Pause for Thought: A Pilot Comparative Study of Pause Placement Amongst Native, Heritage, and Non-Native Speakers

by

Mary Kathryn Kennedy

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Approved April 2021 by the
Graduate Supervisory Committee:

Elly van Gelderen, Chair
Kathryn Pruitt
Matthew Prior

ARIZONA STATE UNIVERSITY

May 2021

ABSTRACT

Temporal features and frequency of pauses have been studied extensively in the literature, but the interest in the syntactic location of pauses is a more recent development. While previous research has studied the pause patterns of L1 and L2 speakers as well as the effects of pause location on perceptions of fluency, these studies have all utilized a binary approach the categorization of pauses as occurring either between or within clauses or major constituent boundaries. This research attempts to take a look at pause placement with a finer distinction of pause location, including junctures that occur between and within phrases. To accomplish this, two experiments were conducted. The first experiment gathered read-aloud speech samples from native, non-native, and heritage speakers of Mandarin Chinese, which were then manipulated in Praat to contain only a single pause that occurred either between or within phrases. The samples were presented to native Chinese speakers to assess for perceptions of fluency as affected by the pause location condition. Findings of this preliminary pilot study did not find a significant correlation between pause location and perceptions of fluency at the phrasal level. The second experiment gathered spontaneous speech samples from the same speaker population as Experiment 1. The pauses that occurred in the samples were coded according to a system developed by the author to account for eight different syntactic junctions, and the percentage of pause at each location was calculated. Analysis showed a significant correlation with pause location and percentage of pauses ($p < 0.01$), as well as a statistically significant interaction between the effects of speaker status and pause location on percentage of pause ($p = 0.011$). The findings of this study are limited due to the small population size, but research in this fine-grained analysis of pause

location within a clause has implications in the fields of L2 acquisition, psycholinguistics, and natural language processing.

DEDICATION

I would not be here writing this paper if it were not the wonderful, supportive, and loving people in my life. My family has listened to me chatter about grammar since I was in middle school, and they've always encouraged and supported me in pursuing my studies, even if they think I'm crazy to do it. Were it not for my parents who made sure I ate during those days when I got caught up with work, I think I might have starved. Dad, you always brought me dessert when I was feeling low, and it always made me smile. And Mom, you have been my guiding star when I'm lost, my lighthouse when I'm uncertain where I am or what I'm doing, and my safe harbor in a storm. You've been there through all the tough times with a hug and a kiss and the promise that you'll always be there when I need you. Thank you to all my family for always believing me and always being my home to come back to. I love you all, and I'm lucky to have been born into a weird, odd-ball family.

This paper and research would also not exist were it not for my friends and family with their words of support, reminders to breathe, jokes and laughs, and their smarty, smarty brains. To Kyra, who has had the patience of a saint and has been one of my biggest supporters throughout college. Random roommates to buddies of studies for life, my friend. Thank you for listening to me rant, for offering advice, for editing the roughest of rough drafts, and for our nights of ice cream, face masks, and John Mulaney and the Big Fat Quiz. May we get to see the Killers together after COVID is over. To Brandon, for answering my context-less questions without hesitation and for giving an entire background to the answer as well as detailed plan going forward. You operate at ax-level, and I'm mighty glad you're on my team. To Callyn, for the years of friendship and the

more going forward. No matter where we are or how long it's been, I know you're always there ready to pick up where we left (often with a new story, theory, or scheme). To John, for celebrating the happy news and good times, consoling me when the bad news came, and providing comfort and support during the rough times. Any day or situation is was better for having you in it.

And lastly, to my dog, Maizey, who is curled up next to me as I write this. You've spent nearly every minute by my side and have, as such, been present for the better part of this study. No one else would stay up with me to 4 in the morning. You're a needy, goofy, weird little dog, but I love you very much.

P.S. Not long after finishing this dedication, I lost a portion of my thesis in the 11th hour to a corrupted file issue. While this was a rather unlucky experience, I can say that I am anything but unlucky. Everyone circled the wagons to offer what help they could, and I was blown away by how amazingly lucky I am to have these people in my life. I just wrote an entire page thanking you all, but your actions in the past three days have truly humbled me and filled me with so much gratitude to have in my life people who go above and beyond when it comes to helping others. Not only were my friends and family coming to my aid during these times, but entire strangers from tech support were rigorous and dedicated in their efforts to fix the issue, and genuine and sincere in their condolences when they came to the conclusion that nothing could be done. So to the girl at the Apple store who gave me a most-needed hug before I left, and to all the people I've ever met and never met and am yet to meet, your small acts of kindness and compassion help remind me that there is always something and someone good out there if I just go out and look. Thank you for making this world a brighter place.

ACKNOWLEDGMENTS

My committee members have all known me since my undergraduate studies. My first class with Dr. Prior was ENG 312, English in Its Social Setting, where I conducted my very first baby research survey to study non-native speaker's perceptions of American English dialects. Dr. Pruitt taught ENG 313, Phonology and Morphology. And ENG 414, Typology, was my first class with Dr. van Gelderen. These three professors have been involved throughout my academic studies, both at the undergraduate level and beyond. Were it not for the classes I took with each of these professors, this study would not exist. Dr. Prior, your teaching of SLA made me want to conduct a comparative study across various language users. Through the years, you always reminded me to breathe and to be kind to myself, and I was always thankful for a teacher who cared so much about his students. Because of Dr. Pruitt, I have now conducted a study on pause and intend to continue studies of prosody going forward. Thank you for all your help and for directing me towards studies in the syntax-phonology interface. And to Dr. van Gelderen, I will be forever grateful that you were my teacher of all things syntax-related. I came to college with an interest in sentence structure, and through your courses, that interest became a love and an enduring fascination. Thank you for your advice and your support both in this research and throughout the years. My academic studies and who I am as an academic would not be what they are today had I not been blessed with professors as dedicated to their subjects and their students as the three I am honored to have on my committee.

Lastly, I would like to acknowledge Tian-Yi Ni, who has been both an admired scholar and a valued friend to me. Thank you for your help and guidance in the Chinese language. I appreciate and cherish all the laughs, the chats, and the support.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATIONS

Adj(P)	Adjective (Phrase)
Adv(P)	Adverb (Phrase)
Adv_{DEG}(P)	Degree Adverb (Phrase)
A(P)	Adverbial (Phrase)
Asp(P)	Aspect (Phrase)
Cl	Clause
Class	Classifier
CoorConj	Coordinating Conjunction
C(P)	Complement (Phrase)
DM	Discourse Marker
D(P)	Determiner (Phrase)
Infl(P)	Inflection (Phrase)
Mod	Modifier
Neg	Negative
N(P)	Noun (Phrase)
P(P)	Preposition (Phrase)
Q(P)	Quantifier (Phrase)
RelCl	Relative Clause
S_{ADV}	Sentential Adverbial
Spec	Specifier
SubConj	Subordinating Conjunction
T(P)	Tense (Phrase)
V_S	Stative Verb

CHAPTER 1

INTRODUCTION

The inspiration for this thesis topic arose from a personal experience in my 400-level Chinese language course. For the class, we were required to read the text aloud and submit our recording to our professor to grade for our fluency. I struggled to say an entire sentence through without stopping to pause for breath or regroup my thoughts and continue to read the sentence. As such, my grades for these texts were confined to the high B, low A range. However, I began to listen to the textbook's audio recording and annotated where the reader paused before marking the spot in my own textbook. I noted every pause the reader made, and in my own read-aloud session, I would then pause at these junctures myself. I had already been chunking sentences by punctuation points, but clauses can be quite lengthy, and I would still struggle. This new method allowed me to mark pauses *between* those punctuation points. At these non-punctual pause points, I could stop and either take a breath, or regroup my mind enough to continue to read the text before I hit either the next pause point or the next punctuation point. After beginning this practice, I noticed my grades improved by half a letter grade to largely mid A's. Because of this experience, I wanted to conduct a study to see if where we pause syntactically affects how fluent we are perceived to be.

To address this topic, I developed two experiments: one which involved eliciting spontaneous speech samples from L1 and L2 speakers of Mandarin Chinese, and coding where their pauses occurred in the syntax; the second of which involved Mandarin Chinese L1 speakers assessing the fluency of speech samples from L1 and L2 speakers that were manipulated so that pauses occurred at specific syntactic junctures. The former

experiment was constructed to better understand if there is a pause-pattern specific to L1 speakers of Mandarin Chinese and whether L2 speakers reflect or differ from those pause patterns if they do exist. The latter experiment is to determine whether pause placement affects perceptions of fluency and to understand if pauses at certain syntactic junctures are more detrimental to perceptions of fluency than other pause junctures.

My experience in my language classroom was an isolated occurrence, involving only me and my teacher. Through this study, I hope to understand if my own experience was an anomalous, one-off occurrence, or if there is an underlying cause to our perception of language and fluency that interacts or even intersects with the syntax-prosody interface. Though the sample size of this pilot study is limited, prospective findings of this study could have implications in the fields of second language acquisition, psycholinguistics, and natural language processing. The following paper is thus arranged as follows: Chapter 2 provides an review of the relevant literature regarding fluency, pause study, the syntax-prosody interface, considerations for Chinese as a selected language, and how my study addresses the gap as well as its relevance. Chapter 3 details Experiment 1's research design, including the methodology, recruitment procedure (Section 3.1), participants (Section 3. 2), and data collection (Section 3.3); and Chapter 4 reports the results (Section 4.1) and discussion (Section 4.2). Chapter 5 covers Experiment 2's methodology, including participants (Section 5.1) and data collection (Section 5.2), while Chapter 6 reports the results (Section 6.1) and discussion (Section 6.2) for Experiment 2. The thesis thus ends on Chapter 7 with final conclusions regarding the findings as well as insights into future research possibilities.

CHAPTER 2

LITERATURE REVIEW

Because this research concerns several areas of study—that of syntax, prosody, and fluency (both in second-language acquisition and cognitive processing)—the following literature review is sorted into four main sections. As a quick note, I will be referring to my participants as "L1" and "L2" Mandarin Chinese speakers in this study. Though there is a discussion in the community as to whether people should be referred to as "native/non-native" speakers or as L1/L2/L3 users, I will be using the terms "L1/L2 speakers" and "native/non-native speakers" interchangeably, though it is important to distinguish that some of my participants are actually L3 Mandarin Chinese speakers, some are heritage speakers, and some may consider themselves fully bi- or multi-lingual. Because my study does not strictly pertain to the issue of identity in language, I will be referring to the participants as they identified themselves to me when discussing speakers individually.

2.1 Fluency

Fluency is a term used quite often in the field of linguistics, particularly in the area of second-language acquisition (SLA). Despite being used frequently, what fluency exactly is and how it functions can be unclear. In 1990, Lennon argued that fluency is purely a performative production; that is to say, it is a judgement made by the listener of the speaker's speech planning and production based on the speaker's utterance, including factors such as speech rate, pause, and corrections (as cited in Segalowitz, 2010, and Kahng, 2018). A good summary is to say that fluent speech is the result of planning and

executing an utterance nearly simultaneously without or with little error (Rehbein, 1987, as cited in Segalowitz, 2010).

More specifically, however, I will be adopting Segalowitz's (2010) conceptualization of fluency as consisting of three main categories: cognitive fluency, utterance fluency, and perceived fluency. The first category pertains to "the speaker's ability to efficiently mobilize and integrate the underlying cognitive processes responsible for producing utterances with the characteristics they have" (Segalowitz, 2010, 48). That is to say, because fluency is the result of simultaneous planning and execution of an utterance, one must be able to organize all the required information—the lexical words, the grammatical words, the syntactic order, in addition to any necessary or wanted prosodic information—in a short amount of time. Slow or disjunctive gathering and ordering of the information will lead to a decreased cognitive fluency. Utterance fluency pertains more to the actualization of the utterance and the more "physical" aspects of the utterance; as such, utterance fluency factors include temporal features (such as speech rate), pauses, hesitations, and repairs. Lastly, perceived fluency steps outside of the speaker and into the listener who makes assessment of the speaker based on the former two fluencies. As such, perception fluency refers to "the inferences listeners make about a speaker's cognitive fluency based on their perception of utterance fluency" (Segalowitz, 2010). For my own study, Experiment 2 will directly pertain to perceived fluency as native Mandarin speakers are asked to assess how fluent they think the speaker is based on their audio sample, while Experiment 1 will involve a bit more of utterance fluency as the locations of pauses in spontaneous speech samples are coded.

Acoustic fluency measurements can generally be broken into three main categories: speed, breakdown, and repair (Witton-Davies, 2014; Pinget, Bosker, Quené, & de Jong, 2014; Tavakoli & Skehan, 2005), the first of which is generally measured through either the number of syllables or words per minute (Skehan 2009) or the mean length of syllables or words per minute (Pinget, Bosker, Quené, & de Jong, 2014), the second of which is identified through pause (both silent and filled) measurements, and the last of which involves repetitions, false starts, replacements, and reformulations (Skehan, 2009). Previous research varies as to what measure of fluency is the strongest predictor of perceived fluency. Several studies suggest the strongest correlations with perceptions of fluency derive from mean length of run (Kahng, 2014; Prefontaine, 2013), while others find speech/articulation rate (Kahng, 2014; Liu & Wu, 2016; Prefontaine, Kormos, & Johnson, 2016; Saito et al., 2017), average pause time (Prefontaine, 2013), filled pause frequency (Révśz, Ekiert, & Torgersen, 2014), mid-clause pause frequency (Saito et al., 2017; Suzuki & Kormos, 2020), and clause-final pause frequency (Saito et al., 2017; Lambert, Kormos, & Minn, 2016).

As can be seen, pause—in some form or another, through its frequency, its length, or its location—bears a non-negligible impact on perceptions of fluency. For my thesis, I will be focusing on pause location as it is a representation of the syntax-prosody interface (discussed below). While there is a growing body of research in the last few years that shows where a pause occurs is not unlikely to impact the speakers perceptions (Saito et al 2017; Kahng, 2018; Kahng, 2020; Suzuki & Kormos, 2020) in that even clause-final pauses may be perceived more negatively than no clause at all (see Saito et al., 2017, and Kahng, 2018, for affirmative findings; and Suzuki & Kormos, 2020, for contradictory

findings) and that mid-clause pauses are perceived more negatively than clause-final pauses (Saito et al., 2017; Kahng, 2018; Kahng, 2020; Suzuki & Kormos, 2020). To date though, no study has yet researched the importance of location within a clause or at the phrasal level; that is to say, are there places within a clause or phrase that are considered to be worse places to pause from a fluency perspective? This is the focus of the first research question of this study.

Though pauses are considered part of breakdown fluency factors, there is also research to suggest that pauses are part of natural speech prosody. Pauses can generally be split into two realizations—"silent" or "unfilled" and "filled" pauses (such as *uh* and *um* in English)—and serve a variety of roles, such as rhetorical purposes, stylistic choices, interjective utterances, and turn-taking indicators amongst others (O'Connel and Kowal, 2008). Categorically, though, there are two main types of pauses: prosodic pause, which is a part of the discourse construction and typically occurs at major constituent boundaries, such as between clauses or between intonational phrases; and hesitation pause, which generally occurs at minor constituent boundaries or after the first word of an intonational phrase (Kahng, 2018; Cruttenden, 1997). As referenced above, the research is still indecisive as to whether between-clause/clause-final pauses are significantly negatively correlated with perceptions of fluency, though there is a general consensus that mid-clause pauses (which would, by definition, include pauses at minor constituent boundaries) are negatively perceived. That being said, there is something of a double standard; O'Connel and Kowal (2008) mention pause can serve a variety of roles and these "disfluencies" can serve a discourse or socio-pragmatic function. Despite L1 and L2 users both producing disfluencies, L1 users are generally ipso facto considered

fluent (Davies, 2003). To better understand if there is a bias against L2 users based on pause location, I will be using my second experiment to code where pauses occur for L1 and L2 users of Mandarin Chinese to identify if there is a significant difference in where pauses occur syntactically for the two groups.

2.2 The Syntax-Prosody Interface

Relevant to this research is Selkirk's pioneering work on the syntax-prosody interface (2006, 2009b, 2011). Selkirk develops the Match Theory of syntactic-prosodic constituency correspondence, which involves three core units of *match clause* (*i*), *match phrase* (*φ*), and *match word* (*w*), whereby a clause/phrase/word "must be matched by a corresponding prosodic constituent … in the phonological representation" (2011, 439). Under this theory, the phonological domains are expected to reflect syntactic constituents ideally; however, higher-ranked prosodic well-formedness constraints[1], such as markedness constraints, may cause the two systems to differ from each other in that the phonological domains do not sync with the syntactic constituents (ibid). Furthermore, because Selkirk's theory already incorporates Chomsky's work, I will be utilizing the minimalist program of Generative Grammar model for my analysis and parsing of syntax for the purposes of this paper.

For this work, Selkirk (2011) adopts several syntactic definitions. The "standard" clause can be embedded or the matrix clause, is the complement to the functional head of C, includes an overt or implied subject, a predicate, and Tense, and sits in the CP (Complementizer phrase) level; an "illocutionary" clause serves a more discourse-driven

[1] See Prince and Smolensky (1993, 2004) for their work on Optimality Theory, which addresses the concept of constraints in phonological theory.

function and would more currently be called a Force phrase (ForcP), which sits higher in the tree and "looks outside of the clause [to] [indicate] mood" (vanGelderen 2013, 154). It is worth noting that according to Selkirk (2011), an FP is more likely to correspond to an intonational phrase than a CP, particularly when considering embedded phrases (for example, parentheticals are more likely to be phonologically marked in English—and proposedly in all languages—than a clause that serves as a complement). Phrases in Selkirk's work generally correspond to Chomsky's X-Bar Theory, though Selkirk makes the distinguishment that while lexical projections—like NP, VP, and AP—are not phonologically differentiated in terms of prosody, lexical phrases and functional phrases *are* phonologically distinct in that lexical phrases are more likely to conform to the φ-domain while functional phrases are less likely to do so if at all. One can further distinguish a maximal φ where the phonological phrase is undominated by any other, a minimal φ where the phonological phrase does not dominate any other phrase, and a simple φ for all other instances (Itô & Mester, 2007). Following that, lexical words are parsed at the phonological *w*-domain while functional words are not necessarily so, hence the common tendency for functional words to cliticize and undergo phonological reduction. However, the claim that there is some inherent distinction between a lexical and a functional word that affects their phonological and prosodic representation is challenged; Kruger (2019) contends that the differing phonological realization is due to where functional words occur in the syntax, thus making the prosodic difference not a byproduct of some inherent quality of lexical and functional words and phrases, but is rather an inherent prosodic result of particular syntactic constructions and configurations.

This theory is generally in line with much of the previous work, though its detailing is more descriptive and reasoning broader. Previous phonological research has observed the correspondences between phonological groupings and syntactic constituents. As mentioned, Cruttenden (1997) noted that prosodic pauses usually occur at major constituent boundaries while hesitation pauses tend to occur at minor constituent boundaries. It also aligns with the phonological bootstrapping hypothesis that Morgan and Demuth (1996) redefined where infants and children can extrapolate enough linguistic information from a speaker's phonology so as to be able to guess at their syntactic construction even if the meaning or part of speech of a word is not otherwise known to the child; further studies support this theory (Christophe et al., 2008; Hawthorne & Gerken, 2014).

2.3 The Chinese Language

My study examines the Mandarin Chinese language. On the whole, Chinese belongs to the Sino-Tibetan language family, which is alternatively known as the Indo-Chinese language family or the Sinitic languages, all of which have tonal systems (Li, 1973). Chinese's tonal system is lexical phonemic in that though there are certain phonological environments and syntactic structures that can alter a tone (see Simpson, 2014, for further detail), a word's tone is generally assigned an underlying tone (Li & Thompson, 1976). As a language, Chinese is incredibly diverse and consists of a multitude of varieties, some of which are incredibly distinct from each other. Overall, the Chinese language consists of seven main dialectal groups: Mandarin, Wu, Xiang, Gan, Hakka, Min, and Yue (Kurpaska, 2010). The differences of these languages can include phonological variation in pronunciation, different lexical choices, and even different

syntactical configurations. The Mandarin dialectal group has its own subsets of varieties, including Northeastern, Jilu (Beifang), Jiaoliao, Beijing, Zhongyuan, Lanyin, Southwestern, Jianghuai, and Unclassified (Wurm et al., 1988). These varieties then have their own regional varieties, though due to the sheer number of varieties that would derive from nine larger varieties, I will not list them all.

