Second language learner multimodality and linguistic development in naturalistic settings

A study of L2 learners in the Chinese street market

Author’s Name: Jacob Lee Wild
The University of Nottingham

British Council ELT Master’s Dissertation Awards: Commendation
Second language learner multimodality and linguistic development in naturalistic settings
A study of L2 learners in the Chinese street market

Jacob Lee Wild

MA Applied Linguistics and English Language Teaching
Second language learner multimodality and linguistic development in naturalistic settings: A study of L2 learners in the Chinese street market

By

Jacob Lee Wild

2015

A Dissertation presented in part consideration for the degree of MA Applied Linguistics
Abstract

This study has investigated second language (L2) learners of Mandarin Chinese in interaction with vendors in the market place, attempting to understand the roles that multimodal communication plays in these interactions and whether the use of multiple modes of communication can facilitate L2 linguistic development. These interactions were video- and audio-recorded and analyzed for speech, gesture, eye gaze, and facial expression using ELAN software. Interviews with L2 participants were also utilized in order to better understand the purposes behind their actions and communicative strategies as well as to better understand the process of linguistic development if indeed it does occur. Through analysis of interactions and triangulation with interviews, it was found that gestures do play a vital role in mediating these interactions and may aid communication in cases where linguistic devices are not yet fully developed, though this seems to be dependent on a variety of factors related to the context of the interaction. Furthermore, it was found that linguistic knowledge can develop through these interactions with the use of gestures and modes of communication other than speech, and that this may be somewhat attributable to the creation of a 'zone of proximal development'. However, linguistic knowledge gained still seems somewhat limited for L2 participants, possibly related to factors such as learning style and motivation.
Acknowledgements

There are a number of individuals whose help and support proved vital to the completion of this project. I am deeply indebted to those listed below and likely many others for their encouragement.

First, I would like to thank Dr. Simon Harrison for being such an immense help on the project. Without his insights and inspiration, this research would likely not have come to fruition in the manner it has. Lending me his high-quality (and quite expensive) camera, as well as his wisdom and experience, has made working on this project intellectually motivating and quite enjoyable. Without his help I would likely not have delved so deeply into the complex and exciting world of multimodal research, and for that, I am deeply indebted.

I am also extremely lucky to be blessed with such a wonderful friend and translator, Camen Teh. Without her help, this project would never have taken off. For her patience as we wandered around the streets of Ningbo, delayed dinner, chased down moving food carts, and were postponed time and time again due to rain and poor weather, I am forever grateful. Her assistance and careful considerations for the nature of my research and the interactions that were the subject of this research proved indispensable, and I simply cannot imagine how this project would have been completed without her.

The Chinese language learners in this study also deserve praise for their participation. Certainly, speaking to proficient speakers as a beginner is a daunting experience, and some form of bravery is required to allow for such an interaction to be filmed and closely analyzed! Again, dealing with last minute decisions and often uncooperative weather also required tolerance, which was very much appreciated. The insights gained from these interactions and interviews could obviously not be reached without such cooperation on the part of the participants in this study, and I am very thankful for their time and effort.

Finally, and perhaps most importantly, I am eternally grateful for the participation of the vendors in this study. That they even permitted our cameras to film them working was clearly vital to the research. In fact, they often invited us into their work space with a smile and a laugh (whether they were laughing with us or at us is a matter of debate among those participating in the study). Having such warm-hearted people as our proficient speaking participants made gathering data pleasant and gratifying. I am extremely indebted to the vendors in these markets around Ningbo and whole-heartedly wish them the best, in business and in life.

There are likely others who have influenced me, encouraged me, or helped me get through the rough patches of the process of dissertation writing. I thank you all.

University of Nottingham Ningbo, July 2015
Jacob L. Wild
Chapter 1 - Introduction

While second language learning and acquisition research (SLA) has seen a surge of interest in the last half of a century, there still seems to be many domains of inquiry to pursue. While some recent research has focused on the role of gestures in the second language (L2) learning process, much of this research has taken place in the classroom or laboratory. Thus, the goal of this research is to examine the role gestures play in L2 learning in a naturalistic setting. While some have examined the gestures of L2 learners in such environments (e.g., Streeck 2009), there is little to no research on the role gestures play in L2 development in such environments, as is the focus of this study, making this research mostly exploratory.

While this study is based in interactionist approaches to language learning, it is also heavily influenced by other areas of research that have been shown to impact the language learning process, most notably gesture studies. In that human interaction is seen as fundamental to language learning in this study, it falls within the 'constructionist' paradigm of modern SLA, language developing in use and for communicative purposes (e.g., Veronique 2014). Central to this view is Vygotsky’s (1978) assertion that language development begins externally, and that through human interaction, this process then becomes internalized and later leads to production. A key concept developed by Vygotsky is the notion of the 'zone of proximal development' (henceforth, ZPD). Through interaction, more capable or proficient learners scaffold or guide less proficient learners, thereby allowing for learning development. Relevant to this study, some have claimed that modes of communication other than speech, notably hand gestures, may contribute to the ZPD, and therefore aid in the language development process (McCafferty 2002). However, this study does not attempt to provide an "idealized" gestural ZPD model for learners in an instructional setting, but rather attempts to examine how gestures are used by learners in a certain context, if they contribute to language learning, and if this learning can be attributed to the notion of a "ZPD".

In fact, research on human gesture, be it with the hand or any other body part, has increasingly gained acceptance within the academic literature and the field of applied linguistics as vital to a more holistic understanding of the human communication system. Following important early studies on the role of gesture in relation to speech, gesture has been shown to be intricately tied to speech in both timing (Kendon 1980) and in semantic content (McNeill 1985), with many now viewing speech and gesture as an 'integrated system' (McNeill 1998; Kendon 2004) and as a window into the mind (McNeill 1992). While this field has been termed gesture studies, there is some debate as to what actually constitutes a gesture and in which circumstances. Many in the field have settled on the term
multimodal communication to refer to this gesture-speech integrated system (e.g., Streeck 2009; Calbris 2011; Seyfeddinipur & Gullberg 2014), in that multiple 'modes' are being used in interaction. In fact, numerous studies have shown gesture to be a key component in understanding language development and the field has grown quite rapidly in recent years (e.g., McCafferty & Stam 2008; Gullberg & de Bot 2010).

In other interactionist approaches to L2 development, some within the SLA sphere have argued that current research has focused too heavily on traditional instructional and classroom settings, while suggesting that the field must expand to account for the diversity of L2 learning experiences worldwide (e.g., Tarone 2000, 2009). The term naturalistic has been used to describe these experiences, though definitions of what this actually means seem to vary. Therefore, the term naturalistic will refer to two aspects of interactions in this study:

1) A physical space or environment outside of the classroom or laboratory, and

2) The process by which communicative input and interaction may allow for linguistic development within this space or environment.

