in certain phonological or pragmatic contexts.

Prolongations and filled pauses are the most common hesitation phenomena in spontaneous speech (Eklund, 2001). In addition, both filled pauses and prolongations are hesitations which make use of duration and vocalization to fill space (Deme, 2013), and both have been observed to mark the ends of turns or the intention to hold the floor (Du Bois et al., 1993). As a result of the similarities between filled pauses and prolongations in features, frequency of use, and functional overlap, they are often grouped into the same category (White, 1997) though direct evidence to justify such a pairing is scant (Eklund, 2001; Deme, 2013).

Both prolongations and filled pauses are *disfluencies*, detritus from the speech planning process which buy the speaker additional time to formulate and articulate the utterance—*disfluencies*. Much research has noted that the higher the cognitive load the more disfluencies emerge (Maclay & Osgood, 1959; Goldman-Eisler, 1968; Clark & Fox Tree, 2002; Gosy, 2006). If this is the case, then the equivalence in frequency between filled pauses and prolongations should signal that they are of equal importance in highlighting areas of difficulty the speaker encounters during speech planning and production (Esposito, 2006). It may also be the case that

the distinction between prolongations and filled pauses is one of individual habit or a carry-over from L1 patterns and functions—but the fact that this distinction remains largely unexamined means that much work remains to be done to fully understand prolongations.

Identifying prolongations

Contributing to the lack of significant research in prolongations is the continuing difficulty in determining what actually constitutes a prolongation. Deme (2013) levels the criticism that much of the research done on prolongations to date has relied on the auditory judgments of the researchers in identifying prolongations (Eklund, 1999; Bell et al., 2000), but it is difficult to imagine how a prolongation might be identified otherwise. Indeed Deme herself identifies prolongations somewhat impressionistically with regards to length if not intonation (f_0). Even in recent work (Collard, 2009; Gut, 2012) prolongations are almost exclusively first identified as prolongations and then measured, rather than the other way around.

On deeper inspection, however, initially identifying prolongations subjectively can be effective for two reasons. First, prolongations depend on both phone lengthening and a level pitch in equal measure. Where a level pitch track (f_0) is relatively easy to spot, the lengthening of a segment may not be. Second, as with other disfluencies, prolongations are subject to a high degree of individual variation. Some speakers may use prolongations infrequently and some not at all, or may use them in some registers and not others. Furthermore, individual variation means that the average length of one speaker's syllable can be slightly or very different than the average for another. Therefore, lengthening that syllable might result in a wide variety of durations, making it difficult to determine on the basis of measurements alone when exactly a segment is “prolonged” (Eklund, 2001). Unlike computerized acoustic scripts, the human ear is remarkably adept at adapting to and making predictions based on the speech performance of the interlocutor,

allowing for more fine-grained filter for spotting prolongations. This adaptability is fortuitous as phone-lengthening serves a number of very crucial pragmatic functions.

Functions of prolongations

Phonological and syntactic lengthening

Any survey of prolongations must recognize that syllables are lengthened as a result of various phonological processes operating above the level of the segment (Savithri, 1984). For example, stressed syllables are longer than unstressed syllables as a general rule (Oller, 1973). Syllables are also longer than average at the end of an utterance, similar in duration to words spoken in isolation, a phenomenon known as *prepausal lengthening* (Gaitenby, 1965; Klatt, 1976). Finally, segments may be longer in certain syllable positions as opposed to others (House & Fairbanks, 1953; Klatt, 1976) or as a result of the segment itself—consonant clusters, for instance, have greater duration than simplex segments (Haggard, 1973; Klatt, 1976).

Phonological prolongations interact with syntactic lengthening. Phrase structure lengthening occurs at the end of constituent clauses (Martin, 1970; Klatt, 1976). One study showed an increase of 30% in vowel duration at the boundary between noun and verb phrases and between main and embedded clauses (Klatt, 1975). Other studies have shown that far from being a uniform process, there is a great deal of variability in what types of clauses give rise to phrase-final lengthening (Klatt & Cooper, 1975). There may also be interactions between phonological and syntactic lengthening—for example, a consonant in a preferred position for lengthening may be lengthened because of the syntactic position of the word.

Other interactions may occur between phonological, syntactic, and pragmatic lengthening—another functional aspect of syllable lengthening.

Pragmatic-functional lengthening

The lengthening of segments has long been noted to serve a pragmatic function in conversational turn-taking. Du Bois et al. (1993) note that syllables or segments within syllables are lengthened to indicate the speaker's intention to hold or relinquish their conversational turn, a phenomenon which appears to hold true across languages, with similar findings for Finnish (Ogden, 2001), French (Vaissiere, 2002), Japanese (Koiso et al., 1998) and others. The floor-holding or floor-ceding function of prolongations critically interacts with intonation: a final rising or falling tone on a lengthened syllable at the end of the utterance can indicate turn-yielding, while turn-holding is signaled by keeping a level intonation (Duncan & Fiske, 1979; Ford & Thompson, 1996).

Aspects of turn-taking also play a role in the function of lengthening identified in conversational analysis (CA). One way in which CA documents the reflexive and highly organized nature of English conversation is to study turn exchanges using the documentary method of interpretation (Schegloff, 1978). The notion of adjacency pairs is foundational to any CA study of turn-taking. An *adjacency pair* is a unit of conversation involving two turns in which both interlocutors are responsible for one of the turns. Examples of common types of adjacency pairs include question-and-answer, greeting-greeting, compliment-response, and offer-acceptance or offer-refusal (Schegloff & Sacks, 1973). Each adjacency pair is made up of turns, or *parts*, identified as either first or second parts. First parts include questions, greetings, offers, or anything which constitutes the opening gambit to the conversational turn. Second parts are the responses to first parts and are *preferred* when the response is considered favorable (i.e., likely to be well-accepted by the first speaker), or *dispreferred* when they are considered unfavorable (Levinson, 1983). Speakers can put off issuing a dispreferred second part by way of audible

devices (e.g. , “umm, sure, yeah, thanks, but...”) or by lengthening the syllable immediately preceding the dispreferred second part. The example below shows an example of lengthening in a dispreferred second part. The lengthened segment is immediately followed by a colon (:).

Speaker A: This wasn't my best work was it?

Speaker B: Um, **we:ll**, no, it wasn't.

Another pragmatic function of prolongations is in marking emphasis (also referred to as *focus*). Emphasis is an acoustically complex set of features which includes alterations of pitch and intensity and can also involve lengthening of phones or whole syllables (Ladd, 1996; Kohler, 2006). These temporal correlates of emphasis are somewhat consistent across languages (English: Cooper et al, 1985; Dutch: Cambier-Langeveld, 2000; Swedish: Heldner & Strangert, 2001; Chinese: Lee et al., 2004).

One example of this might be emphasis which denotes a topic-shift or signals new or unexpected information:

Speaker A: John bought a new car for his daughter.

Speaker B: No, John bought the new car for his **so:n**.

Lengthening can serve as an intensifier, particularly in the case of common modifiers. Consider the following two examples as responses to the question “How is John feeling today?”:

A. He is very tired.

B. He is **ve:ry** tired.

While the surface meaning is identical in both utterance, (3A) is more straightforward while (3B) is meant to underscore the degree of tiredness John is experiencing.

It is evident from these examples that intentionally lengthening a sound, syllable, or word can serve a variety of pragmatic functions. As a result, studies dealing specifically with the behavior and distribution of prolongations must make a distinction between functional and disfluent prolongations. Hereafter, references in this dissertation to prolongations will refer *only* to those judged to be disfluencies in natural speech, and not lengthened syllables which may be phonologically or syntactically induced or serve a pragmatic function.

Prolongations as disfluencies

Studies in prolongations as disfluencies fall into two major categories: 1) studies which give an account of prolongations in a first language; 2) studies which compare prolongation distribution between two language groups. The following section explores key studies in both categories.

Most studies address prolongation distribution patterns in the speech of native speakers of a given language. Eklund and Shriberg (1998) showed similar disfluency patterns in English and Swedish L1 speech. In a follow-up study, Eklund (2001) contrasted the distributions of prolongations in three Swedish corpora and one Tok Pisin corpus and found that the prototypical Swedish prolongations occur on a final continuant of a function word. Tok Pisin data revealed a similar preference for prolonging closed classes (function words) but the segmental and distributional data revealed deep cross-linguistic differences. Speakers of Swedish and Tok Pisin prolong most often at the end of a syllable, but the initial-medial-final distribution ratio is different for both languages, which the researchers suggest is a consequence of the fact that phonological length exists in Swedish but not in Tok Pisin. In a study of prolongations in Mandarin Chinese, Lee et al., (2004) found a similar preference for word-final prolongations, but this study engaged only four speakers for data samples and so these results may not hold up with

a broader corpus. Den (2003) identifies six strategies in prolongation use for Japanese speakers including prolongations of discourse markers, final vowels of content words at the beginning of an utterance, final vowels of both content and function words at the end of an utterance, and continuants at the disruption point of a word fragment (Den 2003: 90). In a study of filled pauses and prolongations in European Portuguese, Moniz, Mata & Viana (2007) showed that these two disfluency phenomena occur in complementary distribution and when used correctly are highly useful in buying planning time before producing complex syntactic constructions.

Others have tried to establish differences between speaker attributes/characteristics and their influence on PR distribution. Andrade & Martins (2011) studied the disfluency patterns of 136 adult speakers of Brazilian Portuguese and found that there was no difference in prolongations between male and female speakers. A small difference was found between the prolongation rates of participants with higher education versus those with only high school education, but this was shown to be an interaction with certain syntactic constructions and not thought to be indicative of heightened disfluency as a result of less education. Deme (2013) examined the speech of eight Hungarian children and found that prolongations were used more frequently than filled pauses as a discourse management strategy, a finding which raises the possibility of age-related differentials in the pragmatic uses of prolongations. In a similar age-based study in Turkish, Kuruoğlu & Altıparmak (2015) compared the distribution of prolongations in the speech of four age groups (4-8 years old, 18-23, 33-50, and over-50). Results showed that participants over the age of 50 produce more prolongations than any other group in prepared speech. In spontaneous speech, both the over-50 group and the 4-8 group were more disfluent than the 18-33 and 33-50 groups. This extensive study by Kuruoğlu & Altıparmak

(2015) both lends support to a significant age variable in the production of prolongations, and also shows that neither gender nor educational background seem to have a significant effect.

Very few studies have looked at the distribution of prolongations in non-native speech. One example of this is Vural (2008) who found that there were no prolongations in the second language English of Turkish L1 learners of English when engaged in a free conversation task on concrete (e.g., giving directions) and abstract topics (e.g., effects of religion). Other findings involving prolongations are buried in larger second language fluency studies (Flege, Munro & Fox, 1994; Gilquin, 2008, among others). These studies have concluded that prolongations are common disfluencies but challenging to engage with and have gone no further in examining the differences between native and non-native prolongation patterns.

In order to address this research gap, the following study looks at duration, position, phonology (syllables and segments), and position within the utterance of native and non-native prolongations.

Methods

Measurements

Prolongations were collected from the full corpus of learner speech compiled and analyzed in Chapter 3, where they had been impressionistically identified by the researcher. In this survey, each prolongation was first analyzed acoustically using Praat (Boersma & Weenink, 2015). Unless otherwise stated, all measures in this chapter are given in seconds.

Measures of individual prolongations were compared against the individual speaker's mean length of syllable (hereafter, MLS). Only those lengthened syllables which were at least 2 standard deviations (SD) longer than the average MLS across the utterance were considered to be

true prolongations. PR were measured from the onset of the initial segment of the word containing the prolongation to the point where phonation ceased. (Note: though disyllabic prolongations were rare in this corpus, the choice was made to measure both syllables of a disyllabic word containing prolongations as it was found that in most cases the lengthening occurred to some degree across the whole word).

To demonstrate this measure, an example is provided from a male participant whose L1 was Korean, and whose mean length of syllable for this particular utterance was .43 seconds. The following image (Figure 3) shows two instances of the word *she* in the utterance. The left side of the image shows the average or unprolonged word, measuring .32; the prolonged version of the word, measuring .48 seconds is shown on the right.

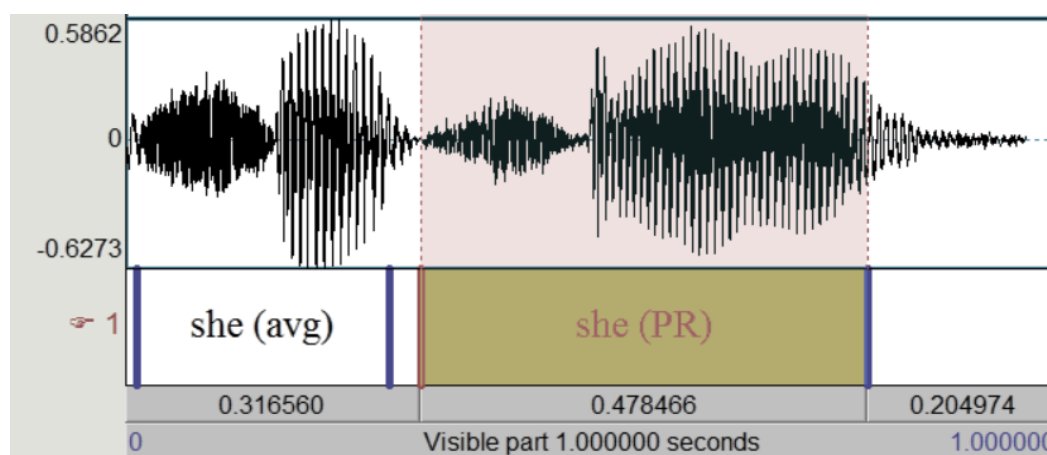


Figure 3. Spectrographic representation of average and prolonged 'she' (in milliseconds)

An impressionistically-identified prolongation was only considered a true prolongation (and included in this survey) if its duration was at least .05 seconds (50 milliseconds) above that participant's MLS for the utterance. This threshold was chosen because the MLS already represents a great deal of variance within itself and so any syllable longer than the mean by even 50 milliseconds is noticeably lengthened. Instituting the .05 difference from MLS eliminated ten

prolongations produced by non-native speakers, and no prolongations produced by native speakers. In the instance of the L1 Korean participant who produced the two syllables in Figure 3, this threshold is quite obviously met and so the impressionistically identified prolonged *she* is used for the broader survey reported in this chapter.