The supergroup of Mandarin includes three large groups of Northern, Eastern, and Southwestern Mandarin, though these can be further subdivided into Northeastern Mandarin, Beijing Mandarin, Beifang (Jilu) Mandarin, Jiaoliao Mandarin, Zhongyuan Mandarin, Lanyin Mandarin, Southwestern Mandarin, and Jianghuai Mandarin (Wurm et al., 1988). While there are decent and varying levels of mutual unintelligibility amongst Southern Chinese dialects, Mandarin varieties (along with and including Taiyuan, Beijing, and Jinan, along with the regrouped Xi'an, Hannkou, and Chengdu) are generally mutually intelligible amongst each other at both the word and sentence level (van Heuven & Tang, 2009). Because of this mutual intelligibility, I have opted to study the broader category of "Mandarin Chinese." Furthermore, the establishment of 普通话 *pǔtōnghuà*, a variety of Chinese based off the Beijing dialect, as the national language and language of education lead to the spread of Mandarin amongst Chinese (Xia, 2017), leading to a world-wide speaker population of 1.12 billion speakers and serving as one of if not *the* official language of People's Republic of China, the Republic of China (Taiwan), and Singapore (Eberhard, Simons, & Fennig, 2021).

Typologically speaking, Chinese is a mono-morpho-syllabic language in that most morphemes contain only a single syllable, though this is not to say that most words in Chinese are monosyllabic; many are in fact disyllabic with a fair number of multisyllabic

words (Norman, 1988). Though originally thought to be an OV language (Li & Thompson, 1989), the grammar of Chinese is generally agreed to be SVO, and despite being an analytic language with little (arguably no) inflection, is a pro-dropping language (Paul, 2014) with a heavy use of the Topic structure (Shi, 2000).

In Chinese prosody, a prosodic word is often disyllabic as many lexical words themselves are disyllabic, and the prosodic unit often begins with a stressed, full-toned syllable, shortening of the word-initial syllable, boundary lengthening, and pitch discontinuity (though usually there are no pauses); however, a prosodic word[2] can (but not necessarily always does) include several words that operate together, such as (1) (Wang, 2003, as cited in Yang, 2016).

1) 他　每天　　(Reproduced from Wang, 2003)
Tā　*měitiān*
3.S　everyday

A singular prosodic word can in and of itself constitute a phonological phrase, but more often, a phonological phrase contains a small pause, pre-boundary lengthening, and pitch reset (Yang, 2016). Above that, the intonational phrase group is offset by pre-boundary lengthening, a pause that generally corresponds to a major constituent boundary, and a pitch reset (Peng et al., 2005, as cited in Yang, 2016). Overall, research has shown that silent pause is a frequent (though not required) marker of discourse boundaries, particularly in Taiwan Mandarin (as opposed to filled pause as a marker of discourse boundaries), and that length of pause generally corresponds with boundary hierarchy in that a higher discourse level was more likely to have a longer pause (Fon,

[2] For further reading on prosodic words in Chinese (specifically, Shanghai dialect), please see Selkirk and Shen (1990).

Johnson, & Chen, 2010; Xie, Xu, & Wang, 2012; Yang, Shen, Li, & Yang, 2014). And while Chinese's status as a tonal language can cause some phonological effects that need to be taken into account when considering prosody[3] (Shen, 1990), tones do not affect silent pauses.

Lastly, the following study considers only pause placement despite other prosodic features—such as syllable length and pitch reset—often co-occuring as a means to demark prosodic units such as intonational phrase or prosodic phrase. This is because while Taiwan Mandarin uses syllable and pause duration in spontaneous speech to distinguish discourse hierarchy, Mainland Mandarin uses only pause, indicating that pause alone is a sufficient phonological marker (Fon, Johnson, & Chen, 2010).

2.4 The Gap

Prosody, pausology, syntax, and fluency are all well-researched areas in the field of linguistics, but the intersection of all four is a little more scarce and yet unexplored. Previous research has found that speakers who pause less frequently and for shorter durations are generally considered more fluent (Bosker et al., 2013; Kormos & Deńes, 2004; and Cucchiarini et al., 2002), and that learners typically pause more frequently as compared to native speakers (who also pause, but not with as great a frequency as language learners) (Götz, 2013). Furthermore, pauses do not affect speaker comprehensibility as much as perceived fluency, meaning what the speaker says is understandable, though not phonologically fluent (Suzuki & Kormos, 2020). However, research into the effect of pause location on pause is a little more scarce. In general, while

[3] For further reading on the interaction between tone and intonation in the Chinese language, please see Shen (1990).

research finds that utterances without pause were considered most fluent, utterances with pauses between clauses were judged more fluent that pauses within clauses (Chen, 2015; Saito et al., 2017; Kahng, 2018; Shea & Leonard, 2019; Suzuki & Kormos, 2020). Other research furthermore has shown that increased proficiency in a language (either as a native speaker compared to a non-native speaker, or as a comparison of proficiency amongst language learners) generally corresponds with an increased preference for pauses at clausal boundaries (Chen, 2015; Lambert, Kormos, & Minn, 2016; Saito et al., 2017; Kahng, 2018; Shea & Leonard, 2019; Kahng, 2020; Suzuki & Kormos, 2020). Within that, though, research has not yet been done on the effect of pause placement *within* clause, such as at the phrasal level. From this gap organically emerges RQ1: does various pause placement within a clause (specifically between or within a phrase) affect perceptions of fluency? Following previous literature with between-clause and within-clause pauses, I hypothesize that pauses within phrases are more likely to be perceived as more disfluent than pauses between phrases. Thus, the null hypothesis (RQ1 H_0) is stated as: pause placement (at the phrasal level) does not affect perceptions of fluency.

Additionally, the literature has found that in regards to pause patterns, prosody research shows that intonational phrases (*i*) generally correspond with clauses, and that such intonational phrases are often demarked by pause amongst other prosodic features, such as pre-boundary lengthening and pitch reset (Peng et al., 2005, as cited in Yang, 2016; Fon, Johnson, & Chen, 2010; Xie, Xu, & Wang, 2012; Yang, Sheng, Li, & Yang, 2014). In regards to L1 and L2 pause patterns, previous research has found an increase in the number of within-clause pauses amongst L2 speakers as compared to L1 speakers (de Johg, 2016; Duran-Karaoz & Tavakoli, 2020). Again though, the literature does not

distinguish all the various syntactic junctures that exist, such as the various syntactic locations *within* a clause and even various syntactic locations that occur *between* clauses. It is from this gap that RQ2 emerges: where are pauses likely to occur (e.g. between a Head and Complement, or between a sentential adverb and its Head, and so on)? RQ2 H_0 is stated as: there is no predictable location of a pause within a sentence, and therefore, no analytic measurement correlating pause location to speaker status. While Experiment 2 is more exploratory in nature, I do still have expectations pauses at clausal boundaries will be most common (this follows with prosodic research) while pauses between syntactic constituents with a close relationship (such as Heads and their Complements) will be less likely to have a pause separating them.

CHAPTER 3

EXPERIMENT 1: METHODOLOGY

3.1 Recruitment

For this study, there were two groups of participants: speakers and raters. Speakers were recruited through purposive sampling from Chinese language courses and other area studies courses via email, text, or social media (see Appendix A). As speakers were expected to give a short speech about themselves, speakers were required to either have studied Chinese for a minimum of two years at the college level or to speak Chinese as one of their first languages. Participants were required to be over 18, and other protected classes were excluded from this study. Additionally, as my study was only approved for use in America and Taiwan, all participants had to be located in either the US or Taiwan at time of participation.

Speaker participants were asked to engage in a series of tasks, including answering (in English or Chinese) some biographically information questions, answering two open-ended prompts for approximately a minute (in Chinese), and to read three sentences in Chinese aloud (simplified, traditional, pinyin, and translation provided) (Appendix C). The length of speaker participation was participant-driven in that the participant could spend as little or as much time providing their audio samples as they liked. The shortest session took about 10 minutes, and the longest session took a little under an hour as some speakers wished to practice several times or to simply chat. A full session was expected to take around 15 minutes (and many did), and as it was expected to be a short time, participants were made aware that they wouldn't be compensated for their time. These participants were made aware their participation was entirely voluntary

and optional and that they could withdraw from the study at any point and they could opt to have their audio samples deleted at the end of the study.

Raters were recruited by purposive and voluntary response sampling via email, text, social media post (see Appendix B), or through English language courses taught on Zoom, various international student organizations, personal contacts and friends of friends, as well as several word-of-mouth participants. All participants—regardless of how they heard of the study—were provided an overview of the study in either the form of the email, text, or social media post. In order to participate as a rater, participants had to be over 18 and speak Mandarin Chinese as (one of) their first language(s) in addition to being located in either the US or Taiwan at the time of the survey. Raters were provided a link to an online survey that took an average of 28 minutes to complete (Appendix C). As it was known the survey would take roughly half an hour to 45 minutes, raters were offered a $5 compensation. If they lived in Taiwan at time of completing the survey, this compensation was offered through a PayPal transfer (equal to 130–150 NTD, dependent upon exchange rate for that day); if the rater was located in the US at time of participation, they were offered a PayPal transfer of $5, or a $5 e-gift card to Amazon, Target, or Starbucks. Like speakers, participants were made aware that participation was entirely voluntary and optional, that they could withdraw at any time, and that they could opt for their responses to be deleted at the end of the survey.

3.2 Participants

For this study, there were two groups of participants: speakers and raters. Speakers provided stimuli for Experiment 1 and speech data for Experiment 2.

3.2.1 Speaker Participants

A total of 24 speakers were recruited, and of the 24 speakers, 5 were native speakers of Chinese, 3 were heritage speakers (meaning they grew up in a household that included Chinese as one of if not *the* sole language, but in an area or country that did not speak Chinese), and the remaining 16 were non-native speakers. The ages ranged from 19 to 34 with an average of 24.1 and a median of 24 years old; 17 participants were female and 7 were male (Table 1).

Table 1
Speaker Participant Demographics (Exp. 1)

Speaker	Age		N. Females	N. Males	Total Subj.
	Mean	S.D.			
Native	26.6	4.72	3	2	5
Heritage	20.67	1.53	3	0	3
Non-native	24	3.29	11	5	16
Total	24.13	3.75	17	7	24

Speaker refers to whether the speaker is a native, heritage, or non-native speaker. Age is presented in Mean and Standard Deviation in years. N. Females and N. Males represent the number of each gender present for the three speakers, while Total Subj. represents the number of subjects for each speaker group.

Of the 19 non-native and heritage speakers (average age: 23.5), 16 identified their first language as English, 1 as Italian, 1 as Spanish, and listed both English and Cambodian as their first language, hence the total reaching "20" for first-languages.

Table 2
Non-Native and Heritage Speaker Language Experience (Exp. 1)

Speaker	Age at Start of Learn		Time at College		ILR	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Heritage	13.67	5.03	3	1	32.67	5.77
Non-native	15.5	2.71	3.5	1.05	23.88	6.17
Total	15.22	2.99	3.42	1.03	25.26	6.81

Age refers to the age (in years) at which the speaker began learning Chinese in a formal setting (re: in a class). Time at college refers to the number of years the speaker studied or has studied the language in college. ILR refers to the speaker's self-assessment of their speaking ability in the language according to the Interagency Language Roundtable.

All 19 listed either Mandarin, Standard Mandarin, or 普通话 *Pǔtōnghuà* as their dialect of Chinese. Table 2 addresses the non-native and heritage speakers' language background, including when they first began learning the language in the formal setting of a classroom, how long they spent learning the language at the college level, and their self-assessment of their language ability. It is worth noting that all three heritage speakers attended formal language instruction at some point in time, but two heritage language speakers indicated that they began learning the language from an earlier age through family members; the third heritage speaker did not disclose from whom they first learned the language, but that it was spoken in the home. All participants had spent a minimum of two years studying Mandarin at the college level. Not all participants were currently studying the language at time of this research; some no longer studied the language, some had graduated, and others no longer studied the language in the formal sense but spoke Chinese either in their personal life or at their job. Though perhaps not relevant, it would have been prudent of me to ask participants to estimate when they last spoke Chinese or the frequency they used the language. The total time spent learning Chinese was not recorded due to the varying situations and difficulty in specifying time spent learning the language as some heritage speakers did not necessarily consider their time speaking with family to be study, and other speakers had non-consecutive periods of learning the language of using the language.

Non-native speakers (including heritage speakers) were asked to self-assess their speaking ability in Chinese based on the ILR scale (Interagency Language Roundtable, n.d.). The scale runs 0–5 with 0 being no proficiency and 5 being native proficiency, with the options of each number having a "+" level (e.g. 0+, 1+, 2+ and so on, though level 5

does not have a 5+). For the purposes of data, the ILR has coded the levels with a point-value system of 0–50, which I will adopt here. The whole numbers correspond with a value of the level times 10 (i.e. a level 2 would code as 20, and a level 0 would code as 0), and the L+ would code as the level times 10 and plus 6 (i.e. level 4+ would code as 46, and level 0 would code as 6). It is worth noting that the ILR scale measures communicative ability and accuracy in terms of grammar and breadth of vocabulary, *not* fluency per se, in that it does not specify docked points for fluency breakdowns, pronunciations, speech rate, and so on. It is also relevant to note that while this was phrased as a self-assessment, some participants (to varying degrees of recency) had actually taken proficiency and aptitude tests whose results were presented according to the ILR scale. Who had received ILR scores and who had self-assessed their own scores were not noted in the data.

Table 3
Study Abroad Demographics

Location	N. Participants	Time	
		Mean	S.D.
Mainland	*20*	*4.99*	*4.16*
Beijing	5	2.2	0.45
Hainan	1	3	
Hangzhou	1	5	
Nanjing	1	7.67	3.77
Xiang/Shanxi Sheng	3	12	
Shanghai	4	4.5	3.79
Sichuan	1	2	
Suzhou	1	1	
Xi’an	2	6.38	7.96
Xiang	1	12	
Taiwan	*8*	*2.67*	*1.41*
Tainan	1	1	

Taipei	5	3.2	1.64

Speaker	Time	
	Mean	S.D.
Heritage	11	7.55
Non-native	4.92	3.85
Total	5.88	4.87

N. Participants in the first half of the table refers to the number of participants who studied at a particular location. Mainland and Taiwan represent the total number of participants who studied in those locations. Time represents how long participants studied abroad in months, and it is represented in Mean and Standard Deviation. The lower half of the table is speaker-specific and lists the Mean and Standard Deviation of time in months each group of speakers spent studying abroad.

Heritage and non-native speakers were additionally asked if they had spent any time abroad and if so, where and for how long. All participants had done so, and the results are presented above (Table 3). The table is split into two halves: the first concerns where the participants studied abroad, and the second half concerns speakers on a more individual level. The time is presented in months as many had done summers abroad. Many of these times are estimated periods as several participants listed numbers that varied by one month for the same study abroad program (e.g. participants would report 2 month and others 3 for a summer program). Several participants studied abroad multiple times, hence why the total number of participants exceeds the 19 speakers for the first half of the table. Notably, 3 participants did not specify a city and simply said "Taiwan," thus, the total number of participants who studied in Taiwan includes the 6 participants who studied at Tainan and Taipei, as well as the 3 participants who simply listed "Taiwan."

Table 4

Native Speaker Dialects

1st Dialect	N. Participants

普通话 *Pǔtōnghuà*	3
北京话 *Běijīnghuà*	1
北方话 *Běifānghuà*	1
四川话 *Sìchuānhuà*	1

2nd Dialect	N. Participants
广东话/粤语 G*uǎngdōnghuà/ yuèyǔ*	2
普通话 *Pǔtōnghuà*	2
None	2

N. Participants refers to the number of participants who speak a particular dialect. The top half of the table concerns the speaker's 1st dialect(s); some speakers listed more than one 1st and 2nd dialects, hence where there are a total of 6 speakers despite the sample size only having 5 native speakers.

Of the native speakers (average age: 26.6), participants identified their native Chinese dialect in various ways. 1 participant identified 普通话 *Pǔtōnghuà* as their native dialect, 1 responded 北京话 *Běijīnghuà*, 1 said Mandarin Chinese, 1 responded 四川话 *Sìchuānhuà*, and the last reported both 普通话 *Pǔtōnghuà* and 北方话 *Běifānghuà*. As mentioned in Xia (2017), 普通话 *Pǔtōnghuà* is the language of education, and is based off of the Beijing dialect (aka 北京话 *Běijīnghuà*) (van Heuven & Tang, 2009), and 北方话 *Běifānghuà* meaning "Northern dialect" also belongs to the Mandarin supergroup (Wurm et al., 1988). Lastly, 四川话 *Sìchuānhuà* or "Sichuanese" is spoken in the region of Sichuan of which Cheng-du is the capital, and is classified as part of the Southwestern Mandarin subsection, and is considered mutually intelligible with other Mandarin dialects, though not perfectly so; van Heuven and Tang (2009) found Beijing speakers were able to correctly perceive 62% of words spoken by 四川话 *Sìchuānhuà*

while 四川话 *Sìchuānhuà* speakers were able to accurately hear 98% of words spoken by *Běijīnghuà* speakers. Furthermore, both speakers who listed 四川话 *Sìchuānhuà* and 北方话 *Běifānghuà* amongst their 1st dialects also included 普通话 *Pǔtōnghuà* as their other 1st dialect. Under this analysis, the five native speakers of Chinese were determined to be acceptable representatives of Mandarin Chinese.

3.2.2 Rater Participants

A total of 14 participants provided full biographical information, and of those, 12 completed the entirety of the fluency assessment (2 did not complete the entire section of the fluency assessment). The general biographical information is presented in Table 5. The age of the raters ran the gamut from early 20s to 50s, though one participant listed "00" as their age—this is assumed to be a typo and was removed for the purposes of calculating Mean and Standard Deviation. The raters were also asked in which country they grew up, all of whom grew up in either China, Taiwan, or in one of the aforementioned *and* the US.

Table 5
Rater Demographics

Raters	Age		Gender		Country		
	Mean	S.D.	N. Females	N. Males	N. China	N. Taiwan	N. US
	32.85	13.01	8	6	9	5	2

The above table presents the age (in years) of raters in Mean and Standard Deviation. N. Females and N. Males report the number of participants who identified as the above genders, and N. China, N. Taiwan, and N. US refers to the number of participants who reported growing up in the aforementioned countries. Some reported growing up in two countries, hence the total exceeding 14.

Raters were furthermore asked to report their native language(s) (Table 6), native dialect(s) (Table 7). Though Mandarin, Taiwanese, and Cantonese are all branches of the

Chinese family, the table represents how the raters chose to identify their first language(s). Asking raters to identify their native dialect (Table 7) was an attempt to further specify what variety of Chinese raters spoke. I categorized these answers into the language branches (e.g. Mandarin, Min, Wu, etc) and further subsets for larger groups (e.g. one respondent reported their native dialect was *Shāndōng* dialect, which belongs to the Jiaoliao branch of the Mandarin supergroup, hence it was counted amongst the Jiaoliao branch).