The naturalistic environment of this study is the Chinese street market. Evers (2014) has called these "more-than-human assemblages" (101), and these impermanent markets act as centers of informal economies in many Chinese cities, with vendors selling a variety of goods from mobile stands that frequently move to meet customers and avoid harassment from law enforcement authorities. While Evers has used the term 'night market' to refer to these, this study will use the term 'street market', in that this is the colloquial term used by the participants in the study. However, these terms can be seen as equivalents. In the Yinzhou District of Ningbo near the University of Nottingham Ningbo where this research took place, vendors in these markets often sell food and other goods from the late afternoon onwards.

The reason for choosing street markets as the site for this study is two-fold. First, these markets are frequented by locals and "foreigners" alike, or as they are commonly referred to in Chinese, lǎo wài, including all of the L2 participants involved with this research. Therefore, interactions between the L2 learners and proficient Chinese speaking vendors are mostly naturalistic in that they are interactions that would occur even without a camera present. Second, there are no translations available, either in writing or in speech. Therefore, L2 learners must communicate with the resources that are available to
them in the moment, linguistic or otherwise, and is therefore similar to L2 learner interactions with proficient speakers in daily life. Moreover still, informal observations of these sites by the researcher have revealed L2 learners to frequently use many modes of communication in attempts at successful interaction with proficient speakers in these markets. However, studies on language learning in naturalistic environments, especially those examining the 'study abroad' experience, have noted that the term 'language learner' is difficult to define (Kinginger 2009: 3). People use languages and travel for a variety of reasons which cannot so easily be categorized, and language learners may have varying proficiencies of a multitude of languages. This study will use the term 'L2 learner' for ease of description, and is used to refer to those whose native language is not Mandarin Chinese, the majority language of the setting of this study, while noting that all participants consider themselves to be 'multilingual'.

Similarly, it should also be noted that these are intercultural exchanges, as all L2 participants in this study originate from either Europe or the United Kingdom. However, in that this study can be seen as exploratory and L2 learner-focused, the role of culture in these interactions was not centerpiece, though it clearly shapes the interaction. Given the size of this project, it was determined that an in-depth analysis of the L2 learner would be more beneficial in answering research questions, though appropriate considerations were made regarding the ethics of such research (see Section 3.2.2). Furthermore, given the time frame allotted for data collection, this study was not longitudinal. Therefore, examining notions such as 'communicative competence' (e.g., Hymes 1984) or proficiency development would have no point of reference and could not be taken into account. Quite obviously, both of these limit the conclusions that can be drawn from this research, however it is hoped that this exploratory study can lay a foundation for future research that examines both the cultural aspects of these interactions as well as how they occur over the course of time.

Noting these limitations, this study most resembles 'micro-ethnographic' conversation analysis of L2 learners (e.g., Streeck 2009) attempting to view the interactions while they happen and determine how communication and language learning occur at the 'micro-analysis' level. Interviews with L2 participants were also utilized as a means to 'triangulate' data from the interactions to better understand L2 participant actions and linguistic development processes related to interactions with proficient speakers, if they do indeed occur.

Thusly, there are two main questions which have motivated this research. The first attempts to examine the salient features of beginning L2 learners' interactions with proficient speakers in the Chinese night market. It seeks to understand the role that gestures play in communication in relation to speech, as well as to develop a better understand as to whether these gestures could be said to
"compensate" for somewhat limited vocabularies of beginning level language learners. Second, the study attempts to understand whether L2 learners can develop linguistic knowledge from these interactions. It seeks to understand whether gestures can aid in this development and begins to examine which factors may encourage or inhibit language learning in this context. This essay will first discuss the relevant literature, subsequently describing methodology, results, discussions and conclusions.

Chapter 2 - Literature Review

2.1 Language Learning Through Interaction

The nature of second language learning and acquisition has long been debated, producing countless theories and approaches which have portrayed sometimes divergent and competing notions as to how language is actually learned, which factors influence this process, and to what degree. One approach that has remained relatively well-accepted in the last thirty years has been the Interaction Hypothesis (Long 1981). While previous hypotheses focused on the comprehensibility of language being used to teach a learner (Krashen 1977), many criticized these approaches for placing too much emphasis on input, (i.e. the language that a learner is exposed to throughout the learning process) while ignoring the output (i.e. the language that a learner is actually producing). Thus, Long's (1981) Interaction Hypothesis claimed that comprehensible input for the learner and L2 development occurs during modified communication requiring both input and output between proficient and non-proficient speakers of a language. It is through the process of modification and adjustment of language that meaning is created and language can develop for the language learner. Following Long's work, some have focused on the role of learner output in L2 development (e.g., Swain 1985; White 1991), while others have examined the roles of feedback and attention in the language development process (e.g., Schmitt 1990), leading to debates of implicit versus explicit instruction and their effect on L2 learning. While it seems that which aspects of language benefit from interaction, what kinds of interaction are most beneficial (including which forms of feedback, in which circumstances and to what extent), and the role of individual variables in L2 learning and acquisition are still in need of further investigation (e.g., Mackey, Abbuhl & Gass 2014), it now seems to be accepted in much of the SLA academic literature that interaction plays an important role in the language learning process (Gass & Mackey 2007).

Central to this interactionist approach is Vygotsky's (1978) assertions that learning occurs as a social activity. It is through gradually accumulated and negotiated social interaction that cognitive and
linguistic development is able to occur, a claim which is central to sociocultural approaches to L2 learning (e.g., Lantolf 1996, 2000). Ohta's (2000) study on two adult L2 learners of Japanese in interaction is often used to illustrate this point. Using micro-analyses of conversation, Ohta shows that through collaboration and subtle requests for assistance, a more proficient learner was able to help a less proficient learner overcome a gap in their ability to produce certain Japanese grammatical structures. Through carefully articulated cues and mutual sensitivity, the learners were able to gradually accumulate implicit knowledge about the structures, with retention recorded after the task. Identifying social interaction as the source of this learning, Ohta emphasized the view that interaction is fundamental to language development. While a number of other studies have since reinforced this claim, others have commented that human interaction entails more than just speech, and in order for a more holistic understanding of how interaction affects language development, other modes of communication (notably, gestures) need to be taken into account (Lantolf 2000, 2010, 2014), especially when considering the variation of learner styles (Purpura 2014). However, the field of gesture studies and language acquisition has developed somewhat independently from sociocultural theories of interactional learning.