The same 60-second utterances used for the corpus study reported in Chapter 3 were used for this survey. However, as the purpose of the study was to examine the features and distribution of prolongations, only utterances containing prolongations were chosen. As not every participant produced prolongations, this resulted in a smaller number of utterances under examination than that reported in the corpus study in Chapter 3. The total number of utterances containing prolongations was 84 (NS=19; NNS=65), from 19 speakers (NS=4, NNS=15).

These 84 utterances contained a total of 52 native and 189 non-native impressionistically-identified prolongations which met the .05 second threshold between MLS and prolongation. The overview of the data for analysis is presented as Table 12 below:

Table 12. *Data overview for prolongation survey*

	NS	NNS
Number of speakers	4	15
Speech events with PR	19	65
Total number of PR	52	189

Following previous research (Eklund, 2000, 2001; Bell et al., 2004; Gilquin, 2008; Andrade & Martins, 2011; Deme, 2013; Kuruoğlu & Altıparmak, 2015, among others) this study examined the distribution of prolongations in duration, syllable position, segment type, word

class, and position within the sentence. Details on the method for addressing each aspect of prolongations is given within the respective sections.

Results

Frequency

Frequency counts were taken for the whole corpus reported in Chapter 3 which showed very little difference between the average number of prolongations for native and non-native speakers (NS=2.13; NNS=2.19). However, these numbers do not tell the whole story. Speakers vary widely with regards to disfluency use, with some avoiding some types while overusing others, only using one kind (Vural, 2008; Gilquin, 2008). Because this study is interested in how people who employ prolongations use them, only samples from prolongation users was selected. This more selected data tells a different frequency story: when only samples with prolongations were considered, the average native speaker used 3.62 prolongations, while the average non-native speaker had 4.2 prolongations. While these numbers are more distinctive than averages when considering the whole corpus, there are other factors to consider.

It is challenging to determine the exact frequency of prolongation, particularly when comparing native and non-native speakers. Native speakers, as has already been noted in Chapter 3, generally speak more because their greater proficiency allows them to produce more language in a given time compared to non-native speakers. Therefore, to better understand frequency one can look at frequency as a proportion of prolonged words to the total number of words spoken in the corpus. In a study comparing the learner English of French L1 speakers to native English speech, Gilquin (2008) noted that learners were far more likely to use prolongations (which she termed *drawls*) than native speakers, using a frequency basis of occurrence per 100,000 words—a typical measure used in corpus linguistics. Although the corpus reported in this dissertation is

far smaller, total word tokens were calculated for each speech sample utilized in this corpus and PR totals were taken as a percentage of that total:

Table 13. *Percentage of prolongations within NS/NNS word tokens*

Lang. Back.	PR	Total word tokens	Percentage
NS	52	3521	1.48
NNS	189	5288	3.57

Another way to approach the frequency question is to look at syllable counts for native and non-native speakers. Prolongations, after all, are lengthened syllables; therefore, perhaps the most representative way to determine prolongation frequency is to a frequency measure which is a ratio of PR syllables to normal syllables within the 60-second speech sample. Here the numbers are quite telling: prolongations occur in 1.6% of syllables occurring in native speech samples while they occur in 4.3% of syllables in non-native speech samples.

To conclude, when frequency is calculated as a total across all data, prolongation use is roughly equivalent for native and non-native speakers. However, the picture becomes more nuanced when looking only at speech samples containing prolongations and when calculating frequency either as a ratio of prolonged words to total words or prolonged syllables to total syllables. Both accounts show that non-native speakers are more likely to have disfluent prolongations than their native counterparts.

Duration

The duration of the prolongation was measured in milliseconds and was a measure of the entire word or syllable in which the prolongation was found. This method of measurement takes into account that the prolongation often extends over multiple segments and can involve

elements of co-articulation. Attributing prolongation to the entire syllable also allows for more accurate comparison with the mean length of syllable (MLS).

The mean duration for all native and non-native prolongations pooled (N=241) was .66 seconds. The 95% confidence interval was .68/.63. Means and confidence intervals for native speaker data alone (N=52) and non-native data (N=189) are given in Table 14 below (all durations shown in seconds):

Table 14. *Confidence intervals for prolongation duration*

	Parameter	Estimate	Lower CI	Upper CI
NS	<i>Mean</i>	.63	.59	.68
	<i>Std. Dev.</i>	.16	.13	.20
NNS	<i>Mean</i>	.67	.63	.68
	<i>Std. Dev.</i>	.17	.16	.20

Although Chapter 3 showed there is a significant difference between mean length of syllable of native and non-native speakers, ($t(59.23)=-12.85, p <.0001$), a paired t-test revealed that there was no significant difference between the mean duration of native and non-native prolongations ($t(86.61)=-0.97, p .34$) shown in Table 4.2. In other words, even though there was a difference between the average length of native and non-native syllables, there was *not* a difference between the average duration of native and non-native prolongations.

In order to explore this finding in more depth, the difference between MLS and the duration of each individual prolongation token was calculated. The mean difference between MLS and prolongation was .28 seconds for non-native speakers and .36 seconds for native speakers. This equates to an overall average percentage of increase of 82.95 percent when all data points are

taken together; however, separated by language background (NS/NNS), the average percentage of increase for non-native speakers is 74.24% from MLS to prolongation, while for native speakers this rate of increase is 131.28%. This discrepancy may indicate that there is an upper articulatory limit for prolongations that prevents them from extending beyond a certain duration.

This data set highlights a very particular lower threshold for prolongations of around .5 seconds (500 milliseconds) per syllable for prolonged syllables. This does not mean that the prolonged segment alone is 500 milliseconds, but that the entire prolonged syllable reliably reaches this cut-off point in 215 of 241 prolongation tokens and holds true regardless of the individual speaker's mean length of syllable. When considered separately, native speakers reach the 500 milliseconds threshold in 77% of prolongation tokens (40:52) and non-native speakers in 93.65% (177:189). This effect remains robust even when disyllabic tokens are removed under the assumption that they are longer by definition and thus would logically exceed 500 milliseconds under normal conditions. The 500 milliseconds threshold is reached for 89% of prolongation tokens with disyllabic items excluded.

In conclusion, there are no significant differences between the durations of native and non-native prolongations, but the data yields interesting observations nonetheless. When data is considered as a whole, the average for both native and non-native prolongation is roughly the same. It should then be considered that the underlying basis for slight differences in prolongation duration is to be found in individual disfluency habits. Still, there does seem to be a reliable preference for having prolongations reach or exceed 500 milliseconds regardless of language background, segment type, or position. It bears mentioning, however, that duration is a notoriously challenging way to study prolongations as Eklund (2001) and others have observed; thus, any durational findings must be subject to further examination in a larger corpus.

Syllables and syllable position

Syllable and syllable position were examined by first coding all prolongation tokens by syllable number as either monosyllabic (1 syllable), disyllabic (2 syllables) or multisyllabic (>2 syllables). No multisyllabic words containing prolongations were found. The overwhelming majority of prolongations were found in monosyllabic words, with only a small percentage (12:242, 4.98%) being disyllabic. Of those which were disyllabic, 11 of 12 were produced by non-native speakers.

In addition to syllables per word, the position of the prolongation within the syllable was also taken into account. Given the amount of variability in how prolongations are produced, there is a degree of “noise” in the data in the form of prolonged rhymes (N=3), single-phoneme words like “I” and “a” (N=3) and words in which the speaker prolonged all segments across the syllable (N=2). For clarity of analysis, these forms were removed from the overall total prior to calculating syllable position, with new overall prolongation totals for NS (50) and NNS (183).

Syllable position could be initial, medial, or final. These terms are not equivalent to *onset*, *nucleus*, and *coda*. Because the overwhelming majority of lengthening occurs on the vowel, it would be almost meaningless to claim that the nucleus is being prolonged—what is more interesting is to determine where within the syllable the prolongation is occurring. In the examples below, a colon (:) indicates lengthening and is placed immediately after the representation of the lengthened segment:

1. Initial: a:nd, i:f
2. Medial: bu:t, the:n
3. Final: she:, he:, in:, then:

Table 15 below shows total tokens and percentages for prolongation by syllable position:

Table 15. *Prolongations by position within syllable*

		Initial	Medial	Final
NS	Tokens	19	3	28
	%	38.00	6.00	56.00
NNS	Tokens	35	33	115
	%	19.13	18.03	62.84

Previous studies have shown a preference for final-position prolongations across languages (Eklund, 2001; Lee et al., 2004). Eklund (2001) claims a 30 (initial)-20(medial)-50(final) ratio for prolongations in American English, Swedish, and Tok Pisin. The current study does not support these numbers for either native or non-native speakers, though native speakers approximate it far more closely than do non-native speakers. While the prevalence of syllable-final prolongation is not in question, it is evident that native speakers rarely prolong medially, whereas non-native speakers prolong in initial and medial positions relatively equally.

Segment type

Given the preference for syllable-final prolongations, instances of prolongation were also examined by segment and position, as shown in Table 16:

Table 16. *Prolongations by segment type and syllable position*

	NS	NNS
Total Number of PR	52	189
Net total excluding: Total-rhyme PR, exceptional cases	46	183
Initial		
Number /Percentage V	18 / 39.13	29 / 15.85
Number /Percent C [+son.]	0	5 / 2.73
Number /Percent C [-son.]	0	0
Medial		
Number /Percentage V	2 / 4.35	32 / 17.49
Number /Percent C [+son.]	0	2 / 1.09
Number /Percent C [-son.]	0	0
Final		
Number /Percentage V	20 / 43.48	103 / 56.28
Number /Percent C [+son.]	6 / 13.04	10 / 5.47
Number /Percent C [-son.]	0	2 / 1.09

This data set shows that both native and non-native speakers demonstrate a preference for lengthening vowels in final position over any other combination. Native speakers seem to be able to prolong in both initial and final position with nearly-equal frequency, but non-native speakers prolong vowels in final position at a much higher rate than they do initial segments. Segments in medial position are lengthened only rarely in comparison with other syllable positions. While results for non-native speakers agree at least somewhat with those for other languages (Eklund, 2001; Den, 2003; Lee et al., 2004), native speakers seem to disregard the medial position while non-native speakers are able to prolong medial segments without constraint. It should be noted that this may be a function of the native samples under analysis—more data is needed for further comparison.

Lengthened consonants were far less frequent than lengthened vowels and most were [+sonorant], the exception being a few tokens which had prolonged fricatives [s] and [f]. This is unsurprising in that segments which allow for continuous airflow are more easily prolonged than those which obstruct airflow. Table 17 below shows the three most common segments by

language background and syllable position. To normalize segment counts, the ratio is given as a percentage out of total segments within prolonged syllables (620 segments total, NS=142; NNS=478).

Table 17. *Total and percentage of frequent segments by syllable position*

	NS			NNS		
	Segment	Total	% of 142	Segment	Total	% of 478
Initial	[æ]	18	12.68	[æ]	17	3.56
	[ɪ]	5	3.52	[ɪ]	3	.63
				[ɔ]	3	.63
Medial	[ə]	1	.7	[ɛ]	7	1.46
	[ɛ]	1	.7	[i]	5	1.05
	[ɪ]	1	.7	[ɪ]	4	.84
Final	[i]	18	12.68	[i]	62	12.97
	[n]	4	2.82	[n]	15	3.14
	[r]	4	2.82	[u]	13	2.72

As previously mentioned, vowels heavily outweigh all other prolonged segments and consonants are mainly [+sonorant]. With the exception of [i]/ [ɪ] no other segment appears frequently in more than one syllable position. It is possible that this indicates a possible interaction between segment and position, or it may be that the interaction is more complex: a large percentage of prolongations are found in function words, and the limited nature of this class may yield an artificially inflated numbers for certain segments. For example, [i] in particular has a high incidence of prolongation because *he/she* and their variants are high-frequency in the data as a result of the narrative structure and topic.

Word class

In order to study the distribution of prolongations across the syntactic boundaries of the utterance, tokens were first classed as either content or function (also referred to as *open* or *closed* classes). Results are shown in Table 18 below:

Table 18. *Content vs. function tokens of prolongations*

	NS	NNS
Content		
Number of tokens	1	31
Percentage	1.92%	16.4%
Function		
Number of tokens	51	158
Percentage	98.08%	83.6%

A Pearson chi-squared (two-tailed) test revealed that the difference in content and function distribution between native and non-native speakers is significant at $p < .001$. Native speakers showed a marked preference for prolonging content words, while the rate of prolonged function word tokens in non-native speech is far higher. This distributional imbalance may also be connected to the difference between prolongations at clause boundaries, which are common in NS disfluencies. This distinction will be examined in the following section.