Several raters reported their native dialect was "Taiwanese," which posed some problems as this phrase can refer to the Taiwanese variety of Mandarin, the Taiwanese variety of Hokkien (a subset of the Min and specifically Southern Min language branch), or the Taiwanese variety of Hakka (which in and of itself constitutes its own language branch in the Chinese language family). Based on the raters' responses for their native language, native dialect, and total dialects spoken, I was able to determine that one rater most likely mean Taiwanese Mandarin, while the other three meant Taiwanese Hokkien or Taiwanese Hakka. Hakka is the language of the Hakka people, an indigenous ethnic group in South East Asia and, in this case, specifically Taiwan; approximately 4 million Hakka people currently live in Taiwan, comprising 15–20% of the country's population (Hakka Affairs Council in Taiwan, 2019). Hokkien meanwhile is a subset of the Southern Min language branch (part of the larger supergroup of Min Chinese), and has been growing in popularity in Taiwan, reaching speaker population equal to 71% of the country (Wu, n.d.). Because there are nearly 3 to 4 times more Hokkien speakers than Hakka speakers in Taiwan, I erred on the side of surmising the other three "Taiwanese" respondents referred to Taiwanese Hokkien (hence why "Taiwanese" is reported under

the Min branch in Table 7); however, it is entirely possible that the respondents could have meant Taiwanese Hakka, in which case Table 7 is not representative of the native dialect. That being said, of the three respondents who responded "Taiwanese" (but not Taiwanese Mandarin), each one reported being bilingual and 2 of the 3 reported Mandarin Chinese as a native language (the 3rd respondent listed "Chinese" as their other native language).

Table 6
Rater Native Language

Reported Language	N. Participants	Perc Mono	Perc Multi
Chinese	6	.83	.17
Mandarin	8	.625	.375
Taiwanese	3	0	1.0
Cantonese	1	0	1.0

The above languages are pulled from respondent's own self-identified first language(s). N. Participants represents how many participants reported the language as one of their first languages. Because some participants identified themselves as bilingual, Perc Mono and Perc Multi represent the percentage of respondents for that language who identified themselves as a mono or bilingual.

Table 7
Rater Dialects

Language Branch	N. Part	Branch Dialect	*Perc Branch*	Perc 1st Dialect	Perc 2nd Dialect	Perc Mono	Perc Multi
Mandarin	10			.42	.58	.55	.45
		"Mandarin"	*.3*	.8	.2		1.0
		Jiaoliao	*.2*		1.0	1.0	
		Jilu	*.3*		1.0	1.0	
		Beijing	*.1*		1.0	1.0	
		Jianghuai	*.1*	1.0			1.0
Min	4			1.0			1.0
		S. Min	*.25*	1.0			1.0
		Taiwanese	*.75*	1.0			1.0
Wu	1	Shanghai	*1.0*	1.0			1.0
Yue	1	Cantonese	1.0	1.0			1.0
Xiang	1	Changsha	1.0	1.0			1.0

Hakka	1	T. Hakka	1.0		1.0		1.0
"Chinese"	2		1.0	.5	.5	.5	.5
Other	2						1.0

This table details the dialects raters reported being able to speak. Language Branch categorizes the responses into the main Chinese language while Branch Dialect specifies the subset of the language branch (see Section 2.3). N. Participants is the total number of raters who can speak each language branch, Perc Branch is the percentage of the language branch who speaks that specific dialect, Perc 1st Dialect is the percentage of participants who speak that dialect and language branch as their native dialect while Perc 2nd Dialect is the percentage who speak that as a non-native dialect. Perc Mono and Perc Multi details the percentage of each language branch and dialect who are monolingual/monodialectal in Chinese and the percentage who are multilingual/multidialectal in Chinese.

The below table reports the raters' experience in foreign language contexts, either as being in a country that spoke a language other than Chinese (the Abroad section of Table 8) or as a teacher of Chinese as a foreign or second language (the Teaching Experience of Table 8). Of the raters who reported having been abroad, all had gone to the US. This section does not note whether the rater was currently abroad or not, simply whether they had gone abroad before. Additionally, teaching experience does not distinguish between people who served as teachers in the past and people who currently teach. Lastly, raters were asked to rank their familiarity with English-accented Chinese on a scale of 1-9, with 1 being not at all familiar, and 9 being extremely familiar. Though one respondent reported an unfamiliarity with English-accented Chinese (3), most reported a moderate to strong familiarity.

Table 8
Rater Linguistic Experience

	N. Participants	Time	
Abroad		Mean	S.D.
Never Been Abroad	4		
Been Abroad	10	9.33	7.94
	N. Participants	Time	

Teaching Experience		Mean	S.D.
No Teaching Experience	6		
Teaching Experience	8	6.83	7.14
	Mean	S.D.	
Familiarity	7.14	2.00	

The above table is split into three sections that details the raters' experience in foreign language contexts. N. Participants reports the number of raters who have or have not been abroad and the number of participants who do or do not have experience teaching Chinese. For those with abroad or teaching experience, the length in years of that experience is reported in Mean and Standard Deviation.

3.3 Data Collection

The 24 speakers provided stimuli for the first experiment and speech samples for the second experiment. The speakers were provided instructions in both English and Chinese and were asked to complete a series of three tasks: to answer several questions pertaining to their background (see section 3.2); to read three sentences in Chinese aloud, which provided the stimuli for Experiment 1; and to answer in Chinese a question about themselves (namely *what is your major?* in the style of Kahng, 2018), which would serve as the data for Experiment 2.

For the read-aloud task necessary for Experiment 1, the three sentences were selected by me from various Chinese textbooks (see Appendix D) and were chosen for their syntactic sentence structure. The sentences were provided in simplified characters, traditional characters, and pinyin, with an English translation provided. For this section, I was not testing reading ability; I was simply attempting to gather stimuli for Experiment 1. As such, participants were encouraged to practice reading the sentence as many times as they would like before they were satisfied that their reading was the most fluent

rendition they could give. I then manipulated each sentence in Praat (Boersma & Weenik, 2020) by normalizing the sample to 70 dB before manually removing every pause (filled or unfilled) greater than 70 ms as 70 ms is the average of the two lowest boundaries for pause duration that I have found cited in the literature (≥60 ms in Campione & Véronis, 2002; >80 ms Levin, Silverman, & Ford, 1967). Of the lowest boundaries for pauses, most cite 100 milliseconds as being the lowest threshold (Kang, Rubin, & Pickering, 2010; Riazantseva, 2001); opting for a threshold just lower than the previously typical threshold decreased the chances of a pause being perceived. I then went in and added pauses to the sentences; a control pause of 300 ms was added to Sentence 1 and Sentence 3 at clausal boundaries as previous literature has shown that the effect of pauses between clausal boundaries as opposed to pauses within clausal boundaries are perceived as being more fluent (Kahng, 2018). A shorter pause of 250 ms was added to randomly assigned to syntactic placements that corresponded with either between-phrase or within-phrase syntactic junctures (see Appendix D for the sentences and their pause location); this served as the independent variable. The reason for 250 ms and 300 ms pauses is because previous literature reports that this durational range reports the highest correlation between pause and L2 fluency (de Jong & Bosker, 2013), thus the control and weightier syntactic point of between-clause was given a 300 ms pause, and the independent variables of pause between-phrase and pause-within phrase were given pause durations of 250.

These specific syntactic junctures at which pauses were placed were, by and large, distinguished by whether or not they occurred between a Head and its Complement; those that did were labeled as “within-phrase” pauses. Appendix D details the various locations

that were coded as either between or within phrases. Due to the number of phrase structures, participants were not all assigned the same pause location. However, I did attempt to equally distribute the speakers. The between-phrase condition received a random assortment of 8 native speakers, 3 heritage speakers, and 25 non-native speakers for a total of 36 samples; and the within-phrase condition received a total of 7 native speakers, 6 heritage speakers, and 23 native speakers for a total of 36 samples. Though the heritage speakers did have an disparate number between the two conditions, this was done to try and control for native speaker disparity. Heritage and non-native speakers together represented 28 of the 36 speakers for the between-phrase condition with 8 native speakers; the two groups together constituted 29 of the 36 speakers for the within-phrase condition with 7 native speakers. Thus, the between-phrase condition had 1 more native speaker than the within-phrase (a total of 8 as opposed to 7), 3 fewer non-native speakers (a total of 3 as opposed to 6), and 2 more non-native speakers (a total of 25 as opposed to 23).

The Random Gauss function in Praat was then overlaid over the samples to provide a white-noise background to obscure that the audio had been manipulated (Kahng, 2018), such abrupt silences absent of even the white noise that had been present in several of the recordings due to the nature of the audio being gathered over Zoom. Because the audio samples were collected via Zoom and not in a controlled environment like a recording studio, the Random Gauss decibel ranged from 50dB to 55dB in an effort to mask the audio such that none could tell the samples had been manipulated.

These samples were then presented to native speakers via an online survey. This survey asked rater participants to provide some background information (see Section 3.2)

before listening to audio samples of Chinese and assessing for fluency. Raters were asked *not* to pay attention to features like pronunciation, grammar, or vocabulary and were instead asked to focus on features such as speech rate, filled and unfilled pauses, hesitations or corrections, and overall flow of speech (Tavakoli & Skehan, 2005), and based on these criteria, assess the fluency of the speaker on a scale of 1 "Extremely disfluent" to 9 "Extremely fluent." The audio samples were presented via PhonicAi online survey (Phonic Inc, 2021) in a double-blind randomized order using the platform's randomization effect (even I do not know in what order participants heard the audio samples). The entire survey took approximately 35 minutes to complete, and rater participants were afterwards emailed their compensation.

CHAPTER 4

EXPERIMENT 1: RESULTS & DISCUSSION

4.1 Experiment 1: Results

Overall, pause location was not found to be significantly correlated with fluency ratings (Fig 1), though analysis of the distribution of fluency ratings (Fig 5) and the mean scores (Fig 6) shows a visible if not statistically significant effect of pause location amongst heritage and non-native speakers.

This experiment had two independent variables to account for: the first being the independent variable of interest to this study, namely where the pause occurs syntactically (either between or within a phrase), the second being an independent variable that needs to be accounted for, specifically the speaker themselves who were either native, heritage, or non-native speakers. To account for this, a 2 Way ANOVA in SPSS version 27 was used to calculate the effect of the two independent variables upon the dependent variable of fluency rating. Because a 2 Way ANOVA requires the dependent variable be scalar rather than ordinal as a Likert scale is, I calculated the mean fluency rating for each speech sample to create scalar data (Statistics Solutions).

With speaker status (native, heritage, and non-native) and pause placement (between clause and within clause) as between-subject factors, the 2 Way ANOVA showed the main effects of speaker status ($f(2, 66) = 32.336$, $p < 0.001$) and of pause placement ($f(1, 66) = 0.326$, $p = 0.570$). The interaction between these two independent variables did not reach statistical significance ($f(2, 66) = 0.156$, $p = 0.856$) (Fig 1). That is to say, pause location did not have a statistically significant correlation with fluency

rating, but whether the speaker was a native, non-native, or heritage speaker did have a statistically significant correlation.

Figure 1

Tests of Between-Subjects Effects

Dependent VariableFluency Rating

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	73.193[a]	5	14.639	13.135	<.001
Intercept	2605.067	1	2605.067	2337.439	<.001
Speaker	72.077	2	36.038	32.336	<.001
PauseLoc	.364	1	.364	.326	.570
Speaker * PauseLoc	.347	2	.174	.156	.856
Error	73.557	66	1.114		
Total	3834.519	72			
Corrected Total	146.750	71			

a. R Squared = .499 (Adjusted R Squared = .461)

It is perhaps unsurprising that whether a speaker is a native speaker or non-native speaker has a strong correlation with their fluency rating as the former are seen as ipso facto fluent (Davies, 2003). What may be a tad more surprising, but still not unexpected, is that heritage speakers were significantly correlated with fluency rating (Fig 2). A Tukey post-hoc test showed native and heritage speakers were not statistically different in their fluency ratings, but that non-native speakers differed statistically from both native and heritage speakers in terms of fluency scores. To verify this, a 2 Way ANOVA was re-run with only heritage speakers and non-native speakers as groups within the independent variable of speaker. Though p-value of pause location unexpectedly rose and reached 0.856, the new analysis did show that speaker status (heritage or non-native) was significantly correlated with fluency ratings (Fig 3).

Figure 2

Multiple Comparisons

Dependent VariableFluency Rating

	(I) Speaker (NS, HS, NNS)	(J) Speaker (NS, HS, NNS)	Mean Difference (I–J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	Native Speaker	Heritage Speaker	.36097	.445121	.698	-.70630	1.42824
		Non-native Speaker	2.24720*	.312279	<.001	1.49845	2.99595
	Heritage Speaker	Native Speaker	-.36097	.445121	.698	-1.42824	.70630
		Non-native Speaker	1.88623*	.383473	<.001	.96678	2.80569
	Non-native Speaker	Native Speaker	-2.24720*	.312279	<.001	-2.99595	-1.49845
		Heritage Speaker	-1.88623*	.383473	<.001	-2.80569	-.96678
Bonferroni	Native Speaker	Heritage Speaker	.36097	.445121	1.000	-.73248	1.45442
		Non-native Speaker	2.24720*	.312279	<.001	1.48008	3.01432
	Heritage Speaker	Native Speaker	-.36097	.445121	1.000	-1.45442	.73248
		Non-native Speaker	1.88623*	.383473	<.001	.94422	2.82824
	Non-native Speaker	Native Speaker	-2.24720*	.312279	<.001	-3.01432	-1.48008
		Heritage Speaker	-1.88623*	.383473	<.001	-2.82824	-.94422

Based on observed means.
The error term is Mean Square(Error) = 1.114.

*. The mean difference is significant at the .05 level.

Figure 3

Tests of Between-Subjects Effects

Dependent VariableFluency Rating

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	28.055[a]	3	9.352	6.793	<.001	.278	20.378	.967
Intercept	1510.701	1	1510.701	1097.306	<.001	.954	1097.306	1.000
Speaker	26.072	1	26.072	18.938	<.001	.263	18.938	.990
PauseLoc	.650	1	.650	.472	.495	.009	.472	.104
Speaker * PauseLoc	.012	1	.012	.009	.925	.000	.009	.051
Error	72.967	53	1.377					
Total	2698.566	57						
Corrected Total	101.022	56						

a. R Squared = .278 (Adjusted R Squared = .237)

b. Computed using alpha = .05

While raters did have a relatively strong degree of agreement on fluency ratings for native and even heritage speakers, the fluency ratings for non-native speakers showed a much greater variation (Fig 4 and 5). Recalling that the data points are all averaged scores across the 12 to 14 respondents, it is easy to see that native speakers had a relatively high and similar fluency score: highest rating of 9 for both between- and within-phrase pause, lowest rating of 8.417 for between-phrase and 8.538 for within-phrase, and a standard deviation of 0.213 for both between- and within-phrase pause.

Heritage speakers also had a relatively high and similar fluency rating (Fig 5), albeit to a lesser extent. The highest rating was a 9 for between-phrase pause and 8.917 for within-phrase; the lowest rating of 8.333 for between-phrase pause and 7.417 for within-phrase pause; and the standard deviation was .371 and .616 for between- and within-phrase pause respectively. Non-native speakers, however, varied quite a bit on scores. The highest fluency rating for between-phrase pause was 8.583 and the lowest was 4.167, making the standard deviation 1.238; the highest for within-phrase pause was 8.692 and the lowest was 4.750, making the standard deviation 1.479.

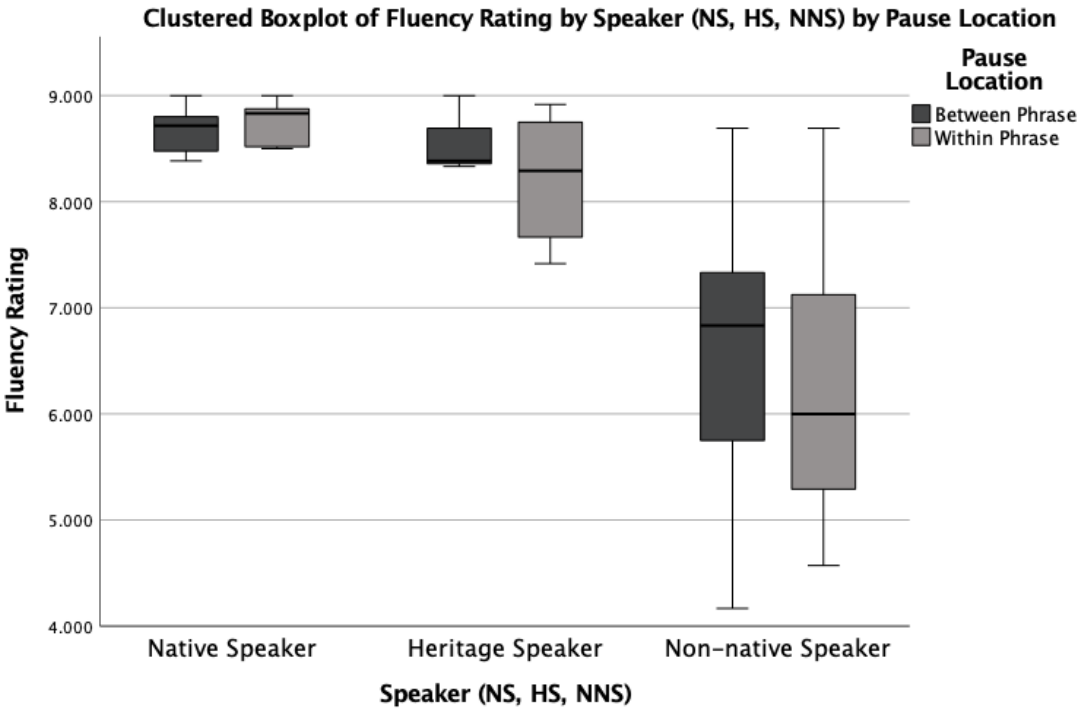
Figure 4

Descriptive Statistics

Dependent VariableFluency Rating

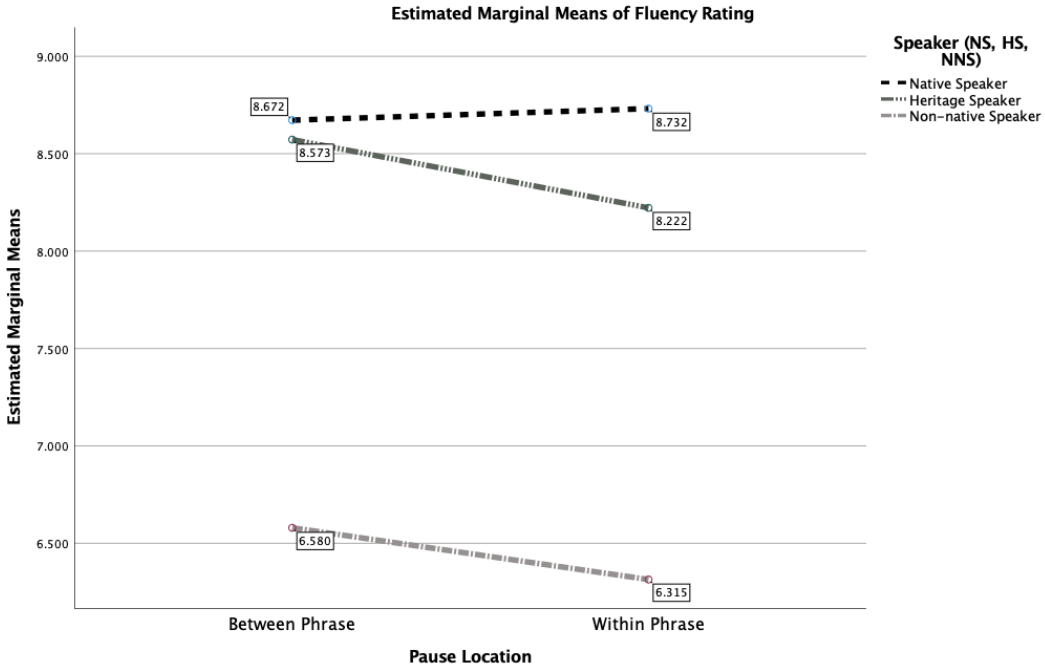
Pause Location	Speaker (NS, HS, NNS)	Mean	Std. Deviation	N
Between Phrase	Native Speaker	8.67228	.213195	8
	Heritage Speaker	8.57265	.370983	3
	Non-native Speaker	6.58000	1.237627	25
	Total	7.21100	1.413745	36
Within Phrase	Native Speaker	8.73168	.212747	7
	Heritage Speaker	8.22222	.616141	6
	Non-native Speaker	6.31454	1.243755	23
	Total	7.10249	1.479234	36
Total	Native Speaker	8.70000	.207521	15
	Heritage Speaker	8.33903	.549887	9
	Non-native Speaker	6.45280	1.234588	48
	Total	7.15675	1.437673	72

Figure 5



Despite all this, we can observe that though the effects may not be statistically significant, there is an observed difference between pause location and fluency rating for non-native and heritage speakers wherein a pause within a phrase is judged less fluent than a pause between phrases (Fig 6). This does not by any means prove a correlation as the 2 Way ANOVA shows quite strongly there is not a statistically significant effect, but it does provide the possibility that pause location may affect fluency in some capacity (this will be discussed further in Section 5.2).