2.2 Gesture and Language Development

2.2.1 Gesture and First Language Acquisition

Evidence for gesture as fundamental to human communication is especially salient in the field of gesture and first language (L1) acquisition. Initial interactions by infants are characterized by a combination of gesture, notably pointing, and vocalizations to create a message described as a "single holistic utterance" (Liszkowski 2010: 46), though one that is intentionally created by the infant and perhaps has a more general social function, possibly related to the development of joint activities and actions. Gestures are also seen as important in first language acquisition in that they help older children track referents in narratives (Cristilli 2014), while others have seen the gestural modality in children as "an alternative system of communication that can be utilized when language is not possible" (Guidetti et al. 2014: 359). Some have questioned the degree to which gesture and speech are actually integrated in infants, in that gesture and speech occur synchronously and with semantic coherence only after the beginning of the one-word phase, converging during this phase, and possibly allowing for the transition to the two-word stage (Butcher & Goldin-Meadow 2000, 2003; Liszkowski 2010). Regardless, after the age of one, gesture and speech seem to develop together (Gullberg & de Bot 2010), and many argue
that after this stage the speech-gesture system becomes increasingly integrated (e.g., McNeill 1998, 2000; Kendon 2004).

2.2.2 Gesture and Second Language Acquisition

In L2 development, gestures provide both communicative and cognitive functions, fostering language learning in various ways (e.g., Gullberg 1998; Gullberg 2006; McCafferty & Galestam 2008; Quinlisk 2008; Gullberg & de Bot 2010). In terms of L2 learner production, some have noted that L2 learners tend to gesture more than native speakers in communication (Kita 1993; Gullberg 1998; Nicolaidis 2007). Some of these studies involve the retelling of a narrative in an individual's L2, comparing frequency and occurrence of gesturing to native speech. In L2 narrative retellings, it seems that L2 learners tend to mark previous topics more explicitly with deictic (i.e. 'pointing') gestures than in native speech (Gullberg 1998, 2003). Similarly, Yoshioka (2010) found that in L2 Japanese learners' narratives, re-introduced referents were gesturally marked more overtly than native speakers' productions, suggesting that this overt marking occurs to allow for hyper-clarity, though it is unclear to what extent this gesturing functions as speaker-internal aid or as communicative clarifications benefitting the listener. Interestingly, L1 gesturing has also been shown to be affected by learning another language. Again in narrative retellings, Japanese speakers who had been learning English tended to gesture more like native English speakers than native Japanese speakers, regardless of whether the narrative was told in their L1 or L2 (Brown 2010). Moreover, in classroom research, Morett & Gibbs (2014) found that L2 students tend to gesture more frequently when an interlocutor (i.e. the teacher) was present than when they were not, suggesting that gestures are used communicatively and for the benefit of the listener. While the issue of whether L2 learner gestures serve mostly cognitive or communicative functions, to which extent, and in what contexts is yet to be settled, studies have approached the issue from a focus on 'thinking-for-speaking' (e.g., Slobin 1991; McNeill 1992) or what gesture can reveal about L2 thought processes (e.g., Stam 2008; Platt 2008), and from a focus on gestures' communicative functions (e.g., Olsher 2008). However, many now note that both views must be studied for a holistic understanding of gesture, speech, and language development (e.g., McNeill 2000).

2.2.3 Gestural compensation for speech
Inextricably tied to the issue of speaker-internal versus interactional gesturing is gestural compensation for speech. Following an earlier observation that gestures do not always simply replace speech in L2 learner productions (Gullberg 1998), Gullberg & de Bot (2010) argue that the over-production of gesture in L2 learner speech should not always be seen as a replacement for missing or inaccessible speech, but rather as a part of the process of language development. In this view, gestures are seen as fulfilling multiple roles related to fluency, working memory and cognitive load throughout this process (Gullberg 2006, Gullberg & de Bot 2010). Yet, it is claimed that gestural compensation can be multifunctional and operate at both speaker-internal and interactional levels (Gullberg 2008: 104, 2010: 21), with learner awareness of their own gesturing likely indicative of the extent to which gestures compensate for speech at each of these levels (Gullberg 2011). Fluency in narrative has been the focus of many studies investigating gestural compensation for speech. While it had been claimed that lack of fluency may encourage more gesture use by the L2 learner (Gullberg 1998, 2008, Gullberg & de Bot 2010), recent studies have shown that gesture in fact co-occurs more often with fluent speech than in non-fluent speech in both L1 and L2 (e.g., Cristilli 2014), supporting earlier studies of stuttered speech (Mayberry & Jacques 2000). Others have argued that gesturing works to allow for lexical retrieval (Krauss et. al. 2000), while more recent research argues that compensatory gestures in narratives allow for easier access to the processing and expression of propositional contents, supplementing and disambiguating incomplete or inaccurate information (Cristilli 2014: 345-346). Regardless, many do admit that gesture can compensate for speech when linguistic devices are used that have not yet been fully acquired (Kita 2000; Goldin-Meadow & Butcher 2003; Gullberg 2011; Guidetti et. al. 2014). At the moment, the majority of these studies have examined narratives of intermediate level or above L1 and L2 learners, and this focus may impact the current research trends in gestural compensation for speech.

2.2.4 The role of input

Studies have also shown that gestures can aid in L2 learning in the classroom and laboratory when used as input for the learner. Using gestures simultaneously with French expressions to teach L2 learners, Allen (1995) found that those who were taught in this manner had greater retention and recall of expressions than those who had only been taught the expressions without gesture. This supported earlier arguments that non-verbal communication may be able to facilitate L2 development (Neu 1990). Furthermore, it has been shown that teachers frequently use many gestures to aid students' understanding in classrooms (Allen 2000), and the benefits of this form of instruction in the classroom have been noted for aiding in vocabulary development (Kellerman 1992; Lanzaraton 2004) as well as
helping to create a 'zone of proximal development' (McCafferty 2002). Gesture is now seen as important in L2 instruction in that it can enhance comprehension (Sime 2008) and negotiate and facilitate language learning by providing "metalinguistic commentary" from teachers to students (Faraco & Kida 2008: 294).

Naturalistic input also seems to have an effect on L2 learning, with some studies showing learning to occur even at the earliest stages of L2 development. Using "naturalistic but controlled audiovisual input" (5), a weather report designed to highlight target words in Mandarin Chinese, Gullberg et. al. (2010) found that adult Dutch speakers with no previous contact to Mandarin Chinese were able to segment a continuous stream of speech and develop some sense of the phonotactic rules in as little as seven minutes of contact with the previously unknown language (16). More importantly, participants were able to map meaning to sounds when gestures accompanied the words, though gestures alone were not sufficient and required 'accumulation' through repetitions (16). That learners were able to discern this information from a continuous stream of speech in such a short time frame suggests that language learning "in the wild" does occur to some extent, and for adult L2 learners, this process is aided by gestures (5). Furthermore, it this focus on naturalistic input in various social contexts and environments that has increasingly been of interest to some sociolinguists within the field of SLA.