Position within sentence

For the sake of simplicity, the definition of clause boundary adopted for this study is that of Foster, Tonkyn, & Wigglesworth (2000) (see also, Kahng, 2014) in which a clause required to consist minimally of a subject + verb (complement). Clauses begin at the subject or its related article and end at the same. If no second part of the sentence exists (i.e., simple sentences as opposed to complex or compound-complex sentences), the boundary is placed at the end of the sentence. For example:

- Clause-initial items (indicated in bold) are those found immediately following major clausal boundaries:

She ate the cake / **they** made for the party

- Clause non-initial words (indicated in bold) are located between but not at the edge of major clausal boundaries:

They **ate the delicious cake** / the chef **baked for Tom**.

- Conjunctions (indicated in bold) were considered separately for first-pass analysis. These were any words which connected independent clauses:

John came to the party **and** he ate cake.

Numbers and corresponding percentages of prolongations in initial, non-initial, and conjunction positions are given in Table 19:

Table 19. *Initial vs. non-initial tokens of prolongations*

	NS	NNS
Initial		
Number of tokens	25	69
Percentage	48.08%	36.51%
Non-initial		
Number of tokens	10	87
Percentage	19.23%	46.03%
Conjunction		
Number of tokens	17	33
Percentage	32.69%	17.46%

The percentage of initial, non-initial, and conjunction prolongations differed significantly between native and non-native speakers, (Chi-squared (3, $N = 241$) = 12.04, $p < .001$). This difference between native and non-native speakers remains even if conjunctions are aggregated within the clause-initial class (NS: 42:52, 80.77%; NNS: 102:189, 53.97%). Position within the sentence is where the greatest difference between NS and NNS distribution lies and may contribute to the impressionistic differences between native and non-native prolongation use. For example, if native listeners are conditioned to hear prolongations mainly in function words at clausal boundaries then an utterance with prolongations located on content words within clauses

may affect listeners' perceptions not only of prolongations but also of the speaker's overall fluency.

Discussion

The results of this survey have shown that there are observable differences in the frequency and phonological and syntactic distributions of native and non-native prolongations.

Phonological differences within this data set support previous findings for the preference of syllable-final PRs of vowels. Consonants are prolonged less frequently than vowels, but those which are prolonged are continuants. While this is not unexpected, a number of questions remain for further study. First, future work should compare distribution of prolongations across different syllable positions and segments with the general frequency of segments in those positions utilizing a calculator of phonotactic probability. This will also provide the basis for an interactional analysis of syllable position, segment type, word class (content/function) and phrasal position (initial/non-initial).

Similar work remains to be done in the syntactic aspects of prolongation distribution. Though it has been shown that non-native speakers are statistically more likely than native speakers to a) prolong content words, and b) prolong words in non-initial positions, this reveals the need for a more nuanced account of the constituents of each utterance. This study was based on a simple syntactic distinction between clause initial and non-initial tokens. Given the extensive work done in phonological and syntactic lengthening (Klatt, 1976; Klatt & Cooper, 1975) it is evident that this particular factor needs to be more carefully accounted for, as there may be gradations in prolongation length resulting from interactions between disfluent prolongations and functional lengthening. Future work will add the following classifications at

the phrasal level and have three levels of position: clause initial, clause final, within-clause; phrase-initial, phrase-final, and within-phrase.

Of the factors under consideration in this study, only duration showed no major differences between native and non-native speakers. However, there seems to be a preference for extending prolongations to a certain threshold, an interesting observation. One initial theory to explain this interesting observation is that the phenomenon has its roots in the need for speakers to disambiguate disfluent prolongations from phonological and syntactic prolongations, which tend to have lengthened durations closer to the median. Much more investigation is needed to corroborate this initial finding and tease apart the many challenging variables which affect syllable duration in context.

This study has several shortcomings which will be addressed in future work. First, the lack of a test for interrater reliability brings into question the objectivity of the measurements. While there are some things to be gained from having a single rater such as uniformity of measurement process and the certainty of having measurements done by a trained and experienced professional, these findings must be confirmed by other listeners. Raters will first be asked to identify prolongations impressionistically from within a selected portion of the data. Then, following a brief training, raters will be asked to measure the duration of prolonged syllables acoustically. Following Fleiss' Kappa Benchmark Scale, interrater reliability in the 75%-100% range will be taken to confirm the researcher's original measurements.

Another shortcoming is the relatively small number of tokens. This will be addressed both by adding to the existing corpus, as well as comparing this peripheral corpus to a prototypical corpus of spoken language annotated for prosody (LeaP). Building upon Gilquin (2008) who applied similar ideas to LOCNEC and LINDSEI, the continued study will instead

survey the disfluent prolongations found in the LeaP corpus to confirm the findings of this study and begin the investigation of prolongations in the speech of speakers of more L1 backgrounds. Additionally, the corpus comparison will provide an avenue for exploring the distribution of prolongations in free or spontaneous speech.

Conclusion

This chapter has shown noticeable differences in the frequency, quality, and distribution of native and non-native prolongations. Native speakers showed a clear preference for prolonging syllable-final vowels in words at clausal boundaries (initial). Non-native speakers, on the other hand, showed some patterning but overall prolonged any segment in any syllable in any sentence position, likely as a result of their need to provide themselves with additional planning time to compensate for areas of low proficiency.

It has been shown that filled pause distributions which deviate from native speaker norms are disruptive to perceptions of non-native fluency (Pinget et al., 2013; Bosker et al., 2013a, 2013b). Since filled pauses and prolongations are frequently classed together, it is to be expected that similar findings would appear in prolongation-focused studies. However, work has yet to be done which controls location and frequency as acoustically manipulated variables to determine how prolongations are perceived in the speech of non-native speakers. Chapter 5 presents the results of one such experiment.

CHAPTER FIVE

PROLONGATIONS AND THE PERCEPTION OF L1 and L2 FLUENCY

This chapter examines the role of prolongations in the perception of L2 fluency (as well as L1 fluency as a necessary corollary) through an experiment within which speech samples are manipulated in order to place prolongations in felicitous and non-felicitous conditions, based on the general findings reported in Chapter 4, as well as previous research which shows that native speakers generally produce fewer disfluencies than non-native speakers (Cucchiarini, Strik, & Boves, 2000).

The following section provides a summary of the background on studies on the perception of L2 fluency and provides support for the methodology used in this experiment.

Background

Fluency is a complex construct with many competing definitions as well as implications for second language acquisition, psycholinguistics, and foreign language teaching. In its most enduring definition, fluency can be conceived of as the ability to formulate and produce speech in a “rapid, smooth, accurate, lucid, and efficient” manner (Lennon, 2000, p. 26). From this definition, Segalowitz (2010) extrapolates three separate but interrelated notions of fluency: *cognitive fluency*, or the speed and ease with which speech planning and production occur; *utterance fluency*, the independent acoustic features which mark (dis)fluency; or *perceived fluency*, which concerns itself with how the listener judges the cognitive fluency of the speaker based on the fluency of their utterances.

Many studies have examined the connection between utterance and perceived fluency for non-native speech, relying on similar methodology which includes first taking exhaustive measures of individual fluency variables in L2 speech, then collecting subjective ratings from

native listeners (typically using a 7- or 9-point Likert scale) and, finally, running correlations to determine which of said fluency variables accounts for the most variance in fluency ratings (Freed, 2000; Cucchiarini et al., 2000, 2002; Riazantseva, 2001; Derwing et al., 2004; Kormos & Denes, 2004; Trofimovich & Baker, 2006; Rossiter, 2009; Kang, 2010). Ratings in such studies have been shown to be reliable regardless of rater background (language teacher, naïve rater, expert rater, NNS rater) level of exposure) (Rossiter, 2009; DeJong et al., 2012), and instructions or task type (Cucchiarini et al., 2000; Kormos & Denes, 2004; Rossiter, 2009).

The goal of such studies is to determine which of these variables most affect fluency ratings, a question which is important both for language testing in general and tests of oral proficiency in particular, both of which heavily rely on perceived fluency. There are also immediate applications for task-based language learning research, which often includes detailed assessments of learners' speech (Skehan & Foster, 1997). Major findings within this body of work indicate that the temporal measures *pruned syllables per second* (the number of syllables in a given speech sample minus repetitions, filled pauses, etc.) (Rossiter, 2009), *pruned syllables per second and pauses* (Derwing et al., 2004), *speech rate* (Kormos & Denes, 2004) and *mean length of run* (Cucchiarini et al., 2002) reliably correlate to fluency judgments.

While these studies do at times place native speech against non-native patterns at the level of acoustic comparison, only a few have carried this research forward to examine the effect of disfluency on the perceived fluency of native speakers (Cucchiarini et al., 2000, Bosker et al., 2014).

Following the definitions set forth by Lennon (1990) and Schmidt (1992) *fluency* may refer to “smooth, accurate, lucid” delivery in either native or non-native speech. Still, it is widely reported that native speakers are thought to be fluent by default (Riggenbach, 1991;

Bosker et al., 2014) and that despite the prevalence of disfluency in the native speech signal (Fox Tree, 1995), these disfluencies are not thought to be disruptive and may even in some instances be helpful in the sense that they signal to the listener that the speaker is having difficulty and attribute this difficulty to the complexity of the upcoming item (Arnold et al., 2007; Watanabe et al., 2008). As such, some work has examined native disfluency in search of clues as to the speech planning process (Goldman-Eisler, 1968; Levelt, 1989).

Though these studies are useful, little work has been done to show which types of disfluencies most affect the perception of fluency of a native speaker by a native listener. Nor, as Bosker et al., (2014) point out, can such a study effectively make use of the same methodology which is commonly used to analyze non-native speech because native and non-native speech differ on so many points. The most effective method, then, for examining how listeners perceive both native and non-native disfluencies is to create an experiment whereby a particular variable (filled pauses, silent pauses, hesitations and, in this case, prolongations) are carefully manipulated and controlled such that the variable under scrutiny is the only main difference between samples.

A series of experiments by Bosker (2014) and Bosker et al. (2013a, 2013b) use this method to examine the effect of different lengths of silent pauses on the perceived fluency of native and non-native speech. The length of silent pauses in the speech samples by native and second language speakers of Dutch was manipulated into LongPause (750-1000 milliseconds), ShortPause (250-500 milliseconds) and NoPause (<150 milliseconds, or the bare minimum of articulatory pause given inspiration/expiration time) conditions. Results show that native speakers were rated as more fluent regardless of condition, with no difference between the three main conditions, results which are in line with previous studies on silent pauses in non-native

speech (Cucchiaroni et al., 2002). However, the speech samples used retained other disfluencies characteristic of non-native speech, such as filled pauses and hesitations, and also did not take into account disparate pause distribution between native and non-native speakers. Finally, Bosker et al. (2014) point out that similar work remains to be done with other types of disfluencies.

Research Questions

The body of research on prolongations is quite limited and in no way as extensive as that dealing with other types of disfluency (Eklund, 2001). As a result, many of the conclusions as to their frequency, duration, location and function are often conflated with data regarding filled and silent pauses (Horvath, 2004; Esposito, 2006).

One example of this occurs with the idea that if prolongations behave like silent and filled pauses they should occur more often at clausal boundaries than within clauses. That prolongations occur more frequently at clausal boundaries was shown in the experiment described in Chapter 4; however, this finding was borne out in the native data set but not within the non-native data, a result which lends preliminary credence to the emerging picture of an NS-NNS differential in prolongation use and distribution. In addition to the clause position distinction, other findings from Chapter 4 include the following:

1. Native speakers and non-native speakers prolong segments at roughly the same rate.
2. Both native and non-native speakers show a preference for prolonging vowels and [+continuant] consonants.
3. Both native and non-native speakers seem to observe a minimum threshold for prolongations of roughly 500 milliseconds.

Taking these observations into account, what remains to be seen is what, if any, effect prolongations have on fluency ratings. Given that the whole of this dissertation has focused on the differences between native and non-native disfluency patterns, the attention now turns to how said patterns affect the perception of fluency for both groups. This chapter addresses the following research questions:

RQ5A: Are native and non-native speakers rated differently on a perceived fluency task?

RQ5B: Does the location of the prolongation affect perceived fluency ratings for native and non-native speakers?

RQ5C: Does the frequency of prolongation use affect perceived fluency ratings for native and non-native speakers?

Predictions

The experiment presented in this chapter studies the effect of prolongations in different conditions on the perceived fluency of native and non-native speakers. As native speakers have been shown to be rated as more fluent than non-native speakers, or to be considered fluent by default, it is expected that this result will hold within this data with native speakers being rated more fluent than non-native speakers.

The case is not as clear-cut for RQ5B and RQ5C.

Clause position

Native speakers are more frequently disfluent at clausal boundaries, at least in the case of silent and filled pauses (Goldman-Eisler, 1968; Anderson-Hsieh & Venkatagiri, 1994). If this pattern extends to prolongations, then it is expected that raters will be sensitive to it and may rate

native speakers more harshly when prolongations do not coincide with clausal boundaries. On the other hand, raters' familiarity with native speech may influence them to rate native speakers as universally more fluent overall, without any distinction for other actors. If this is the case then no difference for native speakers in the clause position condition is expected. Raters may also fail to key into the placement of prolongations in non-native speech because of the general "noisiness" of non-native speech. In this case, no difference for non-native speakers would be expected either.