Figure 6



4.2 Experiment 1: Discussion

To recap, the research question for Experiment 1 was: does pause placement (between phrases or within phrases) affect perceptions of fluency? The null hypothesis was stated as: pause placement (at the phrasal level) does not affect perceptions of fluency. Based on the findings, I fail to reject the null hypothesis and found that there is no correlation between pause placement (at the phrasal level) and perceptions of fluency. The following segment discusses in detail the results and interpretations to explain the above statement.

As already mentioned, whether the speaker is a native, non-native, or heritage speaker is significantly correlated with fluency scores with native and heritage speakers being significantly different from non-native speakers, but not significantly different from each other. We can observe from the data (Fig 5) that even the highest of the non-

native speakers could not reach native-level proficiency. Furthermore, we can observe that heritage speakers who had a pause within-phrases as opposed to between-phrases were rated lower; and the lower half of the heritage speaker scores all came from the between-phrase condition. However, there were a limited number of heritage speaker participants, so this observation should be taken with a grain of salt. Lastly, the mean of heritage speakers (regardless of pause location) was lower than native speakers, and same for non-native speakers to heritage speakers (Fig 6).

Why there were no statistically significant findings for the effects of pause location on perceptions of fluency could be due to several reasons. To start, the lack of findings could be due to the small sample size as only 12–14 responses per stimulus; a larger sample size may have yielded statistically significant results. Another issue could be that the pause effect was not great enough. In my experiment, I had inserted only 1 pause at either a between- or within-phrase juncture for each sentence. It may be that in order to have a noticeable effect, there must be a greater threshold of frequency that needs to be reached. Previous research that found statistically significant results with pause location had included 5 pauses per sentence (Kahng, 2018). If this is the case, it would suggest that pause location alone is not a strong enough effect to be observed without being fortified by an increased pause frequency. A third reason could simply be that there is not a significant difference in how we judge fluency based on a pause between- or within-phrases. Though findings have shown that pauses within clauses are less fluent than pauses between clause (Chen, 2015; Lambert, Kormos, & Minn, 2016; Saito et al., 2017; Kahng, 2018; Shea & Leonard, 2019; Kahng, 2020; Suzuki & Kormos,

2020), listeners may not differentiate between types of pause within a clause (namely between- or within-phrases).

Though the findings were not statistically significant, I would like to take a moment to address the data as it appears. We can observe from Figure 4 that though mean fluency rating for heritage speakers in both pause conditions (8.572 for between and 8.222 for within), the effects of the within-phrase condition may have had a greater likelihood of decreasing fluency scores as the lowest within-phrase pause score (7.417) was nearly a full point lower than the lowest between-phrase pause score (8.333). This could suggest that while there may be other acoustical and fluency features that may overcome the effects of pause location, a pause between a phrase is less likely to be observed as disfluent as compared to a pause within a phrase. However, this theory has the caveat that due to randomization, there were only 3 heritage speakers for the between-phrase pause group. This is an incredibly small sample size, and a larger study would need to address this effect.

This observation was not seen in the non-native speaker pool. The highest fluency rating for the two groups was nearly identical (8.583 and 8.692 for between- and within-phrase pause), and the lowest rating for the two groups unexpectedly favored the within-phrase pause condition by over half a point (4.750 for within-phrase and 4.167 for between-phrase). This again may be due to the idea that while pause location may have a more aggressive disfluent effect depending upon where it occurs, pause location alone is not enough to sway fluency judgements to the point of overcoming speech rate, pause frequency or duration, or other such factors. Varying lower and upper bound thresholds aside, the mean fluency rating dropped from 6.580 in the between-phrase group to 6.315

in the within-phrase group. This is a mild but noticeable change, and one that may become more present if more pauses are added to a sentence in order to observe the effect.

All this being said, we must also discuss the paradoxical *increase* in fluency of native speakers in the within-phrase pause group as compared to the between-phrase group. The increase was incredibly small, less than 1/10 of a point from 8.732 to 8.672, but it does still run against expectations. That the native speaker population trends differently than the heritage and non-native speaker populations suggests that there are other fluency factors affecting this population at these high levels of fluency.

To summarize, this study did not find statistically significant correlations between fluency ratings and pause location as it set out to do. While there is a minute but observable effect of pause location in heritage and non-native speaker populations, this research cannot statistically show that a pause within a phrase is perceived as more disfluent than a pause between phrases. That being said, I do believe that due to the limitations of this study (i.e. small sample size) and short-comings of this research design (i.e. the inclusion of only one condition pause per each auditory stimulus) lead to inconclusive and statistically insignificant results. However, the observable effects in Figure 5 suggest that re-testing the effects of between- and within-phrase pauses on perceptions of fluency is worthwhile as I do not believe this study conclusively showed that there is no correlation between the two. Rather, I believe the study's own limitations failed to provide enough evidence to either support or refute a correlation between fluency ratings and pause location between and within phrases.

CHAPTER 5

EXPERIMENT 2: METHODOLOGY

5.1 Recruitment & Participants

Participants for Experiment 2 were gathered from the same speaker pool as in Experiment 1. As part of their tasks, speakers had been asked to answer an open-ended prompt about their major in order to elicit a spontaneous speech sample (Kahng, 2018). Though 24 responses were gathered for this prompt; time and resource constraints meant only ten could be analyzed. Rather than splitting the three speaker groups as three, three, and four, five non-native speakers were selected at random, as were three native speakers and two heritage speakers. This was done in part because the results of Experiment 1 showed that there was no significant difference in fluency ratings of native and heritage speakers, thus I opted for an equal distribution of five non-native speakers against a collection of five native and heritage speakers.

The ages of the speakers ranged from 21 to 33 with an average of 24.4 years, and a near even distribution of men and women.

Table 9
Speaker Participant Demographics—Experiment 2

Speaker	Age		N. Females	N. Males	Total Subj.
	Mean	S.D.			
Native	28.67	4.04	1	2	3
Heritage	21.5	0.71	2	0	2
Non-native	23	1	3	2	5
Total	24.4	3.63	6	4	10

Speaker refers to whether the speaker is a native, heritage, or non-native speaker. Age is presented in Mean and Standard Deviation in years. N. Females and N. Males represent the number of each gender present for the three speakers, while Total Subj. represents the number of subjects for each speaker group.

Excepting one non-native speaker whose native language was Spanish and one who spoke Cambodian as one of their two native languages (the other being English), all the other non-native and native speakers spoke English as their native language. As with the sampling from Experiment 1, the self-reported ILR score (see section 3.2.1 for discussion on ILR scale) varied by approximately 10 points (equivalent to one level) between the heritage and non-native speakers.

Table 10
Non-Native and Heritage Speaker Language Experience—Experiment 2

Speaker	Age at Start of Learn		Time at College		ILR	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Heritage	16	4.24	3.5	0.71	31	7.07
Non-native	14.2	1.64	3.4	0.89	21.6	4.34
Total	14.71	2.36	3.43	0.79	24.29	6.47

Age refers to the age (in years) at which the speaker began learning Chinese in a formal setting (re: in a class). Time at college refers to the number of years the speaker studied or has studied the language in college. ILR refers to the speaker's self-assessment of their speaking ability in the language according to the Interagency Language Roundtable.

Regarding native speakers, two of the three were monodialectal Mandarin speakers who spoke Mandarin Chinese (普通话 *Pǔtōnghuà* one speaker specified) as their native and sole Chinese dialect. The other native speaker spoke 四川话 *Sìchuānhuà* as their native dialect, with Mandarin Chinese as their second dialect. As mentioned in Section 3.2.1, the Southwestern Mandarin variety of 四川话 *Sìchuānhuà* is largely mutually intelligible with other Mandarin dialects (van Heuven & Tang, 2009).

5.2 Data Collection

5.2.1 Gathering the Speech Samples

The question about themselves was to prompt the participants to provide spontaneous speech samples. Participants were asked to aim for about 60 seconds of

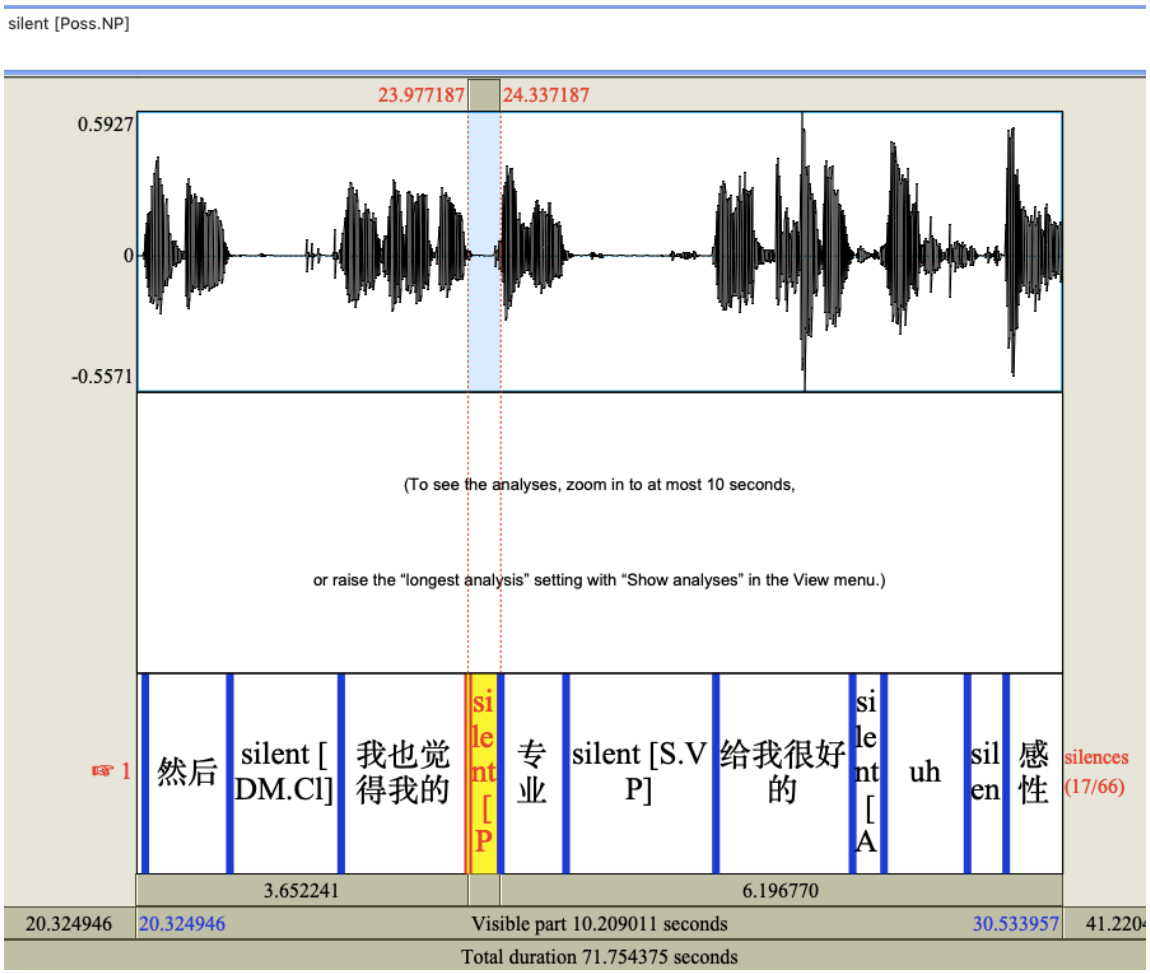
speech (I would give them a thumbs-up to signal the time was reached) and were allowed to think about their response for as long as they'd like. They were also allowed to decide at the end of their response if they were satisfied with their answer or whether they'd like to try again. The reason I allowed this was to encourage participant turnout. Several participants in my study were not willing to participate unless they had some control over their final speech sample. Allowing participants to re-record their responses with me (but *not* allowing participants to script a response) was a means through which I tried to still gather spontaneous speech samples from all participants. Furthermore, I was attempting to gather authentic speech samples, not necessarily speech samples that displayed nervousness or jitters from being recorded.

5.2.2 Analyzing the Data in Praat

For this experiment, I was focusing on researching the likelihood of a pause occurring at a particular syntactic juncture. To gather this data, I transcribed the spontaneous speech samples of 10 speakers (3 native speakers, 2 heritage speakers, and 5 non-native speakers) using Praat's Annotate to Textgrid (silences) function with a Minimum Pitch set to 60 Hz and the minimal silent interval duration set to 0.25 s (Southee, 2020) as 250 ms has been found to be the optimal minimal silent pause duration that best correlates with fluency measures (de Jong & Bosker, 2013). I then went through to verify that Praat's annotated silent pauses were indeed silent, and that the "pause" hadn't mistakenly captured a sustained sonorant or quieted speech. After correcting any mislabeled silences, I transcribed the speech data and defined where each

pause occurred syntactically[4] (Fig 7). Pauses that were preceded, followed, or interrupted by filled pauses were recorded as a single silent pause for the purposes of this experiment (Kahng, 2014).

Figure 7



The above figure shows a sample of how I annotated and identified pauses in Praat. This segment was selected as it does not give identifying information about the participant who provided this speech sample. Translated, the sentence reads *Then (silent [DM.Cl]) I also think my (silent [Poss.NP]) major (silent [S.VP]) gives me really good (silent [Adj.NP]) (uh) (silent) feelings.* DM stands for discourse

[4] When transcribing the samples and labeling where pauses occurred syntactically, I did my best to ignore prescriptive grammar in favor of syntactic grammar. That is to say, though someone may have been grammatically "wrong," their utterance may still have had a clear syntactic structure. For example, "She think good story" is grammatically incorrect because "think" takes a CP as its Complement, not a DP, and because "story" cannot function as a bare noun in English without a determiner of some capacity. That being said, one can still understand that "she" operates as the subject, "think" as the verb, and "good story" as the Complement, even if the sentence is grammatically flawed and the meaning unclear.

marker, under which I included sentential adverbs like "then" (note, 然后 is not operating as temporal adverb here, more as a narrative adverb that is connecting two thoughts), Cl means clause, S means subject, Adj adjective, and VP verb phrase. Under my analyses, the run of pauses *(silent [Adj.NP]) (uh) (silent) would all be counted as one single pause.*

It is worth noting that though my study is on silent pauses, for this particular experiment, I did note in my data filled pauses that occurred *without* silent pause. While I originally kept these filled pauses separate in my original gathering of the data, with the exception of one solo filled pause, all filled pauses occurred in syntactic locations where the speaker had paused silently.

The co-occurrence of filled pauses with silent pauses is not surprising and is well documented in the literature (Stenström, 1990 as cited in Stenström, 2011; Swerts, 1998; O'Connel & Kowal, 2008). Though filled and silent pauses carry different sociolinguistic connotations—filled pauses can serve as a bid to hold the floor in the Discourse Analysis paradigm (Belz & Reichel, 2015), though the need to hold the floor is a little less pressing in a one-way interview—previous research has shown that neither filled nor unfilled pauses are significantly correlated with whether the speaker is a native of non-native speaker. In other words, one's speaker status does not predict the occurrence of filled pauses nor silent pauses, and that both are significantly correlated with pause location between and within an Analysis Speech Unit (essentially a major constituent boundary) (de Jong, 2016) with some studies suggesting a filled pause unaccompanied by silence is more likely to appear phrase medially (Swerts, 1998). Because of this and because my noting of silent pauses did not differentiate between a truly silent pause and a pause that included periods of silence and hesitation sounds (e.g. *um, uh*), I opted to ultimately include these isolated filled pauses in my analysis as they speak to disfluencies through

an interruption of speech. In total, 19 of 230 total pauses were filled pauses that occurred without a contiguous silent pause, representing 8.26% of the pauses.

Additionally, to make the task of coding (see Section 3.2.2.3) simpler, I did not count pauses that occurred as part of restarts and reformulations because reformulations in particular change the syntactic environment.

5.2.3 Coding the Data

Pauses recorded, labeled, and counted, I then needed to develop a coding system to categorize and group the pauses. Previous research that has looked at pause location has only looked at it binarily: either between or within clauses, major constituent boundaries, Analysis Speech Unit, and similar large-branching differences that distinguish pauses at the clausal level and within the clausal level (de Jong, 2016; Kahng, 2018; Shea & Leonard, 2019). Ultimately, I wanted to look at pauses with a finer toothed comb, so I created a more complex coding system than the previous binary between-within clause one. This system was developed using Generative Grammar as an underlying base for its model and reasoning.

In total, I organized the pauses into 10 categories: 1) between clauses, 2) between a discourse marker or a sentential adverb and the following clause that it precedes, 3) between coordinated phrases (this includes phrases that were coordinated without a conjunction being present), 4) between the Specifier and the VP (for simplicity's sake, the TP is not noted as Chinese does not have a Tense system), 5) between a Head and its Complement, 6) between a Modifier and its noun Head, 7) between a degree adverb and its corresponding Adjective or Verb Stative, 8) between a Modifier and its verbal Head, 9) imparsable/unintelligible pauses, namely pauses that occurred in environments where

what was said was unclear, or pauses that occur in locations that are either grammatically correct or have uncertain interpretations, and lastly 10) pauses that occur in miscellaneous locations, which were largely in connection to discourse markers.

I have several reasons for this particular categorization, which I will detail below. For one, between-clauses was counted as a pause that follows the end of a clause and the start of another, regardless of whether that next clause begins with a subject, a discourse marker, a sentential adverbial (including conjunctive adverbs like "however," temporal adverbials like "during the meeting, he took studious notes" or "while attending high school, she was enrolled in six AP courses," and prepositional adverbs situated at the sentential level like "At home, I relax and watch TV"), a clausal coordinating conjunction (such as 但是 *dànshì* "but"), or a clause adverbial of reason or a temporal clause adverbial (e.g. "I went the store because I needed milk" provides the reason, or "My dog woke up when she smelled bacon" provides the *when* adverbial). For simplicity's sake, "sentential adverbial" will henceforth include all sentential adverbials discussed above, as well as clause adverbials and clausal coordinating conjunctions. The term "discourse marker" can reasonably apply to all of the above as discourse markers refers to words or phrases that can serve one of four functions—temporal, contingency such as cause or reason, comparison, or expansion, such as conjunction or specification Zufferey & Degan, 2013, as revised by Crible, 2018)—but for the purposes of my analyses, discourse markers will refer specifically to set phrases like for example, 那就是说 *nàjiùshìshuō* "that is to say," 怎么说 *zěnmeshuō* "how to say," 希望是说 *xīwàngshìshuō* "speaking hopefully," etc). Ultimately, as the two are categorized together, distinguishing sentential adverbials and discourse markers is not particularly relevant; the important feature they

both have is the sit high in the syntax tree above the clause. Furthermore, previous research supports the frequency with which a pause co-occurs before a discourse marker (Crible, 2019).