2.3 Language Learning "Out-of-class"

While there does seem to be quite substantial evidence for the importance of examining both interaction and gestures to better understand human communication and language development, the vast majority of studies in both fields have taken place in the laboratory or classroom. While some have argued that a change of setting would likely not have a significant effect on L2 development (e.g., Long 1998), other studies have reported that social context does impact the language learning process (e.g., Schumann 1978; Tarone 1988, 2009). As Tarone (2000) has argued, variation occurs in the language learning process by individual and social context, affecting outcomes. Furthermore, if we consider that many of the language learners worldwide are illiterate, then the current SLA approach to a focus on the classroom is not representative of the range of language learning circumstances or learner types outside of these contexts (Bayley & Tarone 2014). Moreover still, classroom environments are conducive to mostly formal interactions and these interactions may not transfer well to more informal, 'out-of-class' interactions and environments (Tarone and Swain 1995). Therefore, it has been argued that SLA research must broaden its scope.
These variationist approaches to SLA have called for language learning and acquisition to be studied "in a wide range of social contexts" (Bayley & Tarone 2014: 42), as have some sociocultural approaches (see Veronique 2014: 259-264). While considered by some to be separate sub-fields of sociolinguistics, both have attempted to study naturalistic language learning to some extent, with this term typically referring to the various factors, processes and social environments in which language learning can occur outside of the classroom or laboratory. Some of the L2 studies which have moved beyond the classroom have examined issues such as identity and social integration in immigrant communities (e.g., Norton 2000), while others have examined technology and computer-mediated communication (e.g., Lamy & Hampel 2007; Chik 2014), while others still have focused on L2 learner interactions with native speakers in the 'study abroad' experience (e.g., Adamson & Regan 1991; Bayley 1996; Richards 2014; Arnold & Fonseca-Mora 2015; Macalister 2015).

What is mostly obvious with all of these studies is that the benefits and drawbacks to such out-of-class L2 learning are highly variable and dependent on numerous individual (i.e. gender, motivation, willingness to communicate, etc.) and contextual factors (i.e. power structures, access to technology, host family personality, etc.) (e.g., Bayley 1996; Norton 2000; Paige et. al. 2004). It is because of the range of factors that influence language learning in various contexts, variationists argue, that studies on language acquisition must extend beyond the classroom.

In fact, language is learned in a multitude of different social environments and both variationist and sociocultural approaches have called for more research on this process. This research, while diverse, seeks to understand the ways in which language is learned outside of a classroom and with naturalistic input. While these studies have pushed for L2 development occurring 'out-of-class' to be included within the field of SLA research, again, the role that multimodal forms of communication play in L2 learner interactions in such environments has generally been ignored.

2.4 Multimodality in Micro-ethnography

Influential to each sociocultural theory, variationist approaches, and gesture studies is Vygotsky's (1978) assertion that learning is socially situated, beginning as external interaction and later becoming internalized within an individual's cognition, an idea later reinforced in sociolinguistics and linguistic anthropology with Lave's (1988, 1991) concept of 'communities of practice'. To those utilizing multimodal and micro-ethnographic approaches, understanding these processes requires that human communication is situated within this external environment. Research on multimodal communication has increasingly applied a number of methodologies to study gesture and speech outside of its
traditional settings - including ethnography, micro-ethnography, and conversation analysis - all of which emphasize gesture as an important part of interaction as it exists in natural settings.

Early studies in the field focused on culture as a variable in micro-ethnographic research. One of the earliest of these was Efron's (1972 [1941]) micro-analysis of gestures among Italian and Jewish immigrant groups in New York during the 1940's, providing evidence for a cultural origin for gestural forms that can change over the course of time. While the role of culture has been included in research on gesture and speech in more recent work (McNeill & Duncan 2000: 151-154; Kendon 2004: 348; McNeill 2008: 25, 60; Brookes 2014), for the purposes of this research, Jürgen Streeck's (2009) view that gestural differences between cultures are "stylistic" and not attributable to any single culture (38) will also be the view taken for this study. While not entirely unproblematic, this view allows the study to focus more intently on aspects of language learning and interaction, while acknowledging that cultural variables may have an impact on the interaction itself.

Streeck's work is further relevant to this research in his visual praxeological approach to studying multimodal communication, defined by Streeck as "the visual study of human practices" (2009: 6). Influenced by both conversation analysis (e.g., Goffman 1961) and context analysis (e.g., Scheflen 1974; Kendon 1990) of interaction, this 'micro-ethnographical' research views gesture as embedded in the natural setting in which it occurs. If the context changes, Streeck argues, so will the gestures. His research therefore uses micro-analyses of interactions to show how gesture, speech, and context work together to create group learning. This notion was explored earlier in Lebaron & Streeck's (2000) study of a do-it-yourself workshop. Through the step-by-step manipulation of objects, gesture forms evolve, accumulating shared knowledge, and thusly creating a 'community of practice' within that environment (131). It is through the multimodal interaction with not only interlocutors, but also with the environment and objects situated in the material world, that this learning takes place. Moreover, some researchers within the field of gesture studies have taken note that gestures are often only interpretable when the context of the situation is considered. These so-called "environmentally-coupled gestures" (Goodwin 2007: 195) are often used in conjunction with deictic terms (i.e. 'this' and 'that') making them "inextricably tied to environment" (201). This means objects held, pointed to, or referenced must be included in the analysis of gestures and speech. Furthermore, this concept has emphasized earlier studies which have revealed the complex ways in which gestures, speech, and objects influence each other's behavior in human interaction (e.g., Haviland 1993, 2000; Streeck 1996; Goodwin 2000, 2003; Mondada 2007a), situating gestures as a key connection between environment, cognition, and social learning.
However, to date few studies have examined how multimodal communication occurs in L2 learner interactions in naturalistic settings, despite both the accepted importance of studying language learning in such environments as well as the role of gesture in linguistic development. One notable example is Streeck's (2009; 2014) 'visual praxeological' studies of mechanics working in a car-repair shop. Through interactions between the owner of the shop and various employees, Streeck reveals the multitude of ways in which L2 speakers can communicate complex ideas and actions, vital to the proper functioning of the shop, using hands, speech, eye gaze, and reference to objects at hand and farther afield. While linguistic development is not the focus of Streeck's studies, this research exemplifies the importance of many elements of human communication being used in conjunction to provide meaning that is mutually understood by participants, even when linguistic resources to communicate such goals may be limited.

This section has reviewed a diverse set of literature regarding interaction, language learning, gesture, and context. To briefly summarize, interaction is seen as key to language learning. Gesture is part of this interaction. Many have claimed that language learning may occur in interactions outside of classrooms and that this learning process should be studied. Micro-ethnographical studies which examine both gesture and speech in naturalistic settings may be the "meeting point", allowing for studies of L2 language learners to take into account the multitude of environmental factors and modes of communication which may lead to linguistic development. However, exactly what is meant by gesture in this study must first be defined and specified.