Frequency and interactions

Because disfluencies have been shown to occur more frequently in non-native than native speech (Cucchiaroni et al., 2000), a general prediction is that items in the high frequency condition will be rated as less fluent than low frequency items for both native and non-native speakers. However, the data presented in Chapter 4 indicates that the frequency of prolongations is similar for both native and non-native speakers. This could lead to native raters essentially dismissing frequency as a factor, instead choosing to focus on overall native-likeness (in which case there would be a robust NS/NNS difference) or the position of the prolongation within the sentence, in which case non-native items which had native-like prolongation patterns (clause-initial) should be rated as more fluent than those with clause non-initial prolongations. Finally, native raters may rely on their expectations of native speakers' disfluency patterns to guide their judgments of non-native fluency—in which case there should be an interaction between clause position *and* frequency.

Methods

Participants

Raters for this experiment were 100 native speakers of English. All participants were university students between the ages of 18 and 25³. Recruitment and participation was done via a panel within Qualtrics, a professional survey platform.

Stimuli

Sample selection

Speech samples from the narrative tasks in the corpus reported in chapter 3 were used as a basis for stimuli creation. This allowed for topic control and, to a lesser extent, control of vocabulary range and register. Native and non-native raw stimuli were taken from the central portion of a 60-second speech sample; that is, the extracted sample began at least ten seconds after the start of the narrative/speech event and ended with at least ten seconds remaining. This was done to avoid “extra” disfluency due to false starts and any awkward phrasing or pacing as speakers were closing the narrative.

In order to avoid the possibility that fluency ratings could be related to a particular accent or that raters could become acclimated to a particular accent over time, speech samples were taken from speakers from four different L1's (Chinese=2; Turkish=1; Spanish=1; Korean=2). Three male and three female speakers from higher levels were selected to minimize interference from segmental pronunciation issues and, to the extent it was possible to do so, minimize the

³ The protocol for this study was approved by the Internal Review Board at the University of Texas at Arlington (#2013-0865).

distance (and, therefore, the dissonance) between native and non-native samples. For each speaker, 2 speech samples were selected for manipulation.

Native samples were also taken from the larger corpus reported in Chapter 3. Because native speakers spoke more quickly only a small number reached the necessary 60-second threshold to be considered for stimuli creation. Two speakers (1 female, 1 male) were selected. Again, 2 speech samples per speaker were selected. All together, 16 samples (12 NNS, 4 NS) were selected.

Manipulations

The first pass of manipulations on the stimuli removed silent and filled pauses in an attempt to minimize possible interference from other forms of disfluencies. First, all long silent pauses (>250 milliseconds) were trimmed to 250 milliseconds (following DeJong & Bosker, 2013). Pauses which were already below this measure were not altered, as they were considered to be wholly articulatory pauses and modifying them in any way would affect the naturalness of the overall utterance. Second, because clusters of disfluencies are more common in non-native speech than in native speech (Deme, 2013), filled pauses were removed altogether to avoid potential cumulative effects from prolongations and filled pauses. Finally, instances of repetitions, hesitations, and self-corrections were removed. All first pass manipulations were done using Praat (Boersma & Weenink, 2015).

The second pass of manipulations involved the manipulation of segments to create 4 different stimuli from each speech sample, for a total of 64 stimuli that varied by frequency and placement of prolongations. The manipulations were meant to approximate as closely as possible the observed characteristics of naturalistic prolongations (see Chapter 4). Naturalness was the

ultimate aim; to that effect, the following ranked criteria were observed in deciding which words/segments to prolong.

1. Position

The highest priority was preserving the integrity of the clause initial/non-initial condition. In cases where another lower criteria (number of syllables, vowel choice, function/content) conflicted with position, the syllable which best preserved the condition was always selected.

2. Monosyllabic over di/multi-syllabic

Words with a single syllable were prolonged almost to the exclusion of all other word types within the corpus. Therefore, this is the second-order constraint. However, given the brevity of the speech samples this occasionally had to be sacrificed to preserve position.

3. Frequently-prolonged vowels over obscure vowels

The most frequently-prolonged vowel in the corpus was [i]. Whenever possible, this was the vowel chosen for prolongation. Other frequent vowels include [ɪ], [æ], [ɛ], and [ə].

4. Function over content words

While function words were preferred in that they are much more frequently prolonged than content words in both native and non-native speech, this criteria was ranked low as often the boundary constraints limited the options for words to prolong.

Conditions

Frequency and clause position conditions were calculated based on major clausal boundaries. In the placement condition, either every clause-initial syllable (i.e., immediately

following a clausal boundary) was prolonged (*initial*), or one syllable between every pair of clausal boundaries was prolonged (*non-initial*). The frequency condition was realized as all-or-1; in other words, in high-frequency stimuli, every syllable that immediately followed a clausal boundary was lengthened. In the low frequency stimuli, only the first such available syllable was lengthened. Strictly speaking, the low condition could apply at or within any of the boundaries. However, the single instance of prolongation was placed as early as possible in the item to preclude the possibility that a listener might—in not hearing a prolongation early in the item—rate the item as more fluent.

Figure 4 gives an example of a sample stimuli with clausal boundaries marked by forward slashes. Items targeted for prolongations are bolded.

Original Sample:
<p>“The fourth factor related to woman’s stress is her car crash. The fifth factors affected woman’s stress is her bills, and lastly woman becomes stressed”</p>
High/Initial: Every bound syllable gets lengthened
<p>The fourth factor related to woman’s stress is her car crash/ The fifth factors affected woman’s stress is her bills/ and lastly woman becomes stressed</p>
Low/Initial: One bound syllable gets lengthened
<p>The fourth factor related to woman’s stress is her car crash/ The fifth factors affected woman’s stress is her bills/ and lastly woman becomes stressed</p>
High/Non-Initial: One syllable between every boundary to boundary marker is lengthened
<p>The fourth factor related to woman’s stress is her car crash/ The fifth factors affected woman’s stress is her bills/ and lastly woman becomes stressed</p>
Low/Non-Initial: A syllable is lengthened in one segment between two boundaries
<p>The fourth factor related to woman’s stress is her car crash/ The fifth factors affected woman’s stress is her bills/ and lastly woman becomes stressed</p>

Figure 4. Sample item by condition

Manipulations--Prolongations

Once words were selected for prolongation, vowels within the targeted words were lengthened using Apple Logic Pro X software. In order to most closely match the natural prolongation data found in the corpus, the target duration for each prolonged word was set at 500 milliseconds. Wherever needed, pitch contours were leveled or readjusted such that there was no unnatural break between the manipulated segment and the immediately following sections of the utterance. At the tail end of each audio file a volume fade was inserted to provide a more natural end to each speech sample. Finally, individual items were normalized for intensity and mastered for sound quality using iZotope Ozone 7 software. All manipulations were performed by a trained sound engineer directly supervised and instructed by the principal investigator. Both parties had to agree that an item both sounded natural and also met all prolongation-based conditions. In instances where this was not the case, target word manipulations were either corrected (prolongations adjusted for duration while maintaining the 500 millisecond limit) or in cases where no adjustment was possible, utterances were discarded and replaced with other utterances from the same speaker and speech sample. The average length of manipulated stimuli was 10.33 seconds for non-native speech samples and 10 seconds for native speech samples.

Experimental Design

As described above, two samples were taken from the narrative tasks performed by the 8 selected speakers, for a total of 16 raw files (NNS=12; NS=4). Each of the 16 files were manipulated such that they resulted in four experimental items reflecting combinations of the two main conditions Clause (initial, non-initial) and frequency (high, low). These were: High/Initial (HI); High/Non-Initial (HNI); Low/Initial (LI); and Low/Non-Initial (LNI), for a total of 64 working stimuli.

Manipulated items were loaded into the Qualtrics online survey system as 4 counterbalanced blocks of 16 items. Blocks were arranged such that raters might hear the same speaker twice, but they would not hear the same speaker and utterance in more than one condition. Each block contained 16 items representing all conditions (3 items in 4 conditions=12 NNS items; 1 item in 4 conditions=4 NS items) and non-native items in all conditions. Items within each block were randomized for every trial.

Procedures

At the beginning of their experimental session, participants were required to read and click indicating their agreement to the informed consent document (ICD). Participants were then asked to indicate whether they were native speakers of English. Immediately thereafter, participants were instructed in the criteria on which to base their judgments of fluency. First, they were given the general definition of fluency (based on Lennon's *broad fluency*), and then were shown an alternate working definition of fluency taking into account the ease of delivery and the fluidity of the speech sample, including pauses, hesitations, and words that are longer than expected. These instructions were presented as a single block of text (Figure 5 below) which was shown to participants throughout the experiment.

Fluency is often defined as the ability to speak a language well.

However, for this experiment, please rate the following speech samples for THIS alternate definition of fluency:

- 1) The ease with which the speaker conveys his or her ideas
- 2) How fluid the speech sample is (pay attention to things like pauses, hesitations, and words that are longer than expected)

Figure 5. Instructions for rating fluency

All rating for this experiment was done using a 7-point Likert scale, consistent with those used by Derwing et al. (2004), Trofimovich & Isaacs (2012) and other L2 fluency studies. Poles were marked “fluent” (right) and “not fluent” (left), with no markings on the five points between. Participants were asked to click on the point of their choosing to record their rating.

Prior to being shown experimental items, participants were given one test item to familiarize them with the concept of fluency. In this item they were given a comparative task with two non-native speech samples in which they were asked to identify which one was more fluent. The following two test items were rating tasks used to familiarize participants with the Likert-scale task, one containing a native speech sample, and one containing a non-native speech sample. Following these practice items participants saw an attention-check question which asked them to point out the non-absurd answer out of a list of 8 alternatives. This question was included once immediately after the instruction and test portion to minimize the incidence of participants who might simply click the same score throughout without attention to the audio prompt.

Participants then saw the remaining test items in succession. Item presentation was timed such that the software disallowed entering a response until 10 seconds (the average length of the stimuli) had passed. This safeguard was meant to prevent participants from rating a sample having heard only a portion thereof, and also to provide enough time for raters to consider the definition of fluency they were given at the beginning of the experiment. After listening to the entire sample, participants’ responses triggered the experiment to move to the next item. Average total participation time was approximately 10 minutes. As the experiment was relatively brief, participants were not given a break between items or groups of items.

Analysis

All analyses for this experiment were conducted using the scores participants provided on a 7-point Likert scale. It was found that several participants had identical answers for all sixteen items while others had a large percentage of one score. It is assumed in those cases that the participant was either not being attentive to the prompts or simply did not understand the instructions sufficiently to respond. Any participant who rated 10 or more (62.5% or above) of the items exactly the same was excluded from analysis. Implementing this standard excluded 13 raters from analysis, resulting in a final total of 87 participants.

It was also noted that many of those remaining participants exhibited a strong preference for certain numbers or groups of numbers. This is not a novel difficulty with the use of Likert scales. It is often the case that some participants will cling to the poles of the scale while others prefer to hedge their answers and keep to central (or neutral) numbers. To correct for this, a z-score calculation was done which normalized individual scores to a uniform measure.

The z-score was calculated by first subtracting the mean of the individual participant's responses from each item, and then dividing this figure by the standard deviation for the participant's responses. Applying this calculation to all data points meant that those responses which were average now have a z-score of 0, numbers below 0 (negative values) are below average, and those above 0 are above average for that particular rater.

Because of the low number of items in this experiment, raw means of z-scores and descriptive statistics will be used in the discussion of results.

Results

Predictably, a global mean of z-scores showed a noticeable difference between scores for native and non-native speakers, with the native speakers rated as more fluent overall than non-native speakers. As shown in Figure 6 below, the mean for native speakers was .45 ($SD=.97$), while the mean for non-native speakers was -0.15 ($SD=.92$):

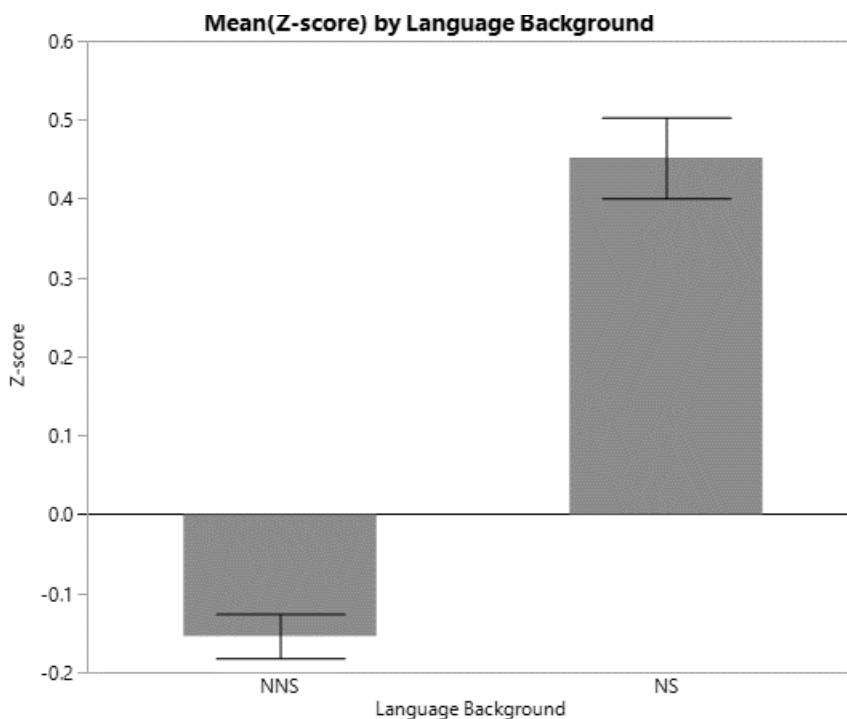


Figure 6. Global means of z-score by language background

There was a noticeable difference for native speakers but not non-native speakers in the clause position condition, as shown in Figure 7 below. The mean of native speech samples with prolongations in clause initial position was .34 ($SD=1.02$) while the mean ratings for native stimuli with prolongations in non-initial position was .56 ($SD=.91$). This result is surprising

given that much of the literature on silent and filled pauses reports better fluency ratings for speech samples with clause-initial pauses (Zahng, 2014).