Category 2 was identified as including pauses that occurred after discourse markers or sentential adverbs and before the rest of the clause. One could argue they sit in the Spec of the TP[5] and one could argue they sit in an adverbial position. While I am inclined to believe the latter, I do not think determining whether it sits in the Spec or the AP of the TP changes the analysis. Thus, a pause that occurs after a discourse marker or sentential adverbial is coded as introducing a new clause. For this reason, a pause that occurred between the discourse marker 所以 *suǒyǐ* "so" or "therefore" of a new clause and the sentential adverb 现在 *xiànzài* "now" of the same clause was coded as a pause between clauses because it occurred after a discourse marker meant to introduce a new clause.

While rare, the data did show a total of three coordinated phrases (Category 3), two of which occurred with a coordinating conjunction 和 *hé* "and." One of the coordinated phrases with a conjunction was uttered by a native speaker who coordinated two NPs for the Subject, and the other conjunction was uttered by a non-native speaker who coordinated two VPs with the same conjunction. While this is technically ungrammatical in Chinese as 和 *hé* cannot coordinate VPs, I believe this is an effect of the speaker's L1 as English does not specify what sorts of phrases can be coordinated by

[5] Though Chinese does not have a Tense system as mentioned above, I will continue to refer to the TP when speaking of the clausal level of Chinese in order to keep with tradition. When clause distinction is not relevant, I will use TP and VP interchangeably. One can also use the IP/InflP for the Inflection Phrase or AspP for the Aspect Phrase.

the word “and.” The other coordinated phrase did not occur with a conjunction, but two NPs that both operated as the direct object of the same verb contained a pause between them, which I analyzed as coordination as they were both complements to the same verb.

Troublingly, this analysis of coordination being permissible even without a coordinating conjunction led to some problems as there were 4 instances of S VP_1.VP_2. I could have argued this is as two verbs of the same grammatical subject that were coordinated with an unrealized conjunction in the style of [$_{TP}$ Subj [$_{CoorP}$ VP_1 [$_{Coor°}$ Conj] [$_{VP}$ VP_2]]]. In fact, if this were English, I would’ve done just so. But as mentioned in Section 2.3, Chinese is a pro-drop language. Thus, I analyzed S VP_1.VP_2 not as two VPs sharing the same subject, but as two separate clauses with the second clause containing an omitted subject, such as [$_{VP}$ Subj [$_{VP}$ VP_1]] (pause) [$_{VP}$ Ø [$_{VP°}$ VP_2]]. While others may argue against this, I thus categorized these four S VP_1.VP_2 scenarios as belonging to Category 1’s between clause grouping rather than to the coordination of Category 3.

As for Category 4, I labeled it as Spec.VP and not Spec.TP for a couple of reasons. One, Chinese does not have a morphological Tense system, and other than the future Tense (which is arguably an irrealis Mood), Chinese does not have purely grammatical words or morphemes whose function is to indicate when the verb occurs temporally in relation to the present (past can be indicated through lexical words like “yesterday” or “last week” and through context) (Lin, 2006). Furthermore, the distribution of adverbs in Chinese suggests that auxiliary verbs are not realized at the Tense/Inflectional level but instead sit in the complement of the InflP/TP, hence why VP-level adverbs occur between the subject and the auxiliary (Paul, 2015) (see 1). Lastly, according to the VPISH (Verb Phrase-Internal Subject Hypothesis), the subject originates

in the Specifier of the VP and is left-dislocated out to the Specifier of the TP (van Gelderen, 2017, 60).

1) 这么 晚。 它 还 能 来 吗？
Zhème wǎn. [$_{TP}$Tā [$_{T'}$ [$_{T°}$ Ø] hái [$_{AuxP}$ néng lái]]] ma?
So late still can come QuesPart
It's already late. Can he still come?
(Reproduced from Paul, 2014, citing Lü Shuxiang, 2000. QuesPart by author of this paper)

2) 小蓝 明天 才 会 到 北京
Xiǎolán míngtian cái huì dào Běijīng
Xiaolan tomorrow only.then will arrive Beijing
Xiaolan will arrive at Beijing only tomorrow.
(Reproduced from Paul, 2014, citing Ernst, 1994)

Category 5 referred to pauses that occurred between a Head and its Complement. These Heads and Complement pairs included: verbs and their Complements (including direct objects, indirect objects, and prepositional complements), light verbs and their VP Complements (see (3)), prepositions and their Complements (including NPs/DPs as well as any clausal complements like the one that occurred in (3)), determiners (including quantifiers, numerals, and demonstratives with their associated Classifiers, as well as possessives with their associated 的 DE particle), within compound nouns or nouns that operated as a single unit (e.g. math.class), and subordinating conjunctions and their clausal counterparts.

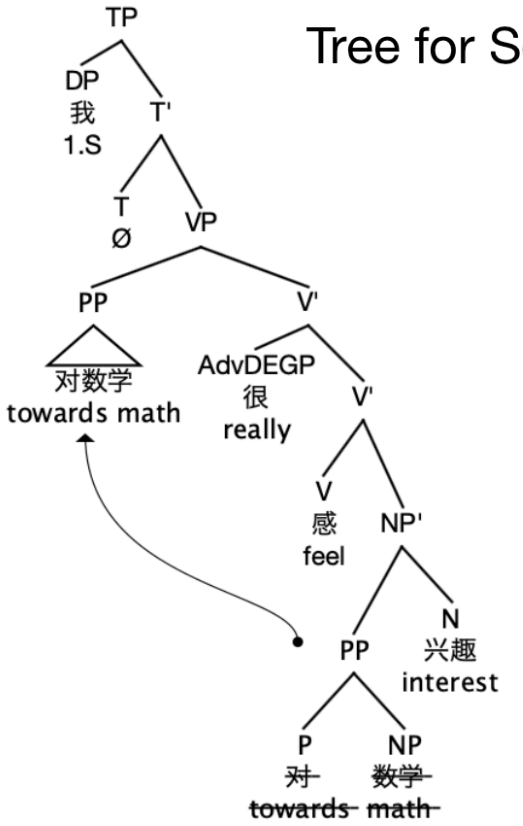
3) 关于 。 如何 去 教 读 书[6]
[$_{P°}$ Guānyú [$_{TP}$[$_{T'}$ [$_{T°}$ Ø]] 。[$_{vP}$ rúhé [$_{v°}$ qù [$_{VP}$ jiào [$_{TP}$ [$_{VP}$ dú shū]]]]]]

[6] All examples taken from this study that contain identifying information (i.e. information about the speaker's major, personalized life situations, or unique aspirations or experiences) have been omitted and replaced with grammatically identical but lexically different words. The lexical verb and the direct object in (3) was specific to that speaker's major, and so I swapped the lexical verb with a verb that also can take a reduced clause as its complement (the verb "teach" in this scenario), then exchanged the reduced clause of a verb and an object with a different lexical verb and object ("read" and "book"). Thus, the sentence maintains its original grammatical construction, but any identifying lexical information has been changed to protect the identity of the participants.

Concerning 。 how go teach read book
Concerning 。 *how to teach reading*
(Taken from this a native speaker's speech sample)

There are a couple things to note here on the topic of Category 4. One, I am making the claim that the subordinating conjunction 因为 *yīnwèi* "because" serves as a Head that takes a clause as its Complement but that coordinating conjunctions like 但是 *dànshì* "but" do not. Coordinating conjunctions can in fact serve as the Head of a CoorP, but I would argue that the relationship between a coordinating conjunction and its clausal Complement is not as close as a subordinating conjunction and its clausal Complement, hence why I've sorted pauses between coordinating conjunctions and clauses into Category 2 while pauses between subordinating conjunctions and clauses I have put into Category 4. Additionally, there are some debatable scenarios that I have chosen to label as pauses between Heads and Complements (see 4 and 5).

4) 我 对 数学 。 很 感 兴趣
Wǒ duì shùxué 。 hěn gǎn xìngqù
1S towards math 。 really feel interest
I really feel interested 。 *in math.*
(Taken from a heritage speaker's speech sample)

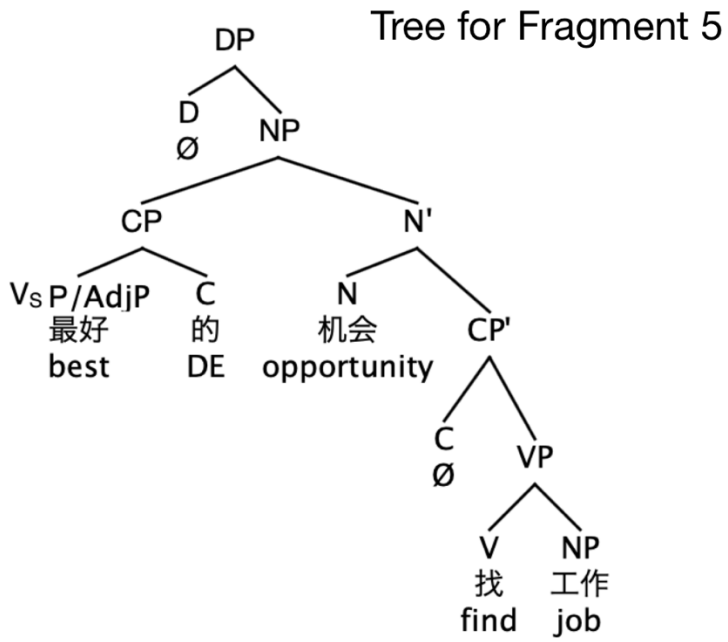


5) 最好	的	机会	。	找	工作
Zuìhǎo	DE	jīhuì	。	zhǎo	gōngzuò
Best	DE	opportunity	。	look for (find)	job

The best opportunity 。 *to look for a job*

Intended meaning: the best opportunity 。 *to find (找到) a job*

(Taken from a non-native speaker's speech sample)



In (4), I maintained that the preposition “towards math” serves as a complement. This is a contentious claim, and it may truthfully be an effect of my own L1 of English. Were I to say “I have an interest in math,” I would claim that “in math” is a complement to the noun “interest;” or if I were to say “I am interested in math,” I would argue that “in math” serves as a complement to “interested.” These intuitions are possibly fueled by the ability for “interest” to serve as a verb in English. But 兴趣 *xìngqù* can only serve as a noun in Chinese, there’s no verbal aspect to it, hence why it must always occur with verbs like “have” or “feel.” So for me to claim that “towards math” PP is a sister to the N that is then shifted out of the NP to sit in an adjunct position may be overstepping my bounds, especially as I do not truly know that the PP 对数学 *duì shùxué* base-generates as a complement to the noun. Furthermore, even if it *did* base-generate there, there would have to be the DE particle that would then disappear somewhere along its journey

northward. All that being said, to simply say “I have an interest” or “I feel interested” has a missing element as it is presumed there is *something* you are interested in.

As for (5), again, I have made some assumptions and claims, namely that I claim that the relative clause “to find a job” serves as the Complement of the noun “opportunity.” An alternative analysis of this sentence would place the relative clause as an adjunct to the NP.

Category 6 (between a Modifier and its noun Head), Category 7 (between a degree adverb and its corresponding Adjective or Verb Stative 8), and Category 8 (between a Modifier and its verbal Head) all are rather similar in that they all are essentially adjuncts to their Heads. Rather than grouping them all together as Adjuncts.Heads, I split them in this manner specifically because I wanted to compare whether they varied in their frequency across speakers. That is to say, of the three pause locations (Mod.NP, $Adv_{DEG}.V_{S}P$/AdjP, and Mod.VP), was one more or less likely to have a pause than another?

As for Categories 9 and 10, this was something of my catchall bag. Category 9 included pauses that occurred in environments where the surrounding speech was unintelligible, making syntactic analysis impossible. This Category also included “imparsable” utterances, meaning that what was said was syntactically ambiguous as in (6) where *zài* may be either a repetition of the last syllable of 现在 *xiànzài* “now,” or it may be the progressive aspect particle. Or it may be like (7), where the intended meaning is unclear. (7) could be read as a word error, a reformulation, or something else.

6) Wǒ xiànzài 。zài xué yǔyánxué[7]
1S now 。? study Linguistics
Interpretation 1: I now 。now study Linguistics.
Interpretation 2: I now 。am studying Linguistics.
(Taken from a heritage speaker's speech sample)

7) a Wǒ zài 。shàng 。Yǔyánxué xūyào shàng yī-mén yīngwén kè
1S in 。on 。Linguistics must attend one-CL English class
Interpretation 1, word error: I must attend one English class in 。at 。Linguistics
b Wǒ zài 。 shàng 。Yǔyánxué xūyào shàng yī-mén yīngwén kè
1S now 。attend 。Linguistics must attend one-CL English class
Interpretation 2, reformulation: I now 。attend 。Linguistics must attend one English class
(Taken from a heritage speaker's speech sample)

As for Category 10, this included three subgroups: a) pragmatic phrases that introduced NPs to either expand upon an earlier NP as in the first two examples, or to hold the floor as in the last example (或者说.NP *huòzhě shuō.NP* "or to say.NP", 比如说.NP *bǐrú shuō.NP* "for example.NP", and 怎么说.NP *zěnme shuō.NP* "how to say.NP"); b) pragmatic markers that provided more information (NP_1.就是 NP_2 NP_1 *jiùshì.NP* "NP_1.which is NP" and NP_1 特别是.NP_2 NP_1. *tèbié shì* NP_2 "NP_1.especially NP_2"); and c) the uncategorizable Okay. 可以了 *Okay. Kěyǐle* "Okay.All done." Because the two constituents in (c) are so small and because I don't know whether to parse 可以了 *kěyǐle* "all done" as a very, very small clause or as a discourse marker (the former would put it in Category 1 while the later reading would put it in Category 3), I've put it with the other misfits.

[7] Characters are not included for these imparsable utterances in order to provide solely a phonetic transcription so as to avoid bias in interpreting the utterance.

CHAPTER 6

EXPERIMENT 2: RESULTS & DISCUSSION

6.1 Experiment 2 Results

Having developed a coding system for pause location, I then converted my data into percentages. For each speaker, I recorded how many pauses occurred in each category, then divided that number by the total number of pauses to get what percentage of that speaker X's pauses occurred in location Y. I input these numbers into SPSS version 27 with speaker status (native, heritage, or non-native) as a categorical variable, and pause location (Categories 1–10) as a categorical variable. Speaker status served as an attribute independent variable, and pause location was another independent variable. The dependent variable was therefore the percentage, which ran on a scale of 0-100%.

With two categorical independent variables and a scalar dependent variable, I opted to conduct a 2 Way ANOVA and found a statistically significant difference in the percentage of pauses by pause location ($f(9,70) = 36.688$, $p < 0.001$) and speaker and pause location both taken into effect ($f(18,70) = 2.168$, $p = 0.011$), but not by speaker ($f(2,70) = 0.000$, $p = 1.000$) (Fig 8), though the latter is expected due to every speaker totaling 100% for all pause locations taken together.

Figure 8

Tests of Between-Subjects Effects

Dependent Variable: Percentage of Pauses

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	14663.978[a]	29	505.654	14.188	<.001	.855	411.439	1.000
Intercept	8709.594	1	8709.594	244.372	<.001	.777	244.372	1.000
Speaker	5.233E-7	2	2.617E-7	.000	1.000	.000	.000	.050
PauseLoc	11768.307	9	1307.590	36.688	<.001	.825	330.192	1.000
Speaker * PauseLoc	1390.885	18	77.271	2.168	.011	.358	39.025	.967
Error	2494.853	70	35.641					
Total	27158.791	100						
Corrected Total	17158.831	99						

a. R Squared = .855 (Adjusted R Squared = .794)

b. Computed using alpha = .05

A Levene's Test was conducted to assess for normal distribution. The Levene's test for equality of variances was maintained in this analysis with percentage of pauses having a normal distribution with unequal variances in percentage of pauses based on mean ($p < 0.001$) and based on mean after accounting for outliers ($p < 0.001$) (Fig 9). This did not hold true for the median or adjusted median ($p = 0.152$ and $p = 0.280$ respectively).

Figure 9

Levene's Test of Equality of Error Variances[a,b]

		Levene Statistic	df1	df2	Sig.
Percentage of Pauses	Based on Mean	3.563	29	70	<.001
	Based on Median	1.354	29	70	.152
	Based on Median and with adjusted df	1.354	29	13.928	.280
	Based on trimmed mean	3.367	29	70	<.001

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: Percentage of Pauses

b. Design: Intercept + Speaker + PauseLoc + Speaker * PauseLoc

A multiple regression analysis was conducted to see whether Speaker status or Pause Location significantly predicted the Percentage of Pauses that occurred. The results showed that the model explained 23.1% of variance ($R^{sq} = 0.231$). Though Pause Location contributed significantly to the model ($B = –2.194$, $p < 0.001$), Speaker Status did not ($B = 4.474E\text{-}5$, $p = 1.000$). Again, this latter is expected as every speaker would have 100% for total pause percentage due to the nature of the research design. A final predictive model was developed as: Percentage of Pauses = 22.067 + (4.474E-5*Speaker Status) + (–2.194*Pause Location) (Fig 10).

Figure 10

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Pause Location, Speaker (NS, HS, NNS)[b]	.	Enter

a. Dependent Variable: Percentage of Pauses

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.481[a]	.231	.216	11.65982

a. Predictors: (Constant), Pause Location, Speaker (NS, HS, NNS)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3971.544	2	1985.772	14.606	<.001[b]
	Residual	13187.287	97	135.951		
	Total	17158.831	99			

a. Dependent Variable: Percentage of Pauses

b. Predictors: (Constant), Pause Location, Speaker (NS, HS, NNS)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	22.067	3.873		5.697	<.001	14.380	29.755
	Speaker (NS, HS, NNS)	4.474E-5	1.337	.000	.000	1.000	-2.654	2.655
	Pause Location	-2.194	.406	-.481	-5.405	<.001	-3.000	-1.388

a. Dependent Variable: Percentage of Pauses

A Pearson's correlation was conducted to assess the correlative effect of each variable. Speaker status did not correlate with pause location ($r = 0.00$, $n = 100$, $p = 1.00$) nor with percentage of pauses ($r = 0.00$, $n = 100$, $p = 1.00$), as would be expected since every speaker was represented by a 100% of total pauses. Pause location, however, was significantly correlated with percentage of pauses ($r = –0.481$, $n = 100$, $p < 0.001$).

Figure 11

Descriptive Statistics

	Mean	Std. Deviation	N
Speaker (NS, HS, NNS)	2.2000	.87617	100
Pause Location	5.5000	2.88675	100
Percentage of Pauses	10.0000	13.16516	100

Correlations

		Speaker (NS, HS, NNS)	Pause Location	Percentage of Pauses
Speaker (NS, HS, NNS)	Pearson Correlation	1	.000	.000
	Sig. (2-tailed)		1.000	1.000
	Sum of Squares and Cross-products	76.000	.000	.003
	Covariance	.768	.000	.000
	N	100	100	100
Pause Location	Pearson Correlation	.000	1	-.481**
	Sig. (2-tailed)	1.000		<.001
	Sum of Squares and Cross-products	.000	825.000	-1810.117
	Covariance	.000	8.333	-18.284
	N	100	100	100
Percentage of Pauses	Pearson Correlation	.000	-.481**	1
	Sig. (2-tailed)	1.000	<.001	
	Sum of Squares and Cross-products	.003	-1810.117	17158.831
	Covariance	.000	-18.284	173.322
	N	100	100	100

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 12 shows the percentage of pauses at each location by speaker group, including the average of all the speakers together. The data is thus presented in descending order from the location of the highest percentage of total speaker's pauses to the location of the lowest percentage of total speaker's pauses.