2.5 Gestures and Multimodal Communication

Despite the plentiful research on gestures, there is in fact not a consensus on what gesture is, perhaps a result of the complexity of the multimodal system. While some have considered gesture to be "communicative action of the hands" (Streeck 2009: 4, emphasis in the original), others have classified the term multimodal communication as "all the oral features of spoken language and all the kinesic features of bodily movement that play a role in communication" (Calbris 2011: 37-38). Due to the limitations of this study, four features of multimodal communication have been taken into account in that they were deemed to be most relevant to the research questions. The first of these is speech and vocalization, and will be explained further in Section 3.4.1. The following categories are hand gesture, eye gaze (along with head movement) and facial expression.

2.5.1 Hand gesture
The hands were linked to speech in Kendon's seminal (1980) study. In fact, numerous typologies have been used over the years to classify gesture types, beginning with McNeill's (1992) four classifications. While this typology has been challenged on a number of grounds with new typologies or variations proposed (e.g., Gullberg 1998, McNeill 2000; Kendon 2004; Streeck 2009; Guidetti et. al. 2014), today there does not seem to be a consensus that one typology or another should be used by all within the field of gesture studies. Moreover, some have commented that, given the nature of gestures, we cannot establish universal categories, making these typologies mostly useful for terminology and analysis (e.g., Kendon 2004; Streeck 2009). This study will therefore introduce terminology as needed.

However, one feature of gestures which has remained relatively constant in the literature has been deixis, gestures that point, either towards an object in the immediate vicinity, beyond the interlocutors' line of sight, or as reference to re-introduced topics in a narrative. In fact, even those deictic gestures which point towards an object interlocutors can see requires quite complex processes in order to achieve joint attention (Haviland 2000; Mondada 2014) and have been relatively under-researched (Kendon 2004). Furthermore, given that deictic gestures have been shown to be used for communication between L2 speakers (e.g., Streeck 2009), they would likely play a role in this research. However, it should be noted that deixis can also be performed with other parts of the body, including the eyes, and is not necessarily tied only to the hands (McNeill 2008).

2.5.2 Eye Gaze

Early studies in what would now be called multimodal research took note of the role of eye gaze in mediating social interactions. Kendon (1967) found that gaze behavior was different during speaking than during listening. When participants began speaking, gaze was turned away from the interaction, while listener's tended to gaze for longer at the speaker. Kendon found this behavior to be related to speaker thinking processes, as well as a means to 'hold the floor', returning towards the end of an utterance possibly as a means to 'offer the floor' to the listener. Listener gaze, on the other hand, serves as response to what is being said by the speaker and as accompaniment to speech. In this sense, gaze direction regulates the behavior of interlocutors during interaction for the purposes of turn-taking.

However, this notion has recently been questioned, with a focus on the "action-sequence", or the kind of interaction (i.e. question and answer versus narrative format), being proposed as the determiner of gaze in interaction (Rossano 2012). Others have claimed that visual attention may mark
aspects of the accompanying speech and is used as meta-communication to indicate salient features of the interaction, with a focus not on attention but rather on recognition of the message (Streeck 2014).

2.5.3 Facial Expression

In fact, as Bavelas et. al. (2014) state, facial expressions in conversation have been mostly neglected in the academic literature. Following Chovil (1991) and Kendon (2004), Bavelas et. al (2014) provide evidence that facial gestures are not just reactions to emotional states. Facial expression has numerous functions in interaction (e.g., Kendon 2004), both intended and unintended. Given the lack of descriptive studies on facial gestures, they should be studied in their own right, however the amount of attention they have received in this project is limited (see Section 3.4.1).

Thusly, this research is not focused solely on the hands (e.g., Streeck 2009) nor will is consider the whole of the multimodal system according to Calbris (2011), though it does take these features of multimodality to be interwoven. While undoubtedly limited, these were the features best able to be analyzed given the nature of the data. The following section explains this data relative to the research design, the participants, and the procedures followed for the collection of data and explains the analysis of these features.

Chapter 3 - Methodology and Design

3.1 Data & Design

This study consists of two phases, both utilizing audio-visual recordings. The first consists of observations of L2 learning participants in interactions with proficient speakers in the street markets, while the second involves interviews between the researcher and the L2 learning participants. This data was gathered between the dates of May 29th and July 6th, 2015 around the Yinzhou District of Ningbo, Zhejiang, China.

3.1.1 Justification of Design

In that this study has proposed to study language use in the natural world and utilizes interviews to triangulate gathered data, it is in some sense ethnographic. Ethnographic research is often qualitative (in that its data cannot be 'counted') and attempts to describe participants and actions in natural settings using a variety of methods, with observation and interviews being some of the most common (Leeds-Hurwitz 2005). Hyme's (1962, 1974) assertions that language should be studied within
the context in which it is used were crucial for language being studied within this approach. Cohesive with Vygotskian claims earlier described (see Section 2.4), ethnographic communication studies have since taken contextual factors to be vital in providing a more holistic picture of language in interaction.

However, the focus of this study is on the interactions themselves, while maintaining that it exists within a social context. Therefore this research is similar to conversation analysis in its micro-analysis of interactions (e.g., Sacks, Schegloff & Jefferson 1974; Ohta 2000), yet it also considers nonverbal forms of communication, and in this sense, it is micro-ethnographic (Leeds-Hurwitz 2005). Thusly, the use of audio- and visual- recordings are vital for repeat viewing and in-depth analysis of multimodal communication used during the interactions. Therefore, this work has much in common with the research previously mentioned by Jürgen Streeck (see Section 2.4), an approach that has also been taken by a growing number of those in the field of gesture studies (e.g., Lebaron & Streeck 2000; Goodwin 2000, 2007, 2014; Mondada 2007a, 2014). These studies examine the multimodal communication system in natural settings using micro-analyses, examining salient features in participant interactions.

While interviews are not commonly utilized in these micro-ethnographic studies of multimodal communication, they were deemed necessary for this study in order to determine the nature of linguistic development of L2 learners. Observation and conversation analysis alone would likely not allow full access to this information given the size of the study. While language learning may indeed be interactional at some level, it is of course internal at another. As some have argued, interviews can help understand both of these processes (Hymes 1974; Goodwin & Duranti 1997). Additionally, in studies of gestural compensation for speech in L2 learners, it has been suggested that determining the level of awareness learners have of their gestures and speech difficulties may be important in determining whether gestures are used as compensatory (Gullberg 2011).