While Figure 7 shows a very slight difference between initial and non-initial conditions for non-native speakers, this effect is minimal: Initial ($M= -0.17, SD=.89$), Non-initial ($M= -0.14, SD=.96$).

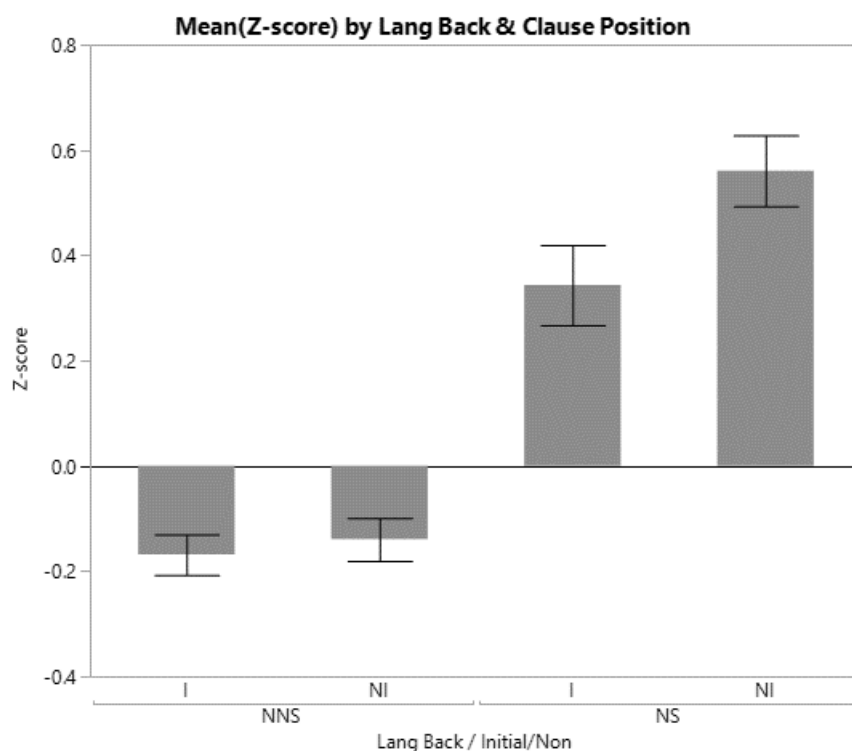


Figure 7. Means of z-score of clause position within language background

The dispreference for clause-initial prolongations in native speech was true in both high and low frequency conditions, as shown in Figure 8 below (High-Initial: $M=.39, SD=1$; High-Non-intial: $M=.29, SD=1$). In other words, participants rated non-native speech samples as less fluent if prolongation appeared immediately after a clausal break, whether they heard one or multiple prolongations clause-initially. This is surprising because previous literature suggests

native speakers are more likely to have disfluencies at clausal boundaries; thus, native-speaking raters should be more tolerant of clause-initial prolongations. Instead, these results suggest the opposite. In addition, the disregard for frequency information is also an unexpected observation given that, in general, speakers are considered less fluent when they have more disfluencies and slower speech rate, both effects which result from having multiple prolongations in a single brief speech sample.

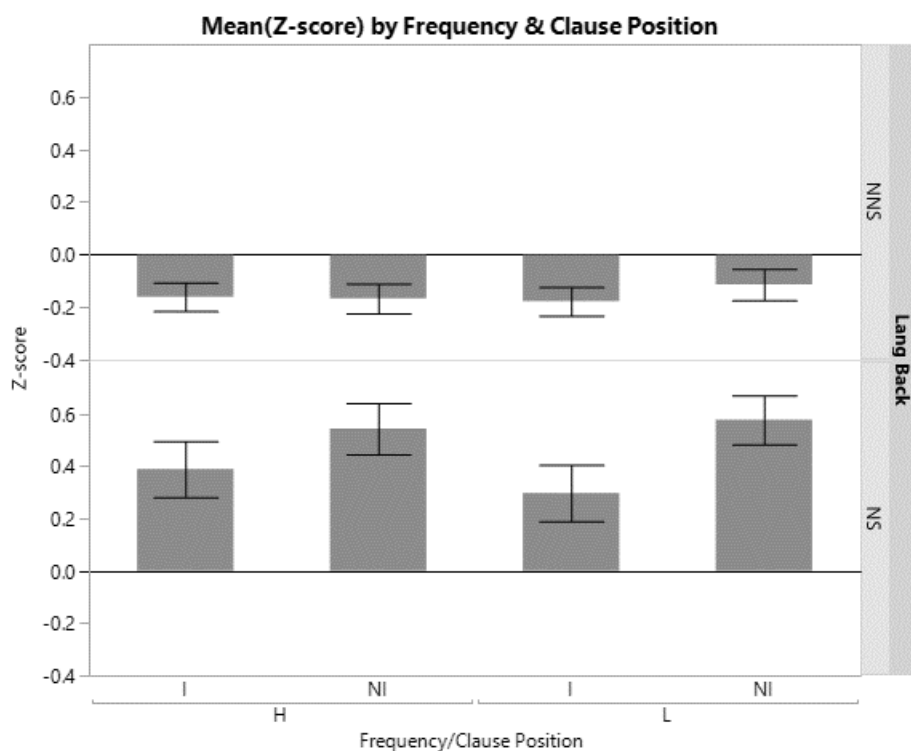


Figure 8. Means of z-score across language background, clause position, and frequency

However, the difference in means of z-scores by condition for non-native speech was negligible (as shown in Table 20 below).

Table 20. *Mean and Standard Deviation of Z-scores across frequency and clause position in Non-native speech samples*

Condition	M	SD
Frequency: High	-0.16	.91
Frequency: Low	-0.15	.94
Clause Position: Initial	-0.16	.89
Clause Position: Non-initial	-0.14	.96

Overall, results indicate that raters in this experiment were sensitive to the placement of prolongations when they heard native speech samples but not when they heard non-native speech. It should be noted that due to the limited number of items in this experiment the results reported above are not generalizable to the population at large. Raters with similar backgrounds are expected to respond similarly to the existing items, but new items may expand the results in unexpected ways.

For the results at hand, the lack of noticeable difference in fluency ratings for non-native speech samples in different conditions raises a number of questions, which will be addressed in greater detail in the following section.

Discussion

This chapter applied the results of the survey of prolongation distribution in Chapter 4 to a perceived fluency experiment which combined manipulated stimuli and a fluency rating task based on a Likert scale. Because comparing native and non-native disfluency patterns using the traditional correlational method used for rating the fluency of non-native speech native speech is ineffective (Bosker, 2014), a method was used in which the number and placement of prolongations were kept constant in speech samples of both native and non-native speakers. Though the utterances themselves were not identical, the structure and lexical complexity of the utterances for both native and non-native speakers were kept restricted by the topic of the narrative task. Further, by removing filled pauses and normalizing silent pause length, the main

temporal variables which affect fluency ratings were removed. It was therefore possible to examine the effect of prolongations on perceived fluency ratings in a controlled environment which still contained natural speech samples. By taking this approach it was possible to measure: 1) how listeners rate the fluency of native and non-native speakers, and 2) how the placement and frequency of prolongations affect these ratings.

Results showed that raters perceived native speakers as being more fluent overall than non-native speakers even when they used frequent and misplaced prolongations. This is in line with what other studies have shown with regards to perceived fluency of native speech. However, one interesting difference emerged in that raters showed more sensitivity to native speakers' disfluencies than they did to non-native speakers'. The greater range of scores in ratings of fluency for native speech samples indicates that listeners are not only able to judge native fluency along a narrow fluency dimension but that native listeners have ingrained expectations about native speech which, when violated, are less tolerable than non-native disfluencies. This contradicts findings in Bosker (2014) which reported that in the case of silent and filled pauses raters weight native and non-native fluency equally and that raters make no qualitative distinction between native and non-native speech when evaluating fluency. Nor is it the case in this study that raters assume native fluency by default, as has been widely reported (Davies, 2003).

Most importantly, this study has shown that raters are sensitive to the placement of prolongations even when all other disfluencies are removed (filled pauses, hesitations, repetitions, self-corrections) or controlled (silent pauses). It is at least a preliminary indication that native listeners have expectations about the placement and use of prolongations which

function independently from their expectations for filled pauses and silent pauses, lending support to the need for disentangling the study of prolongations from that of other disfluencies.

This study also showed that raters dispreferred prolongations in clause-initial positions in native speech samples, a finding which contradicts previous work with filled and silent pauses which has shown that speech rated as disfluent had more instances of silent pauses within a clause compared to L1 speech (Tavakoli, 2001; Bosker, 2013; Kahng, 2014). Related to this, raters did not seem to be sensitive to prolongation frequency or clause position when rating non-native speech samples. This may have occurred as a result of a tendency for raters to make a binary native/non-native distinction early in the utterance. For example, it may have been the case that once a rater determined a sample was from a non-native speaker, they simply gave that sample a uniformly low score to indicate its non-nativeness, reserving distinctions between preferred and dispreferred disfluency patterns for native speech items. Nevertheless, both results are unexpected and may have been influenced by one of several factors.

First, results may have been affected by the confounding of the clause position and frequency conditions. For non-native samples this may have meant that the overall effect of “non-nativeness” arising from differences at the segmental and grammatical levels may have washed out the effects of prolongation placement. For native samples, it may have been the case that when prolongations were most frequent the effect of clause position was effectively lost in the noise of the data, an unlikely interpretation given that there appeared to be no difference between frequency conditions as long as the prolongations appeared in non-initial position. In other words, raters were just as likely to rate an utterance less fluent if it had one prolongation or many prolongations.

In theory the issue of condition conflation could be addressed by studying frequency and position separately; however, closer inspection reveals challenges to doing so. For example, if we wanted to examine *only* prolongation frequency (many vs. some or some vs. none), certain words would still need to be targeted for lengthening. Where these words occur within the utterance may be more or less natural depending on the clause position, and thus some account of clause position would be required. If we were to choose clause position as sole variable, we would still need to decide if we would target every clause or only some clauses, thus introducing an element of frequency into the equation.

Clause position may also be impacting results in other ways. While the items in this study were necessarily quite brief, even within these short utterances there were various levels of syntactic boundaries. As an introductory study, this experiment only targeted main clausal boundaries; however, it may be the case that phrasal categories not taken into account (verb phrases, prepositional phrases, etc.) may be natural targets for prolongation. The hierarchy of target selection described in the Stimuli section prioritized the preservation of clause position condition. However, an unintended consequence of this was that, when combined with the preference for targeting function over content words, this placed many prolongations at phrasal boundaries. In doing so, raters may have perceived multiple non-initial prolongations (which happened to fall at phrasal boundaries) as more fluent than a single clause-initial prolongation. In order to test this, future work will use longer speech samples and test the acceptability of prolongations at different syntactic boundaries.

Furthermore, this study also has several shortcomings which will be addressed in future work, the most immediate of which is the need for more and longer test items. Another limitation is the uneven number of native and non-native speech samples which curtails the scope of

interpretation and the generalizability of results. Third, having tested only short speech samples means that the high-frequency condition may have simply forced too many prolongations into a single utterance. To correct for this, future studies will use longer speech samples (30-60 seconds) from both native and non-native speakers who will be coached on producing prolongations at the targeted locations, thus eliminating the need for external manipulation and thereby increasing the naturalness of intonation and pacing across the utterance while still holding the placement and frequency of prolongations constant. This will also permit greater control within a single variable such as prolongation duration or placement.

Follow-up studies will also specifically test the place of prolongations in the disfluency canon. Distributional surveys have shown some support for the treatment of prolongations as separate from filled pauses (Deme, 2013); however, this has not been tested experimentally. This study has shown that raters are sensitive to prolongations in native speech even when all other disfluencies are removed. A further extension of this test would be to place prolongations and filled pauses of identical durations in equivalent locations in the utterance to see if raters are more sensitive to one over the other. A similar investigation can be done with both disfluency types co-occurring within an utterance to test whether sensitivity to prolongations persists when other disfluencies are present but held constant while prolongations alone are manipulated.

CHAPTER SIX

CONCLUSIONS AND APPLICATIONS

Disfluencies are common features of both native and non-native speech. The study of disfluencies has had an important role in the field of second language acquisition from learner corpus research to evaluative studies of second language fluency and cognitive studies of speech planning and perception. Some disfluency types have received a great deal of attention in the literature (Lennon, 1990; Towell et al., 1996; Cucchiarini et al., 2000, 2002; Kormos & Denes, 2004; Ginther et al., 2010; Bosker, 2013). Prolongations, however, have remained what Eklund (2001) termed the “dark horse of the disfluency stable”.

This persistent gap in the literature means that there are limits to our understanding of narrow fluency (Lennon, 1990). Though prolongations are a measurable feature of utterance fluency, there remained a simple lack of data to build the knowledge base of how this aspect of utterance fluency contributes to the perception of fluency in native and non-native speech.