Figure 12

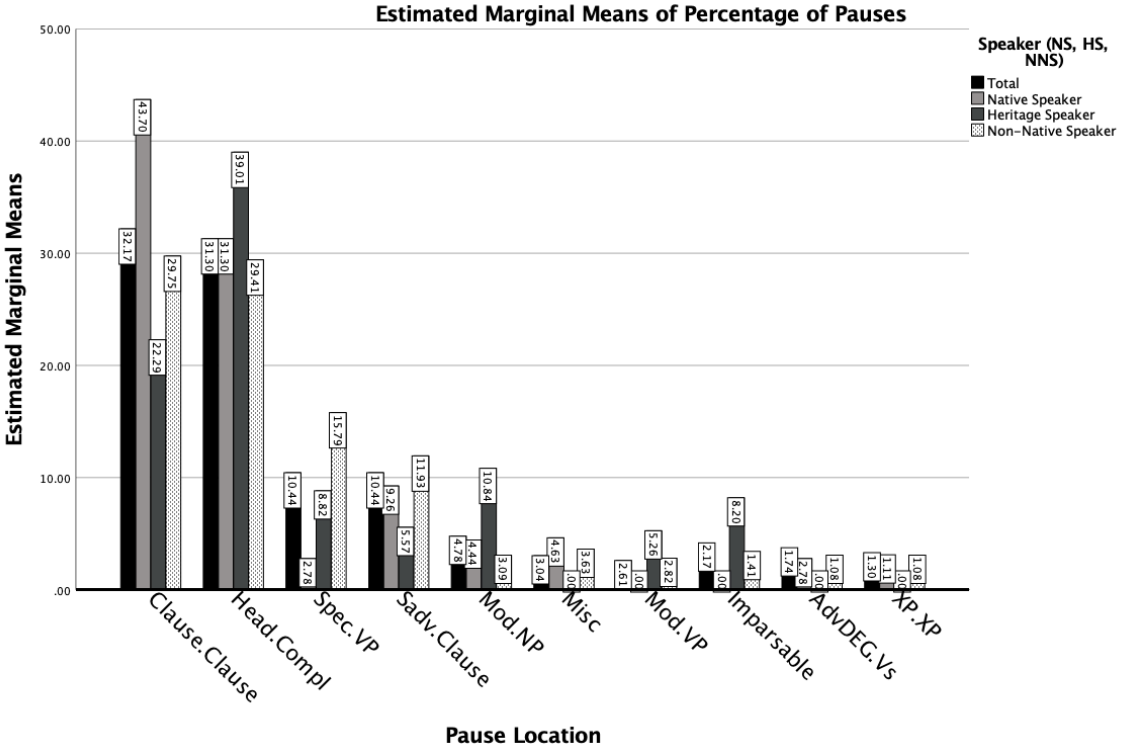
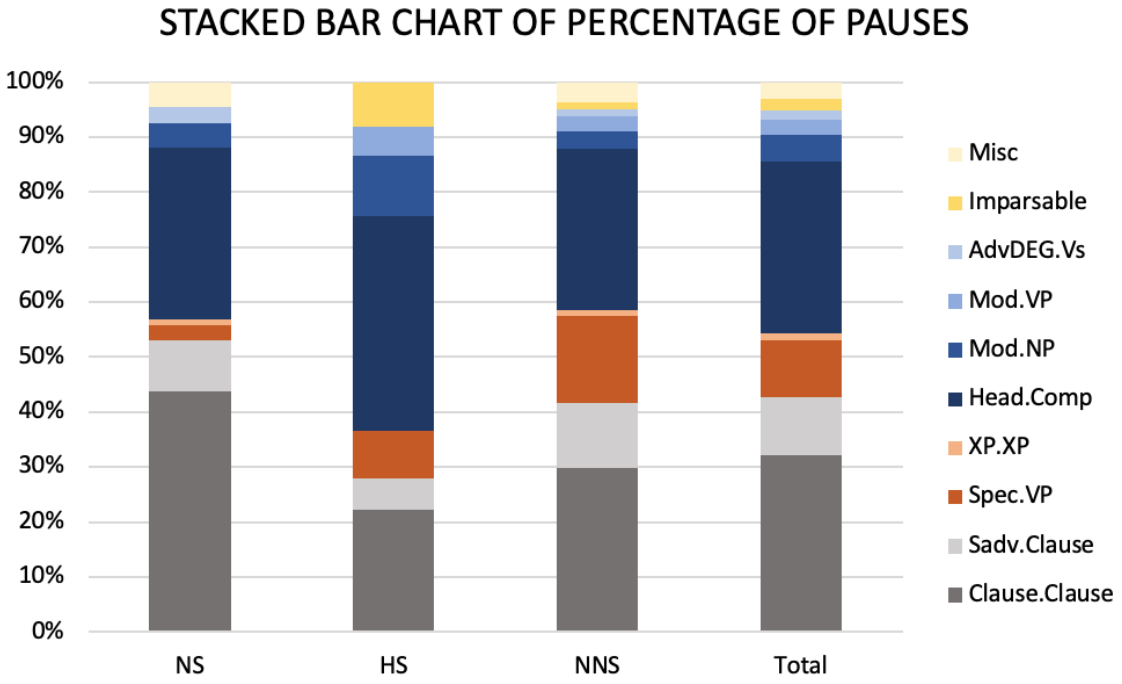


Figure 13



6.2 Experiment 2 Discussion

"Where are speakers (native, heritage, and non-native) likely to pause within a sentence" served as the research question for Experiment 2, and the null hypothesis was stated as "there is no predictable location of a pause within a sentence, and therefore, no analytic measurement correlating pause location to speaker status." Based on the previously reported results, I reject the null hypothesis and found a statistically significant correlation between the pause location and the percentage of pauses to occur at that location, as well as a statistically significant correlation between pause location and speaker status taken together and the percentage of pauses to occur.

Due to the nature of the study design, it is not unexpected to find p-values of 1.00 for speaker status in correlation with percentage of pause. Every speaker is represented by 100% of percentage of pauses at the percentage of pauses across all 10 categories totaled 100% for each speaker participant. Speaker status only becomes a truly relevant variable when taken into consideration with pause location as not all speakers uttered the same percentage of pauses at certain pause locations. The correlations of speaker status alone on pause percentages will thus not be discussed in this section as lends no insights to the research findings.

Pause location was found to correlate significantly with percentage of pauses ($p > 0.001$) , and pause location taken in conjunction with speaker status was also found to have significant correlations with percentage of pauses ($p = 0.011$. This would suggest that pause location's correlation with percentage of pauses is universal to all speakers and not specific to L1 or L2 speakers. By looking at the graph (Fig 12), we can easily see that both the Category 1 (between clauses) and Category 5 (between Head and Complement)

have the highest percentage of pauses across all speaker groups (native, heritage, non-native, and total). That a pause would occur between clauses is expected as it fits with prosody studies of pause that intonational phrase groups are offset by a pause (amongst other phonological features) that typically corresponds to a major constituent or discourse boundary (Peng et al., 2005, as cited in Yang, 2016; Fon, Johnson, & Chen, 2010; Xie, Xu, & Wang, 2012; Yang, Shen, Li, & Yang, 2014). It also falls in line with studies in L2 fluency where clausal pauses are the preferred pause location, particularly at higher levels of proficiency (Chen, 2015; Lambert, Kormos, & Minn, 2016; Saito et al., 2017; Kahng, 2018; Shea & Leonard, 2019; Kahng, 2020; Suzuki & Kormos, 2020).

That such a high percentage of pauses occurs between the Head and its Complement is unexpected. I had expected that constituents that shared a close relationship, such as that close relationship between a Head and its Complement, would be less likely to have a pause interrupt the two. Some of this high percentage can be accounted for through my choice in coding and categorizing (see Section 3.2.2.3), though it would still be one of the top locations for pauses. According to Selkirk's (2011) Syntax-Phonology Interface theory, a clause would include both embedded and matrix clauses in addition to clausal complements of functional C Heads, meaning my choice to include pauses between subordinating conjunctions and their clausal complements amongst pauses between Heads and Complements might have skewed this category to be higher as Selkirk (2011) appears to argue this location as a clausal boundary. Were I to recategorize, I think I would still avoid putting pauses between subordinating conjunctions and clausal Complements into the between clause category, and would instead count these pause percentages amongst those that occur between a sentential

adverbial or discourse marker and the following clause. Thus, while we may analyze Category 1 (between clause) pauses as being prosodic, the next highest percentage of pauses appears to be hesitation pauses occurring at the more minor constituent boundary between a Head and its Complement (Cruttenden, 1997).

From between clause pauses and between Head and Complement pauses, there is a significant drop in numbers before landing at pauses between sentential adverbials and their clauses. As mentioned just a moment ago, this number may increase if I move subordinating conjunction pauses over to this category, but it's still a rather large difference to go from 31.3% of pauses (total speakers) in Head.Complement locations to essentially any other location. While pauses between a Specifier and its VP enjoys a peak for non-native speakers by reaching 15.79%, all the other categories progressively get lower and lower as a whole. This could suggest a disfluency where speakers pause at this juncture to plan the rest of their sentence, which would stand in contrast to native and heritage speakers who paused there so little.

Interestingly, heritage speakers have something of a spike at locations that are otherwise unpopular, namely between modifiers and NPs and VPs (10.84% and 5.26% respectively) where the percentage of pauses by heritage speakers is double or more than any other speaker group. This might suggest that the two structures are mapped similarly in their mind, though whether that similarity is a similarity in degrees of closeness between the Heads and their modifiers or is a similarity between a need to word-search for the Head is unclear. Additionally, that there is a 50% reduction in pauses from between a modifier and its nominal Head to between a modifier and its verbal Head implies a greater need for a pause before a noun than a verb, though this may be owing to

constituency weight: several of the modifiers that received pauses before their nominal Heads were relative clauses, making them syntactically weighty due to the greater number of words a relative clause can hold.

Additionally, heritage speakers' percentage of pauses (both in the between- and within-phrase conditions) were closer to native speakers as compared to non-native speakers at the Sadv.Clause and Spec.VP junctures only. That being said, native speakers' pause percentage at Sadv.Clause was not wildly different from the other two speaker groups, indicating there may be a cross-linguistic acceptability in pauses after a sentential adverb. This is not entirely unexpected as sentential adverbs sit above the main clause and TP and often serve a discourse function, making it more likely for them to offset by some pragmatic, discursive prosody. As for the pauses at Spec.VP, heritage speakers and non-native speakers both showed a notable preference for pauses (8.82% and 15.79% respectively) as compared to native speakers (2.78%). The current study as it stands, however, cannot explain whether this is a disfluency feature of heritage and native speakers, or whether it is an effect of the speakers' L1 prosody as most of the non-native speakers and all of the heritage speakers spoke English as their L1. That is to say, is pausing after a specifier (which is usually a grammatical subject) and before its verbal Head an feature of disfluent prosody, or is it a characteristic prosodic feature of the English language? In order to understand this, the study would need to be re-conducted, ideally using a greater population size, with other languages to see if this pattern is observed.

The findings of this study are preliminary and still significantly limited. Experiment 2 only had a population size of 10 speakers with 230 pauses amongst them.

Though pause location was significantly correlated with percentage of pauses, it only accounted for 23.1% of variance in percentages. In order to find more robust and accurate results, this study would need to be run with a higher population sample and, depending on the researcher, a different coding system or different choices in sorting syntactic constructions into the established coding system.

While the graph in Figure 12 is interesting and prompts further questions—such as why is it so unlikely to have a pause between a degree adverb and its adjective or stative verb or between two coordinated phrases—it does not explain *why* one location might be more likely to have a pause than another. Some we can gather from previous research, such as the prosodic pause between two clauses or the hesitation pause that seems to occur between a verb and its typically highly semantic Complement, but we cannot tell from the data why certain more minor constituent boundaries have a higher percentage of pauses than others. Rerunning this particular study, even with a greater population sample, will not yield reasons for the likelihood of a pause at certain locations. To understand the mechanisms that affect this, an entirely new study would need to be developed and tested.

CHAPTER 7

CONCLUSION

This study was developed and conducted to address the gap in the literature regarding research into the interaction between pause and syntactic location at a detailed level beyond that of the previous binary system of studying pauses between clauses as opposed to pauses within clauses. The first experiment sought to use manipulated speech samples to study pause location on a more minute scale (namely between phrases and within phrases) and how it affects perceptions of fluency in Mandarin Chinese. The second experiment aimed to analyze the syntactic distribution of authentic pauses in spontaneous speech in order to observe trends in the likelihood of pauses occurring at specific syntactic locations.

Though the small population size and study limitations means that the findings in this paper are merely preliminary, the data suggests that further investigation into these phenomena is worthwhile. Though Experiment 1 did not find a statistically significant effect on fluency based on pause location at the phrasal level, findings did suggest that at least amongst heritage speakers and non-native speakers, pauses within phrases did tend to be perceived more disfluent as a whole. A reconduction of this study would benefit from a larger sample size, and from including more condition pauses in each sentence (Kahng, 2018) as opposed to merely one. Experiment 2, meanwhile, found statistically significant correlations between pause location and percentage of pause as well as between pause location and speaker status taken together and the percentage of pauses to occur at a given location for each speaker group. However, the study as it currently stands cannot offer suggestions as to *why* the pauses are distributed the way that they are.

Some of the data fits with expectations of the literature (such as the highest number of pauses occurring between clauses) while other findings are a bit more unexpected (such as the second highest percentage of pauses occurring between a Head and its Complement). A larger sample size may be able to obtain a more normalized distribution of pause percentage, but in order to understand the cognitive mechanisms behind the pause distribution observed, an entirely new study would need to be designed.

This paper was meant to serve as a preliminary pilot test into pauses that occur at and within the phrasal level. The findings of this study encourage further research into the area in order to better understand the results of these experiments, and further study is worthwhile as this area has applications in and implications for multiple fields of research in linguistics. A better understanding of where within clauses are acceptable locations to pause benefits the field of Second Language Acquisition both in terms of improving learners' perceived fluency and in improving their understanding of the underlying syntactic structure. Additionally, research into *why* people are more likely to pause at certain phrasal boundaries than others contributes to the field of psycholinguistics to help us understand how the brain both processes, plans, and produces prosodic features of speech. Lastly, having a more accurate understanding of what syntactic locations can acceptably have a pause and what syntactic locations are likely to have pause has applications in the field of computational linguistics, specifically in natural language processing. Further findings in this area can work to develop more naturalistic computer speech production and improve the accuracy of computer speech processing.

REFERENCES

Belz, M., & Reichel, U. D. (2015). *Pitch charactersitics of filled pauses* [Symposium Presentation]. Duke Industry Statistics Symposium, Edinburgh, Germany. https://epub.ub.uni-muenchen.de/25259/1/BelzReichelDiss2015.pdf

Boersma, P. & Weenik, P. (2020). Praat: Doing phonetics by computer [Computer program]. Version 6.1.09, retrieved January, 2020, from http://www.praat.org/

Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, R., de Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30,* 159–175. doi:10.1177/0265532212455394

Chen, M. (2014). The Pausing Pattern of Spontaneous Speech in Chinese as Second Language Learners (汉语作为第二语言的自发性言语停顿模式研究). *Experimental Linguistics (实验语言言学), 3*(2).

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping Lexical and Syntactic Acquisition. *Language and Speech, 51*(1&2), 61–75. https://doi.org/10.1177/00238309080510010501

Campione, E., & Véronis, J. (2002, April 11–13). *A large-scale multilingual study of silent pause duration* [Conference Presentation]. International Speech and Communication Association, Aix-en-Provence, France. https://www.isca-speech.org/archive_open/sp2002/sp02_199.pdf

Crible, L. (2019). Local vs. global scope of discourse markers: Corpus-based evidence f rom syntax and pauses. In Ó. Loureda, I. R. Fernández, L. Nadal, & A. Cruz (Eds.), *Empirical Studies of the Construction of Discourse* (pp. 43–59). John Benjamins.

Crible, L. (2018). *Discourse Markers and (Dis)fluency: Forms and functions across languages and registers.* John Benjamins Publishing.

Cucchiarini, C., Strik, H., and Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America, 111,* 2862–2873. https://doi.org/10.1121/1.1471894

Davies, A. (2003). *The Native Speaker: Myth and reality* (2nd ed.). Multilingual Matters LTD.

de Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching, 54*(2), 113–132. doi: 10.1515/iral-2016-9993

Duran-Karaoz, Z. & Tavakoli, P. (2020). Predicting L2 fluency from L1 fluency behavior: The case of L1 Turkish and L2 English Speakers. *Studies in Second Language Acquisition, 42*(4), 671–695. https://doi.org/10.1017/S0272263119000755

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2021). *Ethnologue: Languages of the World* (24th ed.). SIL International. Online version: http://www.ethnologue.com

Fon, J., Johnson, K., & Chen, S. (2010). Durational patterning at syntactic and discourse boundaries in Mandarin spontaneous speech. *Language and Speech, 54*(1), 5–32. doi:10.1177/0023830910372492

Götz, Sandra. (2013*). Fluency in Native and Nonnative English Speech*. John Benjamins Publishing.

Hakka Affairs Council in Taiwan. (2019, January 11). *Speak Hakka, our Mother Tongue.* Language Log. https://languagelog.ldc.upenn.edu/nll/?p=41357

Hawthorne, K. & Gerken, L. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency, *Cognition 133*(2), 420–428. https://doi.org/10.1016/j.cognition.2014.07.013

Interagency Language Roundtable. (n.d.). *Language skill level descriptions - Speaking.* Interagency Language Roundtable. https://www.govtilr.org/Skills/ILRscale2.htm

Itô, J., & Mester, A. (2007, January 18-21). *Categories and projections in prosodic structures* [Conference Presentation]. Old World Conference in Phonology, Rhodes, Greece. https://www.researchgate.net/profile/Junko-Ito-3/publication/228348849_Categories_and_projections_in_prosodic_structure/

Kahng, J. (2014). *Exploring the production and perception of second language fluency: Utterance, cognitive, and perceived fluency*. [Doctoral dissertation, Michigan State University]. ProQuest Dissertations Publishing.

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Linguistics, 39*(3), 569–591. doi:10.1017/S0142716417000534

Kahng, J. (2020). Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics 41*(2), 457–480. doi:10.1017/S0142716420000065

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgements of language learner proficiency in oral English. *Modern Language Journal, 94*(4), 554–566. doi:10.1111/j.1540-4781.2010.01091.x

Kormos, J., & Deńes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*(2)*,* 145–164. https://doi.org/10.1016/j.system.2004.01.001

Kurpaska, M. (2010). *Chinese Language(s): A Look through the Prism of the Great Dictionary of Modern Chinese Dialects.* De Gruyter.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition, 39*(1), 167–196. doi:10.1017/S0272263116000085

Lee, K.-S. C., Liang, H.-H., Jiao, L.-W., & Wheatley, J. K. (2014). *The Routledge Advanced Chinese Multimedia Course: Crossing Cultural Boundaries* (2nd ed.)*.* Routledge.

Lennon, P. (1990). Investigating Fluency in EFL: A Quantitative Approach. *Language Learning, 40*(3), 387–417.

Levin, H., Silverman, I., & Ford, B. L. (1967, August 1). Hesitations in Children's Speech During Explanation and Description. *Journal of Verbal Learning and Verbal Behavior, 6*(4), 560–564. https://doi.org/10.1016/S0022-5371(67)80017-3

Li, C. N., & Thompson, S. A. (1989*). Mandarin Chinese: A Functional Reference Grammar.* Berkeley.

Li, D.-D., & Liu, I. (2010). Reading into a New China: Integrated Skills for Advanced Chinese. (3rd ed., Vol. 1). Cheng & Tsui.

Li, F.-K. (1973). Languages and dialects of China. *Journal of Chinese Linguistics, 1*(1), 1–13. https://www.jstor.org/stable/23749774

Lin, J.-W. (2006). Time in a Language Without Tense: The Case of Chinese. *Journal of Semantics, 23*(1), 1–53. https://doi.org/10.1093/jos/ffh033

Liu 刘, Y. & Wu 吴, X.-Y. 汉语二语学习者口语产出的流利度研究 [A Study of Fluency in Oral Output of Chinese Second Language Learners]. *华文教学与研究 [Chinese Teaching and Research], 4*, 32–42.

Mo, C. (2014). The pausing pattern of spontaneous speech in Chinese as second language

learners (汉语作为第二语言的自发性言语停顿模式研究). *Experimental Linguistics, 3*(2). https://www.researchgate.net/publication/281337032

Morgan, J., & Demuth, K. (1996). Signal to syntax: An overview. In J. L. Morgan, & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 1-22). Lawrence Erlbaum Associates.