Another purpose of utilizing interviews was to aid in providing 'thick description' to the analysis (e.g., Geertz 1973). While multimodal analysis of interactions alone can provide some information into the factors which underlie these communicative acts, understanding motivations, intentions, and emotions which accompany the individual throughout the interaction can provide additional layers of meaning from which to interpret the interactions and vice versa.
3.2 Participants

3.2.1 Beginning L2 Learners of Mandarin Chinese

Participants for this study were gathered through personal contacts using WeChat, a form of social media commonly used in The People's Republic of China. In fact, gathering participants by these means has been justified in the past (e.g., Milroy 2002), as using one's own social networks allows for an 'emic' perspective on participant viewpoints. It also addresses the problem of artificially grouping individuals who may not otherwise be grouped, following Hymes (1974) assertion that "there are no masses, only ways to regard people as masses" (8), though it should be noted that emic perspectives are not without their own limitations. Participants were gathered in this fashion throughout the course of data collection. Overall, this led to five participants, all residents at the time of data collection of the Yinzhou District of Ningbo, China, with two identifying as female and three identifying as male. Four of the participants were post-graduate students at the University of Nottingham Ningbo, while one was an English teacher in the city. Two of the participants were native English speakers, two were native German speakers, and one was a native Dutch speaker. However, all participants deemed themselves to be proficient speakers of English, and post-graduate students at the University of Nottingham Ningbo are required to pass the TOEFL (Test of English as a Foreign Language), a widely used test for English university acceptance, with a score of 87 in English proficiency. Because of this, communicating in English with the participants was not determined to be in any way limiting to the participants ability to understand the written ethics forms or communicate with the researcher in conversation or during interviews. All participants were between the ages of 22 and 35 and had lived in China for under one year at the time of data collection. Through unrecorded conversations as well as recorded interviews in the study, all revealed themselves to be at a "beginner" level or lower in Mandarin Chinese. All participants were fully informed of the intent of the study (see Appendix 1) and signed written consent forms (see Appicences 2-6). However, prior to their interactions with the proficient speakers they were not told specifically which aspects of their interactions would be analyzed in an attempt to avoid changing their behavior while on camera. These participants were involved in both recorded observation and interview phases of the study.

3.2.2 Proficient Speakers - The Street Vendors

The other participants in the study were "proficient speakers" of Chinese, and were the various vendors around Yinzhou District, selling goods and food. They were chosen somewhat by L2
participants of the study, or mostly by chance (if they happened to be where the researcher and participants were filming), but there was no precedent to choose one vendor over another by the researcher in an attempt to make the interaction between L2 learner and proficient speaker as natural as possible. Indeed, during the interview phase with L2 learners, it was mentioned that their choice in vendor was based mostly on recognition of the items vendors were selling. Some participants mentioned having frequented some of the vendor stands that were the setting for interactions in this study, possibly affecting the interactions themselves.

All proficient speakers were fully informed of the intent of the study by a translator fluent in both Mandarin Chinese and English (see Appendix 7). This translator also obtained verbal consent from these participants. Written consent forms were deemed inappropriate in these cases by the researcher in that some of these market places are not fully licensed, and vendors and shopkeepers may be unwilling (or unable, due to literacy issues) to read and sign the relevant forms. Vendors were also given the choice to have their faces blurred and voices modified in the video recording, though all declined. Video-audio recordings did not begin until consent was reached. The translator and researcher made concerted efforts to ensure vendors did not feel pressured to participate in filming, allowing these participants to stop the filming at any time. However, verbal consent was reached without difficulty and with a smile on the condition that we were not employed by law enforcement agencies. We were able to reassure them that we were not.

These proficient speaking participants were not involved in the interview phase of the study. Given the size and time constraints of the study, it was deemed more important to engage time and resources into an in-depth analysis of the L2 learners, allowing for the research to focus more directly on the language learning process and the intents of these participants. However, that these participants were not included in the interviewing phase is a limitation of the research and will be discussed in Section 5.

3.3 Procedures

3.3.1 Street Market Interactions

Street market interactions took place at two locations around Yinzhou District, Ningbo, China. The first street market was positioned just outside Gate 3 of the University of Nottingham Ningbo and was located along the side of a main street, while the second was located at about a 15 minute walk from the campus in the plaza of a residential area. Once the researcher, the L2 learning participant, and
the translator were gathered and appropriate documents signed, travel to the one of the locations was organized either on foot or by bicycle. Time of day affected the location the research would take place, as the markets are constantly moving in order to find customers.

Once an L2 participant chose a vendor's stand, the translator was asked to translate for the researcher, explaining details of the study and receiving verbal consent. Again, filming could only begin after this process had taken place. On one occasion, we were told that law enforcement officers were nearby and that it may not be appropriate to begin filming, thus requiring a change in location. However, in all other interactions, the process of verbal consent was obtained without difficulty.

The L2 learning participant was equipped with an audio microphone as a "necklace" in order to better capture sound, as can be seen in the video recording. Once filming began, the camera was unmounted and manipulated by the researcher. The unmounted nature of the camera was necessary in that people and objects are constantly moving in and out of the frame. Similar to what has been argued by Streeck (2009), the unmounted camera allows the researcher to move and capture vital features of the interactions. Key to this movement was keeping all participants' hands, upper bodies, faces, and heads in frame. It was thought that this would allow for the collection of the most relevant data. While these interactions often began as dyadic, or between two people, others, either in the frame of the camera or off-camera, often became involved in the interaction. At times it seemed, the camera and filming attracted much curiosity, as can be seen in the video recordings. However, due to the multimodal nature of the research and the limitations of the camera (it certainly could not film in 360 degrees), only those interacting with the L2 language learner within the camera frame were considered participants, at most allowing for triadic, or between three people, interactions. In all cases, participants other than the L2 learner were those who had given explicit verbal consent to allow for filming and audio recording. In this sense, the "focal event" was defined as L2 learner interaction within the frame of the camera (e.g., Goodwin & Duranti 1997).

Interactions, or 'communicative events', were defined as beginning with initiation of face-to-face interaction between the L2 learner and the vendor and ending when this face-to-face interaction ceased (e.g., Hymes 1974, Saville-Troike 1982). While this delineation is somewhat contrived in that the process of receiving ethical consent to film could be seen to 'break-up' a natural flow of interaction, it is not entirely un-naturalistic, in that without the presence of a researcher or camera these interactions would likely begin and end with the same boundaries. However, an 'observer's paradox' must be noted during these interactions, as can be said for much of ethnographic research (Leeds-Hurwitz 2005). Previously mentioned issues of legality may have modified vendors' actions while being filmed, even
with the promise that these audio-video-recordings would not be shown to law enforcement. The camera would likely have also affected L2 learning participants' interactions, perhaps because of a lack of confidence in speaking Mandarin Chinese as a beginner, leading to participants using only "safe" language and afraid to make linguistic "mistakes" while on camera. While we attempted to mediate these issues with ethics forms, translators, and by allowing any participants to avoid taking part in activities that they might deem too uncomfortable, these issues are still present and somewhat unavoidable. The positioning of the camera itself is also somewhat problematic. By using a video-recorder, there becomes an invisible "boundary" within the line-of-sight of the camera which is often avoided by those moving around the street market. While this boundary is necessary for capturing the focal event, it modifies the surrounding environment and must therefore be noted. The camera's previously mentioned attractive qualities to those in the street market also must affect the focal event, in that there becomes a sort of "set" with participants watched by those off-camera. Our "actors" noted the "odd feeling" of being the focus of so much attention, but it was thought that the task at hand, ordering and acquiring the desired food, would at some level move their focus towards the interaction itself and away from the camera and onlookers. Regardless, it must be noted that studying in natural environments, and with a camera especially, must to some degree always affect the communicative event.