To that end, the goal of this dissertation has been to examine the following research questions:

RQ 1: In what ways do measures of utterance fluency related to prolongations differ from native to non-native speakers?

RQ 2: To what extent do prolongations affect the perceived fluency of native and non-native speech?

Summary of Results

To address these questions, first a corpus of learner speech was compiled and analyzed, followed by an investigation of the distribution of prolongations in both native and non-native speech. Finally, fluency ratings were collected for native and non-native speech samples which

included prolongations manipulated by two conditions: frequency (high, low) and clause position (initial, non-initial).

Chapter 3 presented the process and product of a corpus of learner speech representing a variety of L1 backgrounds and contrasting the disfluency patterns therein with a select group of native controls. Speech samples from a narrative task were measured for nine fluency variables in the three dimensions of fluency (speed, breakdown and repair) as determined by Skehan, 2009) and in accordance with methodology commonly used in the field (Gilquin, 2008; Gut, 2012). Specific measures included mean length of run, speech rate, mean length of syllable, mean length of silent pause, and rates of silent pauses, filled pauses, repetitions, self-corrections, and prolongations. As expected, it was shown that native speakers were significantly faster and exhibited fewer disfluencies across all nine measures, confirming previous findings in similar studies (Lennon, 1990; Riggensbach, 1991; Derwing et al., 2004, among others).

In terms of the less-studied hesitation phenomena (repetitions, self-corrections, and prolongations) it was apparent that prolongations behaved differently in that they were the second most common disfluency, running a very close second to filled pauses in both native and non-native speech. Unlike filled pauses, however, where non-native speakers outpaced native speakers in both frequency of use and average duration, prolongations were similar for native and non-native speakers on both counts when all data was examined together.

Chapter 4 examined the distribution of prolongations in native and non-native speech. Taking a similar approach to those in previous studies on L1 prolongations (Eklund, 2001; Lee et al., 2004), this chapter provided a much-needed account of the ways in which native and non-native distributions of prolongations vary. Because of variability in hesitation use between subjects, only data from those subjects (both native and non-native) who used prolongations was

examined. It was shown that native speakers used had fewer prolongations than non-native speakers when engaged in an identical task, and that those prolongations differed in their placement within the sentence. Native speakers are more likely to prolong at major clause boundaries while non-native speakers prolong roughly equally at and between clausal boundaries. This finding points to the notion that prolongations, like filled pauses, are vocalizations used to buy the speaker additional planning time and while native speakers generally need such boosts to access complex lexical items or grammatical forms, non-native speakers may use the same strategy to compensate for general lack of proficiency, whether syntactic, lexical, or narrative.

In addition, the study of prolongations in Chapter 4 showed that both native and non-native speakers shared a preference for prolonging vowels and continuants, but that the individual segments varied based on position within the syllable. More work is needed to determine whether this is coincidental or a function of the word type frequency. Finally, it was shown that function words were more likely to be prolonged in both native and non-native speech, but that this effect was far more robust for native speakers than non-native.

Chapter 5 applied the results of the survey of prolongation distribution in chapter 4 to a perceived fluency experiment which combined manipulated stimuli and a fluency rating task based on a 7-point Likert scale. Because comparing native and non-native disfluency patterns using the traditional correlational method used for rating the fluency of non-native speech native speech is ineffective (Bosker, 2014), a method was used in which the number and placement of prolongations were kept constant in speech samples of both native and non-native speakers. Though the utterances themselves were not identical, the structure and lexical complexity of the utterances for both native and non-native speakers were kept restricted by the topic of the

narrative task. In doing so, it was possible to measure: 1) how listeners rate the fluency of native and non-native speakers, and 2) how the placement and frequency of prolongations affect these ratings.

Results showed that native speakers were judged to be more fluent overall than non-native speakers, a result which is in line with other similar research on silent and filled pauses (Riggenbach, 1991; Cucchiaroni et al., 2000, Derwing et al, 2004). However, raters demonstrated a sensitivity to the position of prolongations within the sentence for native speakers which they did not show for non-native speakers. Native speech with prolongations in the initial condition (at clausal boundaries) were perceived to be *less* fluent than native speech with prolongations within the clause (non-initial). This finding demonstrates that listeners are able to rate the fluency of native speakers and that they are not simply considered fluent “by default” as has been widely reported in the literature (Raupach, 1983; Riggenbach, 1991). More importantly, the study showed that raters were sensitive to the placement of prolongations even when all other disfluencies were removed, showing that native speakers have expectations for prolongation patterns which are independent of those for filled and silent pauses.

Applications

Foreign language teaching and testing

For the past decades the notion of communicative competence has been the theoretical foundation for most foreign language teaching and testing. First put forth by Dell Hymes (1974) as the speaker’s awareness of the social context for language, it has since been further refined by Canale and Swain (1980), who established grammatical, pragmatic, and strategic competencies as the pillars of overall communicative competence. Traditional language teaching focuses on grammatical competence, the rules, forms, and structures which govern the grammar, and

pragmatic competence, or how language functions in its broader interactional and sociocultural contexts (Canale & Swain, 1980; Dörnyei & Thurrell, 1991).

Strategic competence is the ability to get one's meaning across even in the face of communication challenges brought on by factors such as insufficient knowledge or mastery of the language. The idea is that, although communication difficulties can occur in both first and second language use, speakers instinctively use appropriate strategies when encountering communication difficulties in their first language. In a second language, however, where the difficulties of communication are increased and communication breakdown is virtually assured, language learners should be taught strategies for how to deal with such challenges as they arise.

Recently, explicit classroom instruction in strategic competence has been gaining ground after being overlooked for many years in favor of grammatical and pragmatic competences (Rieger, 2003; Derwing et al., 2004). Examples of strategies which can be taught and practiced in the foreign language classroom include circumlocution and paraphrasing, use of language chunks, and using accepted hesitation patterns such as filled pauses to buy additional speech planning time (Rubin, 1987; Dörnyei & Thurrell, 1991).

Prolongations are as frequent in native speech as filled pauses (as seen in Chapter 3). However, explicit instruction in the appropriate use of prolongations in building strategic competence is lacking. On the whole, syllable lengthening may be a *more* preferable hesitation strategy. Use of L1 fillers in L2 speech has been linked to lower fluency ratings (Matthews, 2014). Filled pauses vary from language to language, and learners have to first be taught and then consciously choose to relinquish their native fillers and use English-based fillers instead (Rose, 2008).

Prolonging syllables, on the other hand, is common across languages and provided that the prolongation is occurring at a preferred position within the sentence, there should be no expected difference between the native and non-native strategies.

In addition to being taught how to use prolongations as a hesitation strategy, it is important that language learners be taught how to recognize them in the speech stream of native speakers, as they can often cause confusion for the non-native listener (Reed, 2000; Griffiths, 1991). While all languages contain hesitations in various forms, hesitations in the mother tongue can more easily and automatically be filtered out, whereas in a second language the added burden of processing hesitations may lead to confusion. Learners should be given opportunities to practice with authentic materials as they learn to process and filter out hesitations in a second language. As the first point of contact between the language learner and the target language, language teachers should also be made aware of their own use of hesitations so that they do not inadvertently create barriers to comprehension and can teach more effectively (Gilquin & DeCock, 2005).

Classroom application and methodology

In order to build strategic competence with prolongations, they must be incorporated into language teaching materials. In addition to inclusion in formal textbooks (a lengthy process at best), prolongations may be taught via authentic examples from native and non-native speech. Learner corpora have been used in the past to expose learners to the difference between native and non-native speech patterns. This is not dissimilar to “noticing the gap,” first proposed by Schmidt and Frota (1986) and since operationalized in the Noticing Hypothesis (Schmidt 1990, 1993). Within this construct, learners must first notice the difference in order to turn input into intake (Gass & Varonis, 1984). This is particularly true for features of language which have

slight but important differences between L1 and L2, since the wholesale transfer of L1 patterns to the L2 could cause errors to arise. Hesitation patterns are just one such phenomena (Dörnyei & Thurrell, 1991), and should be brought to the attention of the learner.

A first step, therefore, is to allow learners access to samples of native and non-native speech and directing them to focus their attention on syllable length and where (within the sentence) and to what extent prolongations are being used. This method, which Seidlhofer (2002) terms *learning-driven data*, has proven to be effective in helping students to notice gaps between their performance and native norms. Moreover, by giving learners access to real-world data, the language teacher can also draw attention to the various ways in which prolongations are used functionally, thereby also contributing to an increase in pragmatic competence.

Strategic competence is particularly important in ensuring learner success in language testing. Since language strategies are primarily used for moments when communication is difficult, they are crucial skills to have when entering a testing environment, since failure to quickly and smoothly address problems with speech planning or comprehension can cause test-takers to panic and for communication to break down entirely. Less severe, but equally as problematic, is the issue that current large-scale language proficiency tests such as the TOEFL and IELTS rest heavily on patterns of hesitation (disfluencies) in judging the overall proficiency (fluency) of the speaker. One recent anecdote even suggests that learners who are otherwise grammatically and pragmatically proficient may receive low test scores as a result of disfluent speech patterns.

Explicit instruction in hesitation strategies—which should include when, where, and which segments to prolong—can go far in helping learners to build the confidence needed to approach high-stress testing environments.

From the test-makers' point of view, there is also the need for a better understanding of how prolongations operate. TOEFL iBT® raters, for instance, are asked to rate test-takers on the overall oral delivery skills, which include pacing and hesitation—both of which are impacted by the use of prolongations. A more complete understanding of how prolongations are distributed in non-native speech can help researchers and test developers refine the descriptions of fluency in order to provide a more representative measure of L2 proficiency (Götz, 2007).

Speech perception

All studies of L2 fluency presuppose that utterance fluency is a surface-level reflection of cognitive fluency, following the model set forth by Segalowitz (2000). That is, regardless of its form (pause, filled pause, prolongation, repetition or hesitation), a disfluency is merely a red flag that some disruption to the speech planning process has occurred.

Non-native disfluencies are almost universally thought to be disruptive to the listener, while disfluencies in native speech trigger prediction and attentional effects which are actually helpful for the listener (Watanabe et al., 2008, Collard, 2009). This is not the case for non-native speech, likely because listeners recognize that disfluencies in non-native speech are somewhat unpredictable or at least do not conform to the native model (Bosker, 2014). What is not yet clear is to what extent these attentional effects are a function of the disfluency itself—in other words, do the effects disappear when native speakers are listening to non-native speakers because their disfluencies are placed in unexpected locations, or because they are perceived as non-native speakers and all expectations, judgments and attentional effects are therefore suspended. While this study did not directly address this question, by placing prolongations at identical spots in brief, similar utterances, the results showed in a very preliminary way that listeners in this study made no distinctions between non-native disfluency patterns in different conditions. That is to

say, once a speaker was judged to be non-native, nuanced gradation of fluency ratings simply ceased. If this effect bears itself out in follow-up studies, it would indicate that attentional effects are not a function of the disfluency, but a function of the language background of the speaker, lending support to an argument put forth by Pinget et al., (2014) and Bosker et al. (2014).

Further, and perhaps more importantly, though prolongations have been classed as disfluencies and disfluencies have been shown to trigger attentional effects in native listeners, prolongations alone have not been previously studied to see if they, too, give rise to such effects. A necessary follow-up to future studies of prolongations would be to test whether native speakers attend to prolongations in a way that is similar to the attentional effects of filled pauses. Given that both are vocalizations used to express hesitation, it is expected that these results will prove to be analogous.

Lastly, it has been established that types and patterns of filled pauses vary somewhat from language to language (Matthews, 2014). Prolongation patterns, on the other hand, are much more similar than they are dissimilar cross-linguistically (Eklund, 2001; Deme, 2013). Since this is the case, an intriguing contribution of prolongations to the field of speech perception would be to test (for the first time, to my knowledge) whether attentional effects are triggered in non-native listeners when hearing native prolongations.

In all, the findings in this dissertation add to the conversation surrounding native and non-native disfluencies in several ways. First, it is clear that raters are able to judge the fluency of native speakers on a gradient scale, thus demonstrating that native speakers are not considered fluent simply for being native speakers. Second, the sensitivity to the clause position for native speakers alone indicates that listeners are more strongly affected when their expectations of native disfluency patterns are not met than when they encounter even heavily disfluent non-

native speech, a result which speaks to the elasticity of the listener's mental perceptual model. Finally, results in Chapter 5 point to the fact that the placement of the prolongations in native speech affected perceived fluency more than the quantity of prolongations present even in a very brief speech sample. This may indicate that the listeners' ability to cope with disfluency in non-native speech does not hinge on being able to parse entire clauses at a time.

Directions for future research

Corpus

As with any growing corpus of learner research, the immediate next step is to expand the base of learners from which the corpus draws in order to also broaden the scope of research into prolongations (and other disfluency features). As the corpus continues to grow, the samples therein will be compared to a broader spoken corpus of learner speech, such as the Learning Prosody in a Foreign Language (LeaP) corpus. The LeaP corpus was excluded for use in this dissertation on the basis of its span and the fact that prolongations (termed *lengthened segments* in the annotation scheme) were both not easily measurable. Additionally, no distinction was made between segments which were lengthened because of disfluency and those which were lengthened as a result of emphasis or end-of-turn. In order to better use such a corpus, a scheme must be proposed which can readily distinguish between the different functions of prolongation.