Norman, J. (1988). *Chinese.* Cambridge University Press.

Paul, W. (2015). *A New Perspective on Chinese Syntax*. De Gruyter.

Peng, S.-H., Chan, M. K. M., Tseng, C.-Y., Huang, T., Lee, O. J., Beckman, M. E. (2005). Towards a Pan-Mandarin system for prosodic transcription. In. S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 1–63). Oxford. doi:10.1093/acprof:oso/9780199249633.001.0001

Phonic Inc. (2021). *PhonicAi* [Online survey tool]. Phonic Inc. https://www.phonic.ai

Pinget, A.-F., Bosker, H. R., Quené, H., & Nivja, H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing, 31*(3), 349–365. https://doi.org/10.1177/0265532214526177

Prefontaine, Y. (2013). Perceptions of French Fluency in Second Language Speech Production. *The Canadian Modern Language Review / La Revue Canadienne des Langues Vivantes, 69*(3), 324–348. doi:10.3138/cmlr.1748

Préfontaine, Y., Kormos, J., & Johnson, D. E. (2016). How do utterance measurements predict raters' perceptions of fluency in French as a second language? *Language Testing, 33*(1), 53–73. https://doi.org/10.1177/0265532215579530

Prince, A. & Smolensky, P. (2004). *Optimality Theory: Constraint interaction in Generative Grammar.* Blackwell Publishing. (Original Work Published 1993).

Rehbein, J. (1987). On fluency in second language speech. In H.W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production*, 97-105. Ablex Publishing Corporation.

Révész, A., Ekiert, M., & Torgersen, E. N. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics,* 37(6), 828–848. https://doi.org/10.1093/applin/amu069

Riazantseva, A. (2001). Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition, 23*(4), 497–526. https://doi.org/10.1017/S027226310100403X

Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S (2018). Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics, 39*(3), 593–617. doi:10.1017/S0142716417000571

Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency.* Taylor & Francis Group. ProQuest Ebook Central. https://ebookcentral-proquest-com.ezproxy1.lib.asu.edu

Selkirk, E. (2011). The Syntax-Phonology. In J. A. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (pp. 435–484). John & Wiley Sons.

Selkirk, E., & Shen, T. (1990). Prosodic Domains in Shanghai Chinese. In S. Inkelas, & D. Zec (Eds.), *The Phonology-Syntax Connection* (pp. 313–338). University of Chicago Press.

Shea, C., & Leonard, K. (2019). Evaluating Measures of Pausing for Second Language Fluency Research. *The Canadian Modern Language Review, 75*(3), 216–235. doi:10.3138/clmr.2018-0258

Shen, X.-N. S. (1990). *The Prosody of Mandarin Chinese*. In M. Langdon, J. Nichols, P. Schachter, W. Shipley, & S. Thompson (Eds.), *University of California Publications in Linguistics* (Vol. 118). University of California Press.

Shi, D.-X. (2000). Topic and topic-comment constructions in Mandarin Chinese. *Language 76*(2), 383–408. doi:10.2307/417661

Simpson, A. (2014). Prosody and syntax. In Huang, C.-T. J. Huang, Y.-H. A. Li, & A. Simpson (Eds.), *The Handbook of Chinese Linguistics* (pp. 476–491). John Wiley & Sons.

Skehan, P. (2003). Task-Based Instruction. *Language Teaching, 36*(1), 1–14. doi: 10.1017/S026144480200188X

Skehan, P. (2009). Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics, 30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Southee, R. (2020). *Language in Trauma: A Pilot Study of Pause Frequency as a Predictor of Cognitive Change Due to Post Traumatic Stress Disorder.* [Master's thesis, Arizona State University]. ASU Library.

Statistics Solutions. (n.d.). *Can an ordinal Likert scale be a continuous variable?* Statistics Solutions. https://www.statisticssolutions.com/can-an-ordinal-likert-

scale-be-a-continuous-variable/

Stenström, A.-B. (2011). Pauses and hesitations. In G. Andersen, & K. Aijmer (Eds.), *Pragmatics of Society* (pp. 537–567). Walter de Gruyter.

Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition, 42*(1), 143–167. doi:10.1017/S0272263119000421

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics, 30*(4), 485–496. https://doi.org/10.1016/S0378-2166(98)00014-9

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal, 65*(1), 71–79. https://doi.org/10.1093/elt/ccq020

Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 133–150. https://doi.org/10.1515/iral-2016-9994

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and Task Performance in A Second Language* (pp. 239-273). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.11.15tav

van Gelderen, E. (2017). *Syntax: An Introduction to Minimalism*. John Benjamins Publishing Company.

van Heuven, V. & Tang, C.-J. (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua 119*(5), 709–732. https://doi.org/10.1016/j.lingua/2008.10.001

Witton-Davies, G. (2014, November 16). *Oral fluency development over four years at a Taiwanese university* [Panel Presentation]. The Twenty-third International Symposium on English Teaching, Taipei. https://d1wqtxts1xzle7.cloudfront.net/37058043/witton-Davies_14_ETA.pdf?

Wurm, S. A., Li, R., Baumann, T., Lee, M. W., Zhongguo she hui ke xue yuan., Australian Academy of the Humanities., & Australian National University. (1988). *Language atlas of China: Parts I and II*. Longman Group (Far East) Limited.

Wu, G. (n.d.). *Taiwanese.* Penn Language Center. https://plc.sas.upenn.edu/taiwanese

Xia, Q.-Y. (2017). The Influence of Regional Dialects on Putonghua Speech. 校园英语 *Xiàoyuán Yīngyǔ (Campus English), 7*(5), 226.

Xie, L., Xu, C.-L., & Wang, X.-X. (2012, December 5–8). *Prosody-Based Sentence Boundary Detection in Chinese Broadcast News.* [Conference presentation]. The 8th International Symposium on Chinese Spoken Language Processing, Institute of Electrical and Electronic Engineers, Hong Kong, China. doi:10.1109/ISCSLP.2012.6423471

Yang, C.-S. (2016). *The Acquisition of L2 Mandarin Prosody: From experimental studies to pedagogical practice.* John Benjamins Publishing Company.

Yang, X.-H., Sheng, X.-R., Li, W.-J., and Yang, Y.-F. (2014). How listeners weight acoustic cues to intonational phrase boundaries, *Public Library of Science (PLoS) ONE 9*(7). https://doi.org/10.1371/journal.pone.0102166

APPENDIX A

SPEAKER RECRUITMENT MATERIALS

Speaker Recruitment Email
Hello, ____

My name is Mary "Katie" Kennedy, I am a Master's student at ASU pursuing my degree in Linguistics and conducting a study on Chinese fluency. You are receiving this email because you may be eligible to participate in this study as person who speaks Mandarin Chinese as either a first or second language. If you are learning Chinese as a second language, a minimum of two years of study is required to join the study. Participation is entirely voluntary and would involve answering two short-answer prompts and reading a few sentences to gather speech samples of spoken Chinese, which is expected to take less than 15 minutes.

The findings of this study will benefit future language learners by identifying means to improve spoken fluency. If you are interested in participating and are over 18 and speak Chinese as either a first or second language, please contact me at (602) 538-1030 or email me at mkkenne3@asu.edu.

Best Regards,
Mary "Katie" Kennedy
Department of English
Arizona State University
_____您好，

我叫孔玛丽 Mary Kennedy，我是亚利桑那州立大学语言学系的在读硕士生/硕士班的在读 生，也正在进行一项有关于听者对说汉语者的汉语语言流利性的感知分析研究。您在收到 这封邮件的原因是您作为十八多岁的以汉语为母语或者为第二语言，可能是合格的来参与 研究。要是您在大学学习汉语作为第二语言，您为了参与研究必须有最少两个在大学学汉 语的经历。参与本项研究是完全自愿的；如果您想参与研究的话，那我为了收集说汉语的 样本会两个你两个短的问题也问您请朗读几个句子。整个过程大约 15 少分钟。如果您是十 八岁多的一汉语也想参与研究，请通过邮件或者电话的方式联系我。我的电子邮箱是 mkkenne3@asu.edu，电话号码哦是 (602) 538-1030.

通过识别怎么提高口语流利性，这项研究将会有益于学第二语言者。如果你是十八多岁以 汉语为母语或者以汉语为第二语言的人，请通过邮件或者电话的方式联系我。我的电子邮 箱是 mkkenne3@asu.edu，电话号码哦是 (602) 538-1030。

祝，

孔玛丽

英语系

亚利桑那州立大学

Speaker Recruitment Text

Hello! My name is Mary “Katie” Kennedy, and I am conducting a study to research spoken Chinese fluency. If you are over 18 and speak Mandarin Chinese as a first language or second language, you are invited to participate in this study. If you are learning Chinese as a second language, a minimum of two years of study is required to join the study. Participation is entirely voluntary and would involve answering two short-answer prompts and reading a few sentences to gather speech samples of spoken Chinese, which is expected to take less than 15 minutes. If you are interested, please message me back at this number or email me at mkkenne3@asu.edu.

您好！我叫孔玛丽，也正在进行一项有关于听者对说汉语者的汉语语言流利性的感知分析 研究。如果你是十八多岁以汉语为母语或者以汉语为第二语言的人的话，我请您参与本项 研究。要是您在大学学习汉语作为第二语言，您为了参与研究必须有最少两个在大学学汉 语的经历。参与本项研究是完全自愿的；如果您想参与研究的话，那我为了收集说汉语的 样本会两个你两个短的问题也问您请朗读几个句子。整个过程大约 15 少分钟。如果您十八 岁多也想参与研究，请用这个电话号码（602-538-1030）给我发短信或者给我打电话。您 也可以通过邮件联系我（mkkenne3@asu.edu）。

Speaker Recruitment Social Media Post

Hello! I am looking to recruit people who speak Mandarin Chinese as either a first or second language to participate in a research study! If you are learning Chinese as a second language, a minimum of two years of study is required to join the study. Participation is entirely voluntary and would involve answering two short-answer prompts and reading a few sentences to gather speech samples of spoken Chinese, which is expected to take less than 15 minutes. If you are over 18 and are interested in participating, please contact me at (602) 538-1030 or mkkenne3@asu.edu.

您好！我在征募十八多岁的以汉语为母语或者为第二语言来参与研究。要是您在大学学习 汉语作为第二语言，您为了参与研究必须有两个在大学学汉语的经历。参与本项研究是完 全自愿的；如果您想参与研究的话，那我为了收集说汉语

的样本会两个你两个短的问题也 问您请朗读几个句子。整个过程大约 15 少分钟。如果您十八岁多也想参与研究，请通过邮 件或者电话的方式联系我。我的电子邮箱是 mkkenne3@asu.edu，电话号码哦是 (602) 538- 1030。

APPENDIX B

RATER RECRUITMENT MATERIALS

Rater Recruitment Email
Hello, ____

My name is Mary "Katie" Kennedy, I am a Master's student at ASU pursuing my degree in Linguistics and conducting a study on Mandarin Chinese fluency. You are receiving this email because you may be eligible to participate in this study as a rater of fluency. Participation is entirely voluntary and will consist of listening to speech samples of Chinese language learners and assessing fluency, which is expected to take less than an hour, and for your time, you would be compensated with a $5 transfer via Paypal or a $5.00 e-gift card to your choice of Amazon, Starbucks, or Target if you live in the US. If you live in Taiwan, you will be compensated with a $5.00 Paypal transfer (which equates to approximate 130–150 NTD).

The findings of this study will benefit future language learners by identifying means to improve spoken fluency. If you are interested in participating and are over 18 and grew up with Chinese as a first language, please contact me at (602) 538-1030 or email me at mkkenne3@asu.edu.

Best Regards,
Mary "Katie" Kennedy
Department of English
Arizona State University

_____您好，

我叫孔玛丽 Mary Kennedy，我是亚利桑那州立大学语言学系的在读硕士生/硕士班的在读 生，也正在进行一项有关于听者对说汉语者的汉语语言流利性的感知分析研究。您在收到 这封邮件的原因是您作为十八多歲的以汉语为母语，可能是合格的来参与研究。参与本项 研究是完全自愿的，随时退出本项活动将不会对您造成任何影响;参与研究意味着根据我 所给您汉语语音材料评定该语音材料中说话者的流利程度。整个过程大约不到一个小时， 并且为了感谢您的参与，您会收到补偿。如果您现在在美国的话，您可以选择收到 PayPal 价值 5 美金的转帐或者收到价值 5 美金的 e-礼品卡，这个 e-礼品卡是来自于亚马逊，星巴克 或者是塔吉特 Target。如果您现在在台湾的话，您会收到 PayPal 价值 5 美金的转帐(是差不 多直接 130至150NTD)。

通过识别怎么提高又语流利性，这项研究将会有益于学第二语言者。如果您是十八多歲以 汉语为母语也想参与研究，请通过邮件或者电话的方式联系我。我的电子邮箱是 mkkenne3@asu.edu，电话号码哦是 (602) 538-1030.

祝，

孔玛丽

英語系

亚利桑那州立大学

Rater Recruitment Text

Hello! My name is Mary "Katie" Kennedy, and I am conducting a study to research spoken Mandarin Chinese fluency. If you are over 18 and grew up with Chinese as a first language, you are invited to participate in this study as a rater. Participation is entirely voluntary and will consist of listening to speech samples of Chinese speakers to assess for fluency, which is expected to take less than an hour, and for your time, you would be compensated with a $5 transfer via Paypal or a $5.00 e-gift card to your choice of Amazon, Starbucks, or Target if you live in the US. If you live in Taiwan, you will be compensated with a $5.00 Paypal transfer (which equates to approximate 130–150 NTD). If you are interested, please message me back at this number or email me at mkkenne3@asu.edu.

您好！我叫孔玛丽，我正在进行一项有关于听者对说汉语者的汉语语言流利性的感知分析 研究。如果您十八歲及以上的以汉语为母语的人，我将邀请您当评级人去参与本项研究。 参与本项研究是完全自愿的，随时退出本项活动将不会对您造成任何影响;本项研究的基 本任务是根据我所给您汉语语音材料评定该语音材料中说话者的流利程度。整个过程大约 不到一个小时，并且为了感谢您的参与，您会收到补偿。如果您现在在美国的话，您可以 选择收到 PayPal 价值 5 美金的转帐或者收到价值 5 美金的 e-礼品卡，这个 e-礼品卡是来自于 亚马逊，星巴克或者是塔吉特 Target。如果您现在在台湾的话，您会收到 PayPal 价值 5 美金 的转帐(大约等值于 130至150NTD)。如果您对参与研究有兴趣，请用这个电话号码 (602- 538-1030)给我发短信或者给我打电话。您也可以通过邮件联系我 (mkkenne3@asu.edu) 。

Rater Recruitment Social Media Post

Hello! I am looking to recruit people who grew up with Mandarin Chinese as a first language to participate in a research study! Participation is entirely voluntary and would involve listening to speech samples from people learning Chinese to assess for fluency, which is expected to take less than an hour, and for your time, you would be compensated with a $5 transfer via Paypal or a $5.00 e-gift card to your choice of Amazon, Starbucks, or Target if you live in the US. If you live in Taiwan, you will be compensated with a $5.00 Paypal transfer (which equates to approximate 130–150 NTD). If you are over 18 and are interested in participating, please contact me at (602) 538-1030 or mkkenne3@asu.edu.

您好！我叫孔玛丽，也在徵募十八歲及以上的以汉语为母语的人来参与本项研究。根据所 给语音材料评定该语音材料中说话者的流利程度，所提供的语音材料均为汉语。参与本项 研究是完全自愿的，随时退出本项活动将不会对您造成任何影响;本项研究的基本任务是 根据我所给您汉语语音材料评定该语音材料中说话者的流利程度。整个过程大约不到一个 小时，并且为了感谢您的参与，您会收到补偿。如果您现在在美国的话，您可以选择收到 PayPal 价值 5 美金的转帐或者收到价值 5 美金的 e-礼品卡，这个 e-礼品卡是来自于亚马逊， 星巴克或者是塔吉特 Target。如果您现在在台湾的话，您会收到 PayPal 价值 5 美金的转帐 (大约等值于 130至150NTD)。如果您大于十八岁并且想参与研究，请通过邮件或者电 话的 方式联系我。我的电子邮箱是 mkkenne3@asu.edu，电话号码是 (602) 538-1030。

80

APPENDIX C

PARTICIPANT TASKS

Biographic Information for Speakers

•What is your age?
•What is your gender? (optional question)
•What is your first language/what are your first languages?
•If Chinese is your first language:
–What is your native dialect?
–How many dialects do you speak? (Please list what they are)
•If Chinese is not your first language:
–What is(are) your first language(s)?
–Are you a heritage Chinese speaker? (Meaning, did you grow up in a household that spoke Chinese but in a country where Chinese was not the primary language)
–If you are a heritage speaker, what dialect(s) of Chinese was spoken in your household?
–When did you start learning Chinese (in middle school, high school, or through extracurricular Chinese language courses)
–What dialect of Chinese did you study?
–How many years have you studied Chinese?
–How many years have you studied Chinese at the college level?
–Have you ever studied Chinese abroad? If so, where did you study and for how long?
–Based on your own assessment, how would you rate your speaking ability in Chinese using the ILR (Interagency Language Roundtable) Scale? (Please see the official description of the ILR scale for reference https://www.govtilr.org/Skills/ILRscale2.htm)
•0: No proficiency
•1: Elementary Proficiency
•2: Limited Working Proficiency
•3: General Professional Proficiency
•4: Advanced Professional Proficiency
•5: Functionally Native Proficiency

•请问您的年龄是多少？

•您的性别是什么？（可选择的问题）

•您的母语是哪一个（哪一些）？

•如果您的母语包括汉语:

-您的母语方言是什么?

-您会说多少方言?（请写下会说的方言）

•如果汉语不是您的母语:

-您的母语包括哪一个（哪一些）语言?

-您是不是“传统演讲者”？（也就是说，您长大在说汉语的家庭，但是所住的国家 主要语言不包括汉语）

-如果你是“传统演讲者”，您家人用哪一个（哪一些）汉语方言?

-您从么时候开始学汉语（国中(台湾的说法)，高中，课外的汉语课）?

-您学汉语多少年了?

-您在大学学汉语多少年了?

-您学哪一种汉语方言?

-您以前去过外国学习汉语吗？如果去过，请告诉我您在哪里学了多长时间?

•根据您自己的评价，如果您评定自己说汉语能力的话，怎么用机构间圆桌会议的语言比 例（ILR）规模来评定说汉语能力？（请参考 ILR 规模正式的描述）

-0: 无基本能力

-1: 基本能力

-2: 有限的工作能力

-3: 普通专业工作能力

-4: 高级专业能力

-5: 母语精通

Speaker Tasks

Instructions for open-ended prompts

Instructions: please do your best to talk about the following prompts for a minute. I will let you know when 60 seconds have passed, at which point you are welcome to continue discussing the prompts if you would like. Please take as much time as you need to think, and when you are ready, we can begin recording.