3.3.2 Interviews

All L2 learning participants were involved in interviews which used both audio- and video-recordings in order to capture a greater amount of data (Dufon 2002) and to account for gesturing. These took place at the researcher’s apartment following the interactions at the street markets, sometimes immediately afterwards and sometimes up to a week after filming at the street markets. In all cases, participants were shown the audio and video footage of their interactions before the interview, allowing the event to be immediate in the participants’ minds. All interviews were conducted in English.

Given the nature of the research questions, these interviews were semi-structured (e.g., Madison 2011). While there were set topics to be discussed beforehand, the interview was not limited to these topics, as motivations, intents, emotions, etc. were all considered to be factors in understanding their interactions with proficient speakers (Hymes 1974; Goodwin & Duranti 1997). The prepared topics are listed below in Figure 1.

Figure 1
1) Could you describe your language learning background/ living abroad experiences?
   1a) How did you learn these languages?
   1b) Is this similar or different to China/Chinese? Why?

2) Could you describe your interactions with proficient speakers of Chinese?
   2a) How easy or difficult is it? Why?
   2b) How do you feel about these interactions?

3) What about interactions with street vendors?
   3a) How do you communicate with them? What strategies do you use?
       3ai) Why do you use them?
       3aii) What about your hands? Your face? Your eyes?
       3aiii) How effective are these strategies?
   3b) How do you feel about these interactions?
   3c) How, do you think, the vendors feel about these interactions?

4) What about language learning? Are there any words/phrases/gestures/etc. you have learned from these interactions?
   4a) If so, how do you think that happened?

5) Is there anything else you would like to talk about regarding interactions with proficient speakers of Chinese?

While detailed, in practice they were not always presented in this order as the interview attempted to follow that of a natural conversation as much as possible. Similarly, some of the questions became redundant and were therefore omitted. For example, if a participant noted that "smiling" was an effective communication strategy in the early stages of the interview, question 3aiii) regarding the face became redundant, or could then be modified to a question of why they believe it to be effective. Furthermore, in some instances there were aspects of the previously video-recorded interaction in the street markets that were explored in greater detail.

Given the focus on multimodal communication and language learning in this study, questions 3) and 4) were seen as the core section of the interview. Question 3) relates to the first research question, attempting to understand to what extent the interviewee is aware of the communicative strategies used in their interactions with proficient speakers. As has been discussed in Section 3.1.1, level of awareness of gesturing could account for some instances of gestural compensation for speech, with follow-up questions attempting to understand to what extent it is effective or ineffective in aiding communication. Question 4) focuses on language learning, and while direct, attempts to understand the nature of linguistic development if indeed it does exist. Other questions attempted to provide background on
motivations, beliefs, emotions, etc. which were thought to provide 'thickness' to descriptions (e.g., Geertz 1973; Hymes 1974). Again, these questions are meant to provide some understanding of their previously recorded interactions with proficient speakers and were not taken as definitive answers to the research questions.

While it is noted that these interviews themselves are somewhat 'context-bound' (Brewer 2000), in that these participants were gathered as personal contacts mediates some concerns regarding the nature of interviewer-interviewee interactions.

3.4 Data Treatment and Analysis

3.4.1 Data from Street Market Interactions

Audio- and video-recorded data was uploaded from both devices onto the researcher’s computer. These data were synchronized using ELAN software (e.g., Lausberg & Sloetjes 2009). Transcription and analysis were heavily informed by Mondada’s (2007b) use of the same software, which allows a variety of communicative features to be analyzed in a parallel fashion with respect to the timing of the recording. The data was then coded by participant for each of the features of voice, hand gesture, and eye gaze and head (including facial expression), with pseudonyms given to all participants. Since these interactions took place in Mandarin Chinese, transcription was done in pinyin, the Romanized writing system for the language, using the Jefferson Transcription Method (2004) in a similar fashion to Mondada's (2014) study of joint attention. Mandarin Chinese is a tonal language and pinyin can account for this feature of the language. However, both tone and pronunciation were transcribed as would be a "standard" dialect, called pǔ tōng huà, even if this was not how they were pronounced by participants except in extreme cases. The purpose of this was two-fold. The area surrounding Ningbo is home to a dialect known as níng bō huà, which may differ in some pronunciation and tonal features, while the L2 learning participants may have knowledge of features of either pǔ tōng huà or níng bō huà, depending on their language learning background. Pǔ tōng huà, however, is the lingua franca. Transcriptions of the interactions were aided by the fluent translator who assisted the researcher during street market interactions (see Appendix 7). Deviations of pronunciation and tone for both proficient and L2 speakers were taken as stylistic (Hymes 1964, 1974), and while potentially important in understanding portions of the interactions, were outside the scope of this study. In some cases, English was used by the L2 participants and was transcribed accordingly.
Hand gesture was coded in terminology mostly from Mondada (2007b; 2014). The focus in this coding was on communicative intent of the gesture and not on micro-features of the gesture phrase or unit (e.g., Kendon 2004), similar to previous micro-ethnographic studies of multimodality (e.g., Goodwin 2007; Mondada 2007a; Streeck 2009), for example, "x points towards y", though Calbris (2011: 104-124) was frequently referenced for terminology. It should be noted however that coding is always somewhat subjective (Calbris 2011). Eye gaze was coded for in terms of directionality, for example, "gaze towards object x" for each participant during interaction. Gaze direction was mostly easily established, though due to lighting and the angle of the camera, it was not always possible to determine, for example, if x's gaze was towards y, y's hand, or the object being held or pointed to by y. The limitations this creates will be discussed in Section 5. Similarly, due to the camera angle needed in order to capture the interaction, facial expressions could not always be fully described. Because of this, and in response to what was revealed through interviews, only one facial expression was noted, that of "smiling" or lack thereof. The head was accounted for only during movement, for example, "head nod, vertical", similar to Mondada (2014). While there are limitations to being able to code aspects of the interactions given lighting, camera angle, and environmental variables (i.e. the movement of people around the camera), the focus, again, is on communicative intent and micro-features were not noted unless they were seen to be vital in understanding a given communicative action.