By conceptualizing and implementing a search function which allows for the distinction between functional and disfluent prolongations, the LeaP corpus would become a viable and valuable tool in furthering this study. Use of the LeaP corpus would allow for a more thorough accounting of the distribution of prolongations across a larger base of L1 backgrounds, as well as some initial exploration into how prolongations factor differ by task type and register.

Prolongations

Related to and based on the proposed extensions to the corpus in the previous section, further research needs to be done to gain a fuller picture of prolongation distribution across many L1 backgrounds. In order to do so first the corpus must first reach a representative number of speakers from the different L1 backgrounds (at this time, Arabic, Turkish, Spanish, Vietnamese, Korean and Chinese) at equivalent levels. From there, analysis into the distribution of prolongations can reveal possible differences by language group and propose potential correlations between prolongation use in the L1 and the L2.

Also, a more thorough account needs to be made of individual differences in prolongation use. Previous research has shown that individuals can differ significantly in their usage of different disfluency types (Fillmore, 1979; Gilquin, 2008), though these studies did not include prolongations as a disfluency under consideration. Nonetheless, it is anticipated that (as with filled pauses), speakers will either show preference for certain disfluencies over others either as a result of L1 transfer (Vural, 2008; Matthews, 2014) or personal disfluency habits irrespective of L1, or may opt to omit the use of prolongations altogether (Rohr, 2015).

Conclusion

This dissertation examined the contribution of *prolongations* to the utterance and perceived fluency of native and non-native speech. Though prolongations have long been studied as a function features of pragmatics and phonology, they have previously been understudied as disfluencies or hesitation types.

This dissertation was a modest but much-needed first step toward a better understanding of the distribution, function, and perceptual effects of prolongations. The results of three main

studies showed that 1) prolongations are as frequent as filled pauses in native and non-native speech (Chapter 3); 2) non-native speakers prolong more frequently (as a percentage of total syllables spoken) and less predictably than native speakers (Chapter 4); and 3) that native raters are indeed sensitive to prolongations even when all other disfluencies are removed but, at least in this preliminary experiment, appear to only be sensitive to differences in the placement and frequency of prolongations in native speech.

These findings open the door for further investigation of prolongations with the aim that by continuing to pursue a better understanding of prolongations and their effects in the perception of native and non-native speech they will, in time, be the dark horse of the disfluency stable no more.

REFERENCES

- Altenberg, B. (2002). Using bilingual corpus evidence in learner corpus research. *Computer learner corpora, second language acquisition and foreign language teaching*, 37-54.
- Anderson-Hsieh, J., & Venkatagiri, H. (1994). Syllable duration and pausing in the speech of Chinese ESL speakers. *TESOL Quarterly*, 28, 807-812.
- Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: the on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), 914.
- Bell, Linda, Eklund, Robert, Gustafson, Joakim. (2000). A comparison of disfluency distribution in a unimodal and multimodal speech interface. *Proceedings of the 6th ICSLP 3*, 626-629.
- Biber, D., & Conrad, S. (2003). 9 Register Variation: A Corpus Approach. *The handbook of discourse analysis*, 18, 175.
- Boersma, P. & Weenink, D. (2015): **Praat**: doing phonetics by computer [Computer program]
- Bosker, H. R. (2014). Research Note: The processing and evaluation of fluency in native and non-native speech.
- Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014a). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of memory and language*, 75, 104-116.

- Bosker, H. R., Quené, H., Sanders, T., & Jong, N. H. (2014b). The Perception of Fluency in Native and Nonnative Speech. *Language Learning*, 64(3), 579-614.
- Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T. J. M., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30, 157-175.
- Brown, G., & Yule, G. (1983). *Teaching the spoken language* (Vol. 2). Cambridge University Press.
- Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy* (Vol. 129, p. 33). Cambridge: Cambridge University Press.
- Cambier-Langeveld, G. M. (2000). *Temporal marking of accents and boundaries*. Holland Academic Graphics.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(3), 517.
- Chafe, W. L. (1980). The pear stories: Cognitive, cultural, and linguistic aspects of narrative production.
- Chambers, F. (1997). What do we mean by *fluency*? *System*, 25(4), 535-544.
- Cheng, W. (2007). The use of vague language across spoken genres in an intercultural Hong Kong corpus. In *Vague language explored* (pp. 161-181). Palgrave Macmillan UK.

- Clark, Herbert H. and Fox Tree, Jean. E. (2002) Using *uh* and *um* in spontaneous speaking. *Cognition* 84, 73–11.
- Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive psychology*, 37(3), 201-242.
- Collard, P. (2009). *Disfluency and listeners' attention: An investigation of the immediate and lasting effects of hesitations in speech*. Unpublished doctoral dissertation, The University of Edinburgh.
- Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America*, 77(6), 2142-2156.
- Council of Europe 2001: Common European Framework of Reference for Languages: learning, teaching, assessment. Cambridge. Cambridge University Press.
- Crawford, W.W. (1987). The pronunciation monitor: L2 acquisition considerations and pedagogical priorities. In J. Morley (ed.). *Current perspectives on pronunciation: Practices anchored in theory* (pp. 101-121). Washington D.C.: TESOL.
- Cucchiari, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, 107(2), 989-999.
- Cucchiari, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer assisted language learning*, 15(5), 441-467

- Cucchiarini, C. de Andrade, C. R. F., & Martins, V. O. (2011). Influence of gender and educational status on fluent adults' speech fluency. *Revista de logopedia, foniatría y audiología*, 31(2), 74-81.
- Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition*, 25(1), 57-71.
- Davies, A. (2003). *The native speaker: Myth and reality* (Vol. 38). Multilingual Matters.
- De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In *The 6th Workshop on Disfluency in Spontaneous Speech (DiSS)* (pp. 17-20).
- De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, 34(05), 893-916.
- DeKeyser, R.M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499-533.
- DeKeyser, R.M. (2012). Age effects in second language learning. In Susan Gass and Alison Mackey (eds.), *Handbook of Second Language Acquisition*, 442-460. London: Routledge.
- Deme, Andrea and Markó, Alexandra. (2013). Lengthenings and filled pauses in Hungarian adults' and children's speech. In *Proceedings of the 6th Workshop on Disfluency in*

- Spontaneous Speech, KTH Royal Institute of Technology, TMH-QPSR 54(1). Robert Ecklund (ed.). Stockholm: Royal Institute of Technology. 21–24.
- Deme, A. (2013). Pragmatic functions of lengthenings and filled pauses in the adult-directed speech of Hungarian children.
- Den, Y. (2003). Some strategies in prolonging speech segments in spontaneous Japanese. In *ISCA Tutorial and Research Workshop on Disfluency in Spontaneous Speech*.
- Derwing, T.M., & Rossiter, M.J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning*, 13, 1-18.
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language learning*, 54(4), 655-679.
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research- based approach. *TESOL Quarterly*, 39, 379-397.
- Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System*, 34(2), 183-193
- Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31(04), 533-557.
- Dörnyei, Z., & Thurrell, S. (1991). Strategic competence and how to teach it. *ELT journal*, 45(1), 16-23.

- Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. *Talking data: Transcription and coding in discourse research*, 45-89.
- Duncan, S., & Fiske, D. W. (1979). Dynamic patterning in conversation: Language, paralinguistic sounds, intonation, facial expressions, and gestures combine to form the detailed structure and strategy of face-to-face interactions. *American Scientist*, 67(1), 90-98.
- Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. *Perspectives on fluency*, 311.
- Eklund, R., & Shriberg, E. (1998). Crosslinguistic disfluency modelling: a comparative analysis of Swedish and American English human-human and human-machine dialogues. In *ICSLP* (Vol. 98, pp. 2631-2634).
- Eklund, Robert. (1999). A Comparative analysis of disfluencies in four Swedish travel dialogue corpora. *Proceedings of Disfluency in Spontaneous Speech Workshop*, University of Berkeley, California. 3-6.
- Eklund, Robert. (2001). Prolongations: A dark horse in the disfluency stable. *Proceedings of Disfluency in Spontaneous Speech*. University of Edinburgh, Scotland. 5-8.
- Esposito, A., & Bourbakis, N. (2006). The role of timing in speech perception and speech production processes and its effects on language impaired individuals. In *Sixth IEEE Symposium on BioInformatics and BioEngineering (BIBE'06)* (pp. 348-356). IEEE.
- Faerch, C., & Kasper, G. (1984). Two ways of defining communication strategies. *Language learning*, 34(1), 45-63.

- Fathman, A. K. (1980). Repetition and correction as an indication of speech planning and execution processes among second language learners. *Towards a cross-linguistic assessment of speech production*, 77-85.
- Fillmore, C. J. (1979). On fluency. In D. Kempler, and W. S. Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85-102). *New York: Academic Press*.
- Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *The Journal of the Acoustical Society of America*, 95(6), 3623-3641.
- Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. *Studies in interactional sociolinguistics*, 13, 134-184.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied linguistics*, 21(3), 354-375.
- Fox Tree, Jean E. (2002). Interpreting pauses and *ums* at turn exchanges. *Discourse Processes* 34(1): 37-55.
- Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, 34, 709-738.
- Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition*, 29(2), 320-326.

- Freed, B. (Ed.). (1995). *Second language acquisition in a study abroad context* (Vol. 9). John Benjamins Publishing.
- Freed, B. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder. *Perspectives on fluency*, 243-265.
- Gaitenby, J. (1965). The elastic word. *Haskins Laboratories Status Report on Speech Research*, 2, 3-1.
- Gass, S.M. & Varonis, E.M. (1984). The effect of familiarity on the comprehensibility of non-native speech. *Language Learning*, 34, 65-89.
- Gilquin, Gaëtanelle (2008) “Hesitation markers among EFL learners: Pragmatic deficiency or difference?”. In Jesús Romero-Trillo (ed.) *Pragmatics and Corpus Linguistics: A Mutualistic Entente* (pp. 119-149). Berlin, Heidelberg & New York: Mouton de Gruyter.
- Gilquin, Gaëtanelle (2011) “Corpus linguistics to bridge the gap between World Englishes and Learner Englishes”. In L. Ruiz Miyares & M.R. Álvarez Silva (eds) *Comunicación Social en el siglo XXI*, Vol. II (pp. 638-642). Santiago de Cuba: Centro de Lingüística Aplicada.
- Gilquin, G., & De Cock, S. (2011). Errors and disfluencies in spoken corpora: Setting the scene. *International Journal of Corpus Linguistics*, 16(2), 141-172.
- Gilquin, G. (2015). From design to collection of learner corpora. In: Sylviane Granger, Gaëtanelle Gilquin & Fanny Meunier, *The Cambridge Handbook of Learner Corpus Research*, Cambridge University Press: Cambridge 2015, p. 9-34.
- Gilquin, G. (2008). Hesitation markers among EFL learners: Pragmatic deficiency or difference. *Pragmatics and corpus linguistics: A mutualistic entente*, 119-149.

- Gilquin, G., De Cock, S., & Granger, S. (2010). The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM.
- Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, 27(3), 379-399.
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.
- Gósy, Mária. (2006). A semleges magánhangzó nyelvi funkciói [Functions of the neutral vowel]. *Beszéd kutatás 2006*. 8–22.
- 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* (Vol. 53). John Benjamins Publishing.
- Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29(3), 311-343.
- Granger, S. (1993) The International Corpus of Learner English. In Aarts, J., P. de Haan and N. Oostdijk (eds.) *English Language Corpora: Design, Analysis and Exploitation*. Amsterdam: Rodopi, pp. 57-69.
- Granger, S. (1994) The learner corpus: a revolution in applied linguistics. *English Today* 39 (10/3): 25-9.

- Granger, S. (2002). A Bird's-eye View of Computer Learner Corpus Research. In Granger, S., Hung, J. and Petch-Tyson, S. (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam and Philadelphia: Benjamins.
- Griffiths, R. (1991). Pausological Research in an L2 Context: A Rationale, and Review of Selected Studies. *Applied Linguistics*, 12(4), 345-64.
- Gut, U., & Fuchs, R. (2012). Exploring speaker fluency with phonologically annotated ICE corpora.
- Gut, U. (2012). The LeaP corpus A multilingual corpus of spoken English. *Multilingual corpora and multilingual corpus analysis*, 14, 3.
- Haggard, M. (1973). Correlations between successive segment durations: values in clusters. *Journal of Phonetics*, 1(1), 111-116.
- Hahn, L.D. (2004). Primary stress and intelligibility: research to motivate the teaching of suprasegmentals, *TESOL Quarterly*, 38, 201–223.
- Heldner, M., & Strangert, E. (2001). Temporal effects of focus in Swedish. *Journal of Phonetics*, 29(3), 329-361.
- Heritage, J. (2005). Conversation analysis and institutional talk. *Handbook of language and social interaction*, 103-147.
- Hird, K., & Kirsner, K. (2003). The effect of right cerebral hemisphere damage on collaborative planning in conversation: an analysis of intentional structure. *Clinical linguistics & phonetics*, 17(4-5), 309-315.