说明：您有一分钟的时间，请尽你所能讨论下述问题。60 秒过后，我会通知您；届时，您 可以选择继续讨论或者结束。您可以花费任意长的时间来思考答案，如果您准备就绪，我 们就开始录音。

Prompt 1:

Please tell me about your major. (E.g. what is your major, what is your major about, why did you pick it, what do you like/dislike about it, what do you hope to do after graduation)

请围绕您的专业叙述。（比如说：专业是什么，专业是关于什么的，为什么选择您的专业， 你喜欢您专业的哪一（些）方面，不喜欢哪一（些）方面，毕业之后打算做什么，等等）

Prompt 2:

Pease tell me about your free time activities. (E.g. what are your hobbies, what are activities you like to do, how frequently do you get to do these activities)
请讨论您业余时的活动。（比如说：爱好是什么，喜欢做什么活动，您多久做这些活动一次，等等）

Read-Aloud:
Please read the following sentences. I have provided the sentences in Chinese (simplified and traditional) along with their translation and pinyin pronunciation. Please practice the sentence as much as you would like, and when you are ready, we can begin recording.
请朗读下述的句子。请朗读下述的句子。我提供了汉语句子（繁体字和简体字）、翻译和拼音。您可以练习朗读这些句子，当您准备就绪，我们便开始录音。

Sentences:
1）中国改革开放初期，政府先在深圳，珠海，汕头和厦门建立了四个经济特区作为试点，进行经济体制改革，并发展市场经济。
中國改革開放初期，政府先在深圳，珠海，汕頭和廈門建立了四個經濟特區作為試點，進行經濟體制改革，並發展市場經濟。
Zhōngguó gǎigé kāifàng chūqí, zhèngfǔ xiān zài Shēnzhèn, Zhūhǎi, Shàntóu hé Xiàmén jiànlìle sì gè jīngjì tèqū zuòwéi shìdiǎn, jìnxíng jīngjì tǐzhì gǎigé, bìng fāzhǎn shìchǎng jīngjì.
At the beginning period of China's Reform and Opening Up Policy, the government first established special zones in Shenzhen, Zhuhai, Shantou, and Xiamen to serve as pilot projects, to carry out economic structural reform, and to develop market economies.
(Text reproduced from Lee et al., 2014, p. 80)

2) 外来的商品在街头商店随处可见，很多品牌更是在媒体的宣传下成为中国人心目中的“世界名牌。”
外來的商品在街頭商店隨處可見，很多品牌更是在媒體的宣傳下成為中國人心目中的“世界名牌。”
Wàilái de shāngpǐn zài jiētóu shāngdiàn suíchù kějiàn, hěnduō pǐnpái gèng shì zài méitǐ de xuānchuán xià chéngwéi Zhōngguórén xīnmù zhōng de "shìjiè míngpái."
Foreign products can be seen in street shops everywhere, and under the publicity of the media many brands even are called "global brands" in the eyes of the Chinese people.
(Text reproduced from Lee et al., 2014, p. 29)

3) 现在，大学毕业生不再由国家分配工作，他们可以通过人才市场自己找工作，实现用人单位和大学生双向选择。

現在，大學畢業生不再由國家分配工作，他們可以通過人才市場自己找工作，實現用人單位和大學生雙向選擇。

Xiànzài, dàxué bìyè shēng bù zài yóu guójiā fēnpèi gōngzuò, tāmen kěyǐ tōngguò réncái shìchǎng zìjǐ zhǎo gōngzuò, shíxiàn yòngrén dānwèi hé dàxuéshēng shuāngxiàng xuǎnzé.
Now college graduates are no longer assigned jobs by the state, they can search jobs through the talent market, actualizing a two-way choice between employers and college students.
(Text reproduced from Li & Liu, 2010, p. 83)

Biographic Information for Raters

•What is your age?
•What is your gender? (optional question)
•What is your first language/what are your first languages?
•If Chinese is your first language:
–What is your native dialect?
–How many dialects do you speak? (Please list what they are)
•In what country did you grow up? (optional question)
•Have you spent time abroad/have you ever moved to another country?
–If you answered yes, please list which country you moved to and for how long you lived there.
•On a scale of 1 (not at all familiar) to 9 (extremely familiar), how familiar are you with English-accented Chinese?
•Do you have any experience teaching Chinese to a non-native speaker?

•请问您的年龄是多少？

•您的性别是什么？（可选择的问题）

•您的母语包括哪一个（哪一些）语言?

•如果您的母语包括汉语:

-您的母语方言是什么?

-您会说多少方言?(请写下会说的方言是什么）•您在哪国家长大?（可选择的问题）

• 请问，您住过外国或者搬到其他国家?

-如果您的答案是对，请写下您去哪国家，也写下在那里住多长时间了。

•请以一（一点也不熟悉）至九（极其熟悉）的尺度来评估您对英文又音的汉语有多熟悉?

•您有没有对非母语者教汉语的经验?

Rater Task

Fluency Assessment:

Please listen to the following speech samples and rate how fluent they are using a 7-point scale:

1 = extremely disfluent
2 = very disfluent
3 = somewhat disfluent
4 = neither noticeably fluent nor disfluent
5 = somewhat fluent
6 = very fluent
7 = extremely fluent

For the purposes of this study, fluency is assessed on the basis of how easily and smoothly the speech was delivered, not on the overall proficiency of the speaker. When rating fluency, please judge *solely* on:

- Speech rate
- Silent and filled pauses (e.g. um, uh)
- Hesitations and corrections
- False-starts, restarts, and repetition
- Overall flow of speech
- NOT on grammar, vocabulary, or pronunciation

请听汉语讲话范例，然后评定说话者的流利性(1 为极其不流利，7 为极其流利)。

1 = 极其不流利

2 = 非常不流利

3 = 有一点不流利

4 = 既不明显不流利也不明显流利

5 = 有一点流利

6 = 非常流利

7 = 极其流利

就本项研究的目的而言，流利性评定以说者讲话的容易和顺利程度为基础，不依赖说者语 言综合运用能力。评定语言流利性，请只考虑仅有下述的:

- 语速
- 无声停顿和填充停顿(比如说，um, uh, 嗯)
- 犹豫和改正
- 错误的开始、重新的开始和重复

- 说话时语气的流动
- 不考虑语法、生词或者发音

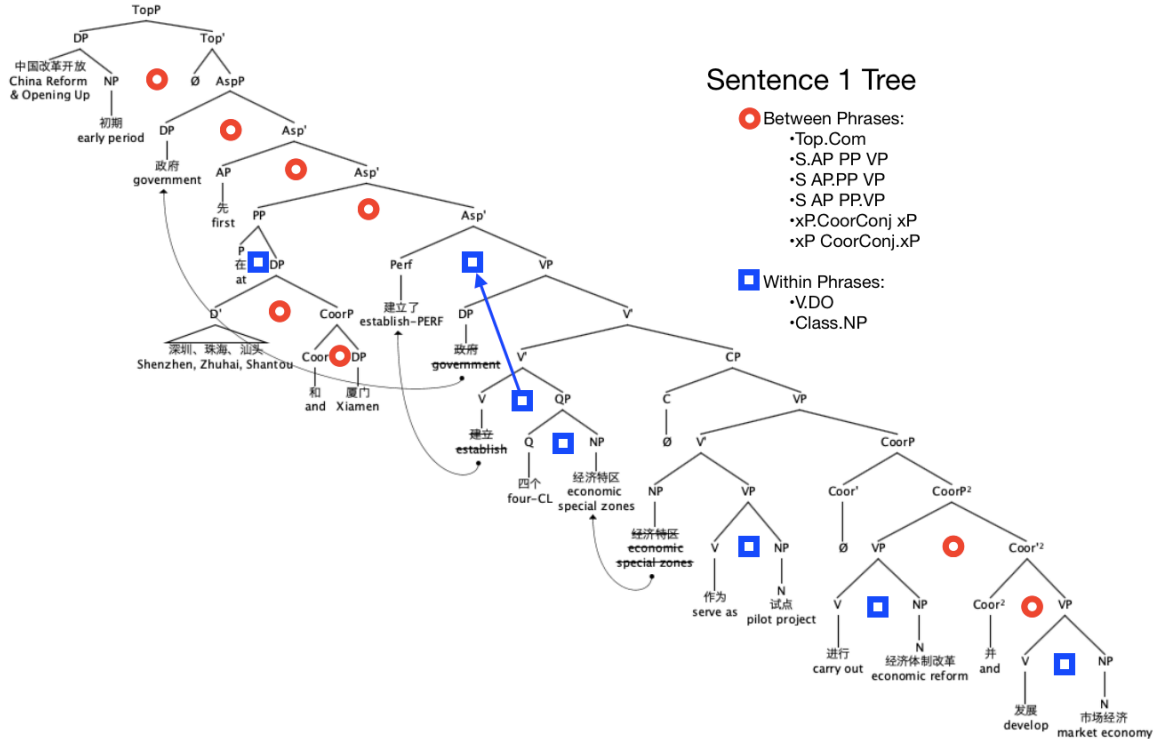
APPENDIX D

EXPERIMENT 1 SENTENCE ANALYSES

1）中國改革開放初期，政府先在深圳，珠海，汕頭和廈門建立了四個經濟特區作為試點，進行經濟體制改革，並發展市場經濟。

中国改革开放				初期	政府		先	在	
Zhōngguó gǎigé kāifàng				chūqí	zhèngfǔ		xiān	zài	
China Reform & Opening Up				early period	government		first	at	
深圳	珠海	汕头	和	厦门	建立了			四个	
Shēnzhèn	Zhūhǎi	Shàntóu	hé	Xiàmén	jiànlì-le			sì-gè	
Shenzhen	Zhuhai	Shantou	and	Xiamen	establish-PERF			four-CL	
经济特区			作为	试点		进行		经济	
jīngjì tèqū			zuòwéi	shìdiǎn		jìnxíng		jīngjì	
economy special zone			serve as	pilot program		carry out		economy	
体制	改革	并	发展	市场	经济				
tǐzhì	gǎigé	bìng	fāzhǎn	shìchǎng	jīngjì				
structure	reform	and	develop	market	economy				

At the beginning period of China's Reform and Opening Up Policy, the government first established special zones in Shenzhen, Zhuhai, Shantou, and Xiamen to serve as pilot projects, to carry out economic structural reform, and to develop market economies.
(Text reproduced from Lee et al., 2014, p. 80)



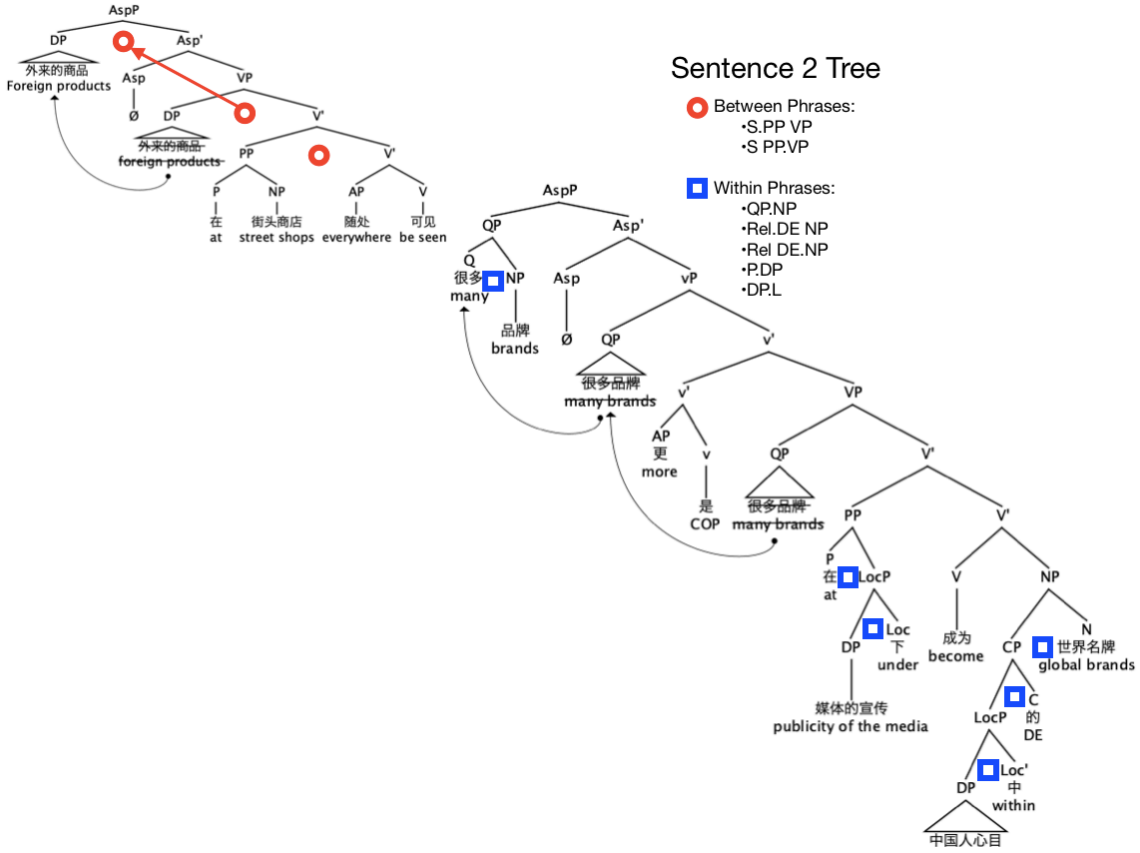
2) 外來的商品在街頭商店隨處可見，很多品牌更是在媒體的宣傳下成為中國人心目中的“世界名牌。”

外来	的	商品	在	街头	商店	随处
Wàilái	de	shāngpǐn	zài	jiētóu	shāngdiàn	suíchù
Foreign	DE	product	at	street	shop	everywhere

可见	很多	品牌	更	是	在	媒体	的	宣传
kějiàn	hěnduō	pǐnpái	gèng	shì	zài	méitǐ	de	xuānchuán
can-see	many	brand	more	COP	at	media	DE	publicity

下	成为	中国人	心目	中	的	世界	名牌
xià	chéngwéi	Zhōngguórén	xīnmù	zhōng	de	shìjiè	míngpái
under	become	Chinese people	eyes	within	DE	globe	famous brand

Foreign products can be seen in street shops everywhere, and under the publicity of the media many brands even are called "global brands" in the eyes of the Chinese people.
(Text reproduced from Lee et al., 2014, p. 29)



3) 現在，大學畢業生不再由國家分配工作，他們可以通過人才市場自己找工作，實現用人單位和大學生雙向選擇。

Xiànzài, dàxué bìyè shēng bù zài yóu guójiā fēnpèi gōngzuò, tāmen kěyǐ tōngguò réncái shìchǎng zìjǐ zhǎo gōngzuò, shíxiàn yòngrén dānwèi hé dàxuéshēng shuāngxiàng xuǎnzé.

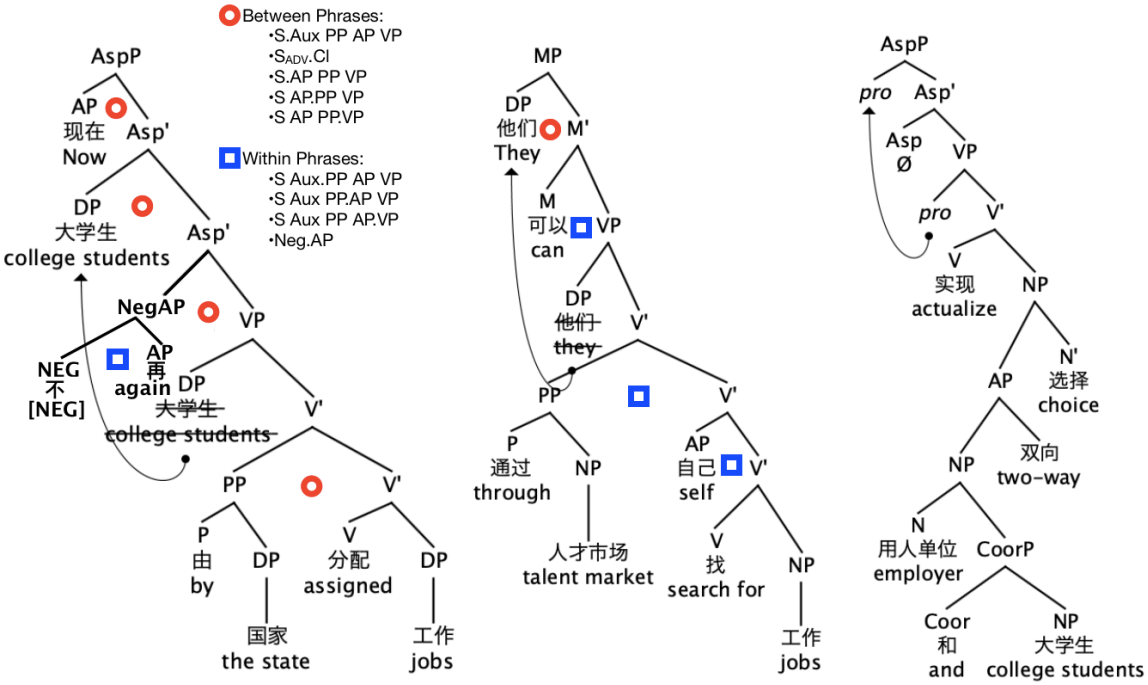
现在	大学毕业生	不再	由	国家	分配
Xiànzài	dàxué bìyè sheng	bù zài	yóu	guójiā	fēnpèi

Now	college graduate student				no longer		by	the state		assign
工作	他们	可以	通过		人才市场			自己	找	
gōngzuò	tāmen	kěyǐ	tōngguò		réncái	shìchǎng		zìjǐ	zhǎo	
job	3-PL	can	through		talent market			self	search	
工作	实现		用人单位			和	大学生			
gōngzuò	shíxiàn		yòngrén dānwèi			hé	dàxuéshēng			
job	actualize		employer			and	college students			
双向	选择									
shuāngxiàng	xuǎnzé									
two-way	choice									

Now college graduates are no longer assigned jobs by the state, they can search jobs through the talent market, actualizing a two-way choice between employers and college students.

(Text reproduced from Li & Liu, 2010, p. 83)

Sentence 3 Tree



APPENDIX E

IRB APPROVAL

APPROVAL: MODIFICATION

Elly Van Gelderen
CLAS-H: English
480/965-3535
ellyvangelderen@asu.edu

Dear Elly Van Gelderen:

On 4/21/2021 the ASU IRB reviewed the following protocol:

Type of Review:	Modification / Update
Title:	The Effect of Pause Placement on Perceptions of L2 Chinese Fluency
Investigator:	Elly Van Gelderen
IRB ID:	STUDY00012380
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• IRB_Social_Behavioral_2019_Kennedy_Pause_Study, Category: IRB Protocol;

The IRB approved the modification.

When consent is appropriate, you must use final, watermarked versions available under the "Documents" tab in ERA-IRB.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator

cc: Mary Kennedy

APPROVAL: EXPEDITED REVIEW

Elly Van Gelderen
CLAS-H: English
480/965-3535
ellyvangelderen@asu.edu

Dear Elly Van Gelderen:

On 9/8/2020 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	The Effect of Pause Placement on Perceptions of L2 Chinese Fluency
Investigator:	Elly Van Gelderen
IRB ID:	STUDY00012380
Category of review:	
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• IRB_Social_Behavioral_2019_Kennedy_Pause_Study, Category: IRB Protocol; • Rater_Consent_Form_Kennedy_Pause_Study, Category: Consent Form; • Rater_Recruitment_Email_Kennedy_Pause_Study, Category: Recruitment Materials; • Rater_Recruitment_Social_Media_Post_Kennedy_Pause_Study, Category: Recruitment Materials; • Rater_Recruitment_Text_Message_Kennedy_Pause_Study, Category: Recruitment Materials; • Speaker_Consent_Form_Kennedy_Pause_Study, Category: Consent Form; • Speaker_Recruitment_Email_Kennedy_Pause_Study, Category: Recruitment Materials; • Speaker_Recruitment_Social_Media_Post_Kennedy_Pause_Study,

	Category: Recruitment Materials; • Speaker_Recruitment_Text_Message_Kennedy_Pause_Study, Category: Recruitment Materials; • Supporting_Documents_Tasks_Kennedy_Pause_Study, Category: Participant materials (specific directions for them);

The IRB approved the protocol from 9/8/2020 to 9/7/2025 inclusive. Three weeks before 9/7/2025 you are to submit a completed Continuing Review application and required attachments to request continuing approval or closure.

If continuing review approval is not granted before the expiration date of 9/7/2025 approval of this protocol expires on that date. When consent is appropriate, you must use final, watermarked versions available under the "Documents" tab in ERA-IRB.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator

cc: Mary Kennedy
Elly Van Gelderen
Mary Kennedy