- Hirst, D. J. (2001). Automatic analysis of prosody for multilingual speech corpora. *Improvements in speech synthesis*, 320-327.
- Horváth, Viktória. (2004). Megakadásjelenségek a párbeszédekben [Disfluency phenomena in dialogues]. *Beszédkutató* 2004. 187–199.
- House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25(1), 105-113.
- Hymes, D. (1974). Ways of speaking. *Explorations in the ethnography of speaking*, 1, 433-451.
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing Comprehensibility. *Studies in Second Language Acquisition*, 34(03), 475-505.
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed). *Conversation Analysis: Studies from the First Generation*. (pp: 13-31). Amsterdam: John Benjamins.
- Jendryczka-Wierszycka, J. (2009). Collecting spoken learner data: Challenges and benefits--A Polish L1 perspective. In *Proceedings of Corpus Linguistics*.
- Kahn, D. (1976). Syllable-based generalizations in English phonology (Vol. 156). Bloomington: Indiana University Linguistics Club.
- Kaisse, E. M. (1985). *Connected speech: the interaction of syntax and phonology*. New York: Academic Press.

- Kahng, J-M. (2014). *Exploring the production and perception of second language fluency: Utterance, cognitive, and perceived fluency*. Unpublished dissertation. Michigan State University.
- Kang, O. (2008). The effect of rater background characteristics on the rating of international teaching assistants speaking proficiency. *Spain Fellow Working Papers*, 6, 118-206.
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301-315.
- 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, 554-566.
- Kang, O. (2012). Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly*, 9(3), 249-269.
- Kang, O., & Pickering, L. (2012). Acoustic and Temporal Analysis for Assessing Speaking. *The Companion to Language Assessment*.
- Kirsner, K., Dunn, J., & Hird, K. (2003). Fluency: Time for a paradigm shift. In *ISCA Tutorial and Research Workshop on Disfluency in Spontaneous Speech*.
- Klatt, D. H., & Cooper, W. E. (1975). Perception of segment duration in sentence contexts. In *Structure and process in speech perception* (pp. 69-89). Springer Berlin Heidelberg.

- Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., & Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Language and speech*, 41(3-4), 295-321.
- Kormos, J. (1999). Monitoring and self-repair in L2. *Language learning*, 49(2), 303-342.
- Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning*, 50(2), 343-384.
- Kormos, J., and Denes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145-164.
- Krashen, S. D., Long, M. A., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL quarterly*, 573-582.
- Kuruoğlu, G., & Altıparmak, A. (2015). A psycholinguistic point of view on prolongations of the native speakers of Turkish.
- Ladd, D. R. (1996). *Intonational Phonology*. Cambridge, UK: Cambridge University Press
- Lee, T. L., He, Y. F., Huang, Y. J., Tseng, S. C., & Eklund, R. (2004). Prolongation in spontaneous Mandarin. In *Interspeech*.
- Lennon, P. (1989). Introspection and Intentionality in Advanced Second-Language Acquisition. *Language Learning*, 39(3), 375-396.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language learning*, 40(3), 387-417.
- Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.). *Perspectives on fluency* (pp. 25-42). Michigan: The University of Michigan Press

- Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41-104.
- Levelt, W. J. M. (1989). ACL.
- Levinson, Stephen C., (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Linell, P. (2005). *The written language bias in linguistics*. London: Routledge.
- Maclay, Howard and Osgood, Charles. E. (1959). Hesitation phenomena in spontaneous speech. *Word* 15: 19–44.
- Martin, E. (1970). Toward an analysis of subjective phrase structure. *Psychological Bulletin*, 74(3), 153.
- Matthews, M. (2014). *The use of first- and second- language fillers by second-language Japanese speakers*. Unpublished masters thesis, University of Florida.
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh University Press.
- Möhle, D. (1984). A comparison of the second language speech production of different native speakers. *Second language productions*, 26-49.
- Moniz, H., Mata, A. I., & Viana, C. (2007). On filled-pauses and prolongations in European Portuguese. In *Interspeech*, 2645-2648.
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning*, 45(1), 73-97.
- Nesselhauf, N. (2004). How learner corpus analysis can contribute to language teaching: A study of support verb constructions. *Corpora and language learners*, 17, 109.

- Norrick, N. R. (2008). Using large corpora of conversation to investigate narrative: The case of interjections in conversational storytelling performance. *International journal of corpus linguistics*, 13(4), 438-464.
- Ochs, E. (1979). Transcription as theory. *Developmental pragmatics*, 10(1), 43-72.
- O'Connell, D. C., & Kowal, S. (1980). Prospectus for a science of pausology. *Temporal variables in speech*, 3-10.
- O'Connell, D. C., & Kowal, S. (1983). Pausology. *Computers in language research*, 2(19), 221-301.
- O'Connell, D. C., & Kowal, S. (1990). Some Sources of Error in the Transcription of Real Time in Spoken Discourse. *Georgetown Journal of Languages and Linguistics*, 1(4), 453-66.
- Ogden, R. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association*, 31(01), 139-152.
- Olynyk, M., D'Anglejan, A., & Sankoff, D. (1987). A quantitative and qualitative analysis of speech markers in the native and second language speech of bilinguals. *Applied Psycholinguistics*, 8 (121-136).
- Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The journal of the Acoustical Society of America*, 54(5), 1235-1247.
- O'Shaughnessy, D. (1992). Recognition of hesitations in spontaneous speech. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on* (Vol. 1, pp. 521-524). IEEE.

- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 317-331). Amsterdam: Elsevier.
- Pinget, A. F., Bosker, H. R., Quené, H., & de Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, 31(3), 349-365.
- Reed, M. (2000). He who hesitates: Hesitation phenomena as quality control in speech production, obstacles in non-native speech perception. *Journal of Education*, 67-91.
- Rehbein, J. (1987). On fluency in second language speech. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 97-105). Norwood, N.J.: Ablex.
- Riazantseva, A. (2001). Second language proficiency and pausing a study of Russian speakers of English. *Studies in Second Language Acquisition*, 23(04), 497-526.
- Rieger, C. L. (2003). Some conversational strategies and suggestions for teaching them. *Die Unterrichtspraxis/Teaching German*, 164-175.
- Riggenbach, H. (1991). Towards an understanding of fluency: A microanalysis of nonnative speaker conversation. *Discourse Processes*, 14, 423-441.
- Ringbom, Håkan. (1987). *The role of the first language in foreign language learning* (Vol. 34). Multilingual Matters Ltd.
- Rohr, J. (2015). "Use of a Learner Corpus in Examining Non-native Disfluency Patterns." Poster presented at the Learner Corpus Research Conference (LCR) 2015.

- 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.
- Rubin, J. (1987). Learner strategies: Theoretical assumptions, research history and typology. *Learner strategies in language learning*, 15-30.
- Rubin, D.L. & Smith, K. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of non-native English speaking teaching assistants. *International Journal of Intercultural Relations*, 14, 337-353.
- Rubin, J. (Ed.). (1987). *Learner strategies in language learning*. Macmillan College.
- Sajavaara, K. (1987). Second language speech production: Factors affecting fluency. *Psycholinguistic models of production*, 45-65.
- Savithri, S. (1984). Timing in speech: A review of literature. *The Journal of the All India Institute of Speech and Hearing*, 15, 7.
- Schegloff, E. A. (1978). On Some Questions and Ambiguities in Conversation" Emanuel A. Schegloff. *Current trends in textlinguistics*, 2, 81
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 361-382.
- Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. *Talking to learn: Conversation in second language acquisition*, 237-326.

- Schmidt, R. W. (1990). The role of consciousness in second language learning¹. *Applied Linguistics*, 11(2), 129-158.
- Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in second language acquisition*, 14, 357-357.
- Segalowitz, N. (2000). Automaticity and attention skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200-219). Michigan: The University of Michigan Press.
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in second language acquisition*, 26(02), 173-199.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.
- Seidlhofer, B. (2002). Pedagogy and local learner corpora: Working with learning-driven data. *Computer learner corpora, second language acquisition and foreign language teaching*, 213-234.
- Sinclair, J. M. (1996). The empty lexicon. *International journal of corpus linguistics*, 1(1), 99-119.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, P. & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1, 185-211. Skehan, P. & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49, 93-120.

- Skehan, P. (2003). Task-based instruction. *Language teaching*, 36(01), 1-14.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510-532.
- Stubbs, M. (1996). Text and corpus analysis: Computer-assisted studies of language and culture. Oxford: Blackwell.
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. Planning and task performance in a second language, 239-273.
- Tortel, A. (2008). ANGLISH : base de données comparatives L1 et L2 de l'anglais lu, répété et parlé, Travaux Interdisciplinaires du Laboratoire Parole et Langage (TIPA), 27, 111-122.
<http://sldr.org/sldr000731/fr>
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17, 84-119.
- Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *IRAL*, 40(2), 117-150.
- Tree, J. E. F. (1993). *Comprehension after speech disfluencies*. Stanford University.
- Tree, J. E. F., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, 62(2), 151-167.
- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, 15(04), 905-916.

- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in second language acquisition*, 28(01), 1-30.
- Vaissiere, J. (2002). Cross-linguistic prosodic transcription: French vs. English. Problems and methods of experimental phonetics. In honour of the 70th anniversary of Pr. LV Bondarko, 147-164.
- Verhoeven, L. T. (1989). Monitoring in children's second language speech. *Second Language Research*, 5(2), 141-155.
- Vural, E. (2008). Disfluency in second language: A quantitative study on Turkish learners of English (Doctoral dissertation, Middle East Technical University).
- 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.
- Wei, N., Li, W., & Pu, J. (2007). Design principles and annotation methods of the COLSEC corpus [J]. *Contemporary Linguistics*, 3, 235-246.
- Wennerstrom, A. (2000). The role of intonation in second language fluency. In H. Riggensbach (Ed.), *Perspectives on fluency* (pp. 102-127). Ann Arbor: University of Michigan Press.
- White, Ron. (1997). Back channelling, repair, pausing, and private speech. *Applied Linguistics* 18(3): 314-344.

Appendix A

Language Background Questionnaire

All information contained herein will be kept confidential.

For investigator use:

Participant research ID number: _____ Initials: _____

Personal Information:

What is your highest level of education completed? (please circle):

some high school high school some college college graduate

Country of origin: _____

1. If you were not born in the U.S., during what ages did you live in your country of origin?

2. If you were not born in the U.S., how long have you lived in the U.S.?

Language Background:

1. At what age did you begin to learn English?

2. Rate your current overall language ability in ENGLISH

1 = understand but cannot speak

2 = understand and can speak with great difficulty

3 = understand and speak but with some difficulty

4 = understand and speak comfortably, with little difficulty

5 = understand and speak fluently like a native speaker

3. If you are currently studying English in an American institution, please answer the following questions:

How many levels are there in the English program you are studying in? _____

What is your reading level? _____

What is your writing level? _____

What is your speaking level? _____

What is your listening level? _____

Appendix B

Sample Items for Free Speech Task

Free Speech Prompts

What will you do this weekend?

What is your favorite food?

What is your biggest dream for your future and why is it important to you?

Give directions from where you are now to the UC/nearest restroom/your classroom/your house.

Describe the plot/story of your favorite movie.

If you were given 1 million dollars, what would you do with the money?

What is the best piece of advice you have ever received?

Appendix C

Sample of Read-Aloud Task

Items_RC

The man that wore the suit talked to Sue in the store.
The girl that made the cake talked to him at the party.
The thief that stole the car walked to the dock in the east.
The priest that wore the cloak walked to the front of the church.
The girl that loved the boy looked at the moon every night.

Items_Pass

The dog was kept in the car at the mall.
Those girls were seen near the store in the mall.
The bread was eaten with the cheese from the store.
The car was bought for the boy with red hair.
The kids were met by their friends from high school.

Items_Interr

Who did the boys hit on the head?
Who did the kids find near the lake?
Who will the cops hunt in the woods?
Who will the girl see at the store?
Who does the priest know at the church?

Items_Decl

The happy couple ate the fruit in their yard.
The laughing kids ate the cupcakes at the party.
A beautiful girl smiled at the man in the suit.
Those barking puppies chewed the leather on my shoes.
Some purple flowers sat in the vase in the window.

Items_Texts

Text 1: Beware of the Dog (Roald Dahl)

But he did not sleep. He wanted to keep his eyes open because he was frightened that if he shut them again everything would go away. He lay looking at the ceiling. The fly was still there. It was very energetic. It would run forward very fast for a few inches, then it would stop. Then it would run forward again, stop, run forward, stop, and every now and then it would take off and buzz around viciously in small circles. It always landed back in the same place on the ceiling and started running and stopping all over again. He watched it for so long that after a while it was no longer a fly, but only a black speck upon a sea of gray, and he was still watching it when the nurse opened the door, and stood aside while the doctor came in. He was an Army doctor, a major, and he had some last war ribbons on his chest. He was bald and small, but he had a cheerful face and kind eyes.

Appendix D

Experimental Design

Speaker	Utterance	Lang. Back.	High/Initial	Low/Initial	High/Non-initial	Low/Non-initial
2 A	NNS		Yellow	Blue	Green	Red
2 B	NNS		Red	Yellow	Blue	Green
3 A	NNS		Green	Red	Yellow	Blue
3 B	NNS		Blue	Green	Red	Yellow
4 A	NNS		Yellow	Blue	Green	Red
4 B	NNS		Red	Yellow	Blue	Green
6 A	NNS		Green	Red	Yellow	Blue
6 B	NNS		Blue	Green	Red	Yellow
7 A	NNS		Yellow	Blue	Green	Red
7 B	NNS		Red	Yellow	Blue	Green
8 A	NNS		Green	Red	Yellow	Blue
8 B	NNS		Blue	Green	Red	Yellow
10 A	NS		Yellow	Blue	Green	Red
10 B	NS		Red	Yellow	Blue	Green
11 A	NS		Green	Red	Yellow	Blue
11 B	NS		Blue	Green	Red	Yellow