3.4.2 Data from Interviews

Interview data took the form of both audio- and video-recordings. These were uploaded onto the researcher's computer and synchronized using ELAN. These interviews were conducted in English and were transcribed using common orthography. In some exceptional cases, they were coded similarly to data mentioned in the previous section to allow for the analysis of multimodality. However, the focus of these interviews was mostly on content, and micro-analyses of the conversation between the interviewer and interviewee were not deemed to be necessary given this focus in most cases. Analysis of both data sets from street market interactions and interviews was therefore used as 'triangulation'. In this sense either data set may influence the analysis and interpretation of the other. However, it is the interaction in the street markets that is the primary focal event being studied. By using this inter-directionality of interpretation, it was thought that a fuller picture of street market interactions may be reached, thus allowing for more relevant information regarding the research questions.
The following section will detail the results recovered from the data and these processes of analysis. They are separated by research question, somewhat artificially, in that they occur within the same interaction and are therefore connected at many levels. These implications will later be discussed.

Chapter 4 - Results

4.1 Salient Features of Interactions

4.1.1 Overview

Multiple modes of communication were frequently utilized by participants in all street market interactions. Particularly salient are features of deictic hand gestures, eye gaze, and facial expressions in that they all seemed to play communicative roles in the interactions. Furthermore, all of these forms of communication, along with speech, can be seen to work together to communicate goals and create meaning in complex ways. Moreover still, data from interviews has revealed that these are sometimes used intentionally by the L2 learners in order to allow for successful communication between themselves and their interlocutors.

4.1.2 Deixis and Eye Gaze in Joint Attention

One feature of these interactions that seemed particularly salient was hand gesturing, chief among these being deixis, or pointing. Commonly, this was done using an extended index finger directed towards an object and was done concomitantly with the utterance "zhè gè", roughly translated to "this (one)" in English. This form of communication was seen in all of the participant interactions to varying degrees. Similarly, this was seen in both L2 learners and proficient speaker productions. The combination of gesture and speech required an object in the immediate environment to be referenced with this combination. Importantly, neither gesture, speech, nor the combination of the two could be understood without the attention of both participants towards the object.

While seemingly a simple task, it required full engagement and group effort of participants in order to achieve mutual understanding. One example shows how this communicative strategy works through sequencing of communicative actions, and through subsequent clarifications from other participants, mutual understanding is achieved. In this example, Rifkin, a native English speaker and postgraduate student, is attempting to order a flat bread. There are a number of possible ingredients that must be chosen from and they lie on a table between himself and a vendor. He begins the selection
process by pointing towards objects with an extended index finger concomitantly with the phrase "zhè gè", while the gaze of other participants follows the movement of this hand, seen in Annotation 1.1.

**Example 1**
1) 1a refers to Rifkin, 2b refers to Vendor 1, 3c refers to Vendor 2 (not speaking in this section)
2) White arrows delineate eye gaze direction, red arrows delineate pointing direction

(see Appendix 8 for video and audio files)

However, the message does not seem to be fully understood and the vendor asks him to clarify. She uses the same combination of deixis and the phrase "zhè gè" towards objects on the table while his gaze now follows the movement of her hand, shown in Annotation 1.2. However, again, there does not seem to be mutual understanding, as he corrects her with speech, "bù yào zhè gè" ("I don't want this") with the extended index finger landing towards the object simultaneously with the phrase "zhè gè", the vendor's gaze now following his hand again, seen in Annotation 1.3.
Example 1 (cont.)

1) 1a refers to Rifkin, 2b refers to Vendor 1, and 3c refers to Vendor 2
2) White arrows delineate eye gaze direction, red arrows delineate pointing direction

He then shows the desired objects with the same combination of deixis and the phrase "zhè gè hé zhè gè" ("this and this") while both participants point towards the corrected item, as the vendor utters, "zhè hǎo: oh hǎo de" ("This ok, oh ok") seen in Annotation 1.4. A second vendor then steps in to finally clarify. He points and asks, "là jiāo yào le a?" ("Did you want chili pepper?") shown in Annotation 1.5. Rifkin replies, "yeah, wǒ yào hěn là, hěn là" ("Yeah, I want very spicy, very spicy") pointing towards the same item as the second vendor in Annotation 1.6. He then lifts his gaze towards the second vendor, smiling.

Though the actual names of objects are barely used in this extract, meaning and mutual understanding are created almost entirely with joint attention, using gestures that are coupled with the immediate environment. In fact, this process requires that all participants' hand gestures and eye gaze are coordinated with one another's, as well as with the objects that are being referred to. When there is misunderstanding, it can be mediated by the same communicative strategy because of the coordinated effort by all participants. Thusly, the L2 speaker is able to communicate his goals with proficient speakers using limited vocabulary. Data from interviews was useful in understanding the purpose of using this kind of communicative act. Rifkin was asked why he was pointing at the objects during the interaction:
Rifkin 8:40

(1) Rifkin: ...I obviously didn't know the nouns for these things,
IR: Ok.
Rifkin: So I had to point to them to demonstrate that I wanted them.
IR: I see.
Rifkin: And I could say the word for "this" which I think everyone can in Chinese
IR: Yeah, yeah.
Rifkin: So I would just point at it and say "this" because that's how I could demonstrate it.

(see Appendix 8 for video and audio files)

This sentiment is echoed by Ana, a native German speaker and English teacher in Ningbo:

Ana 34:45

(2) Ana: ... if she would speak English or I would speak Chinese, I would say, "I would like bread with this mushroom paste, and just a little bit of chili and some herbs."
IR: Mhm, mhm.
Ana: So, I can't say that, so we both know she's just doing breads, so I'm saying I want this. But obviously I can't, by saying I want "this" I need to kind of clarify which one that is, so I'm pointing to it.

It is clear that these 'environmentally-coupled gestures' were used consciously by these participants and as a kind of communicative strategy, supplementing the fact that the L2 participant's linguistic vocabulary may be limited. Quite unsurprisingly, interviews revealed that all L2 participants were well aware that they used this communicative strategy, with most of them feeling that it was effective and rarely caused problems which would impede the interaction as a whole. However, when further analyzed, this seemingly 'simple' way of communicating actually requires quite a complex organization of many elements working together using the entirety of the multimodal system, a fact which becomes clear when there is greater miscommunication between the participants.

In this example, Ana is waiting to claim food which she had just ordered, coincidentally from the same vendor in the previous example. She extends her left index finger towards an object without vocal utterance, the flat bread which she believes to be her own and is now cooling on the edge of the kiln, ready to be claimed. Her gaze is initially towards the flat bread, but quickly shifts towards the first vendor, her intended interlocutor at this moment. The vendor is currently busy preparing another object, and after a short delay her gaze eventually lands towards the flat bread being pointed to, creating joint attention, seen in Annotation 2.1.
Example 2
1) 1a refers to Ana, 2b refers to Vendor 1
2) White arrows delineate eye gaze direction, red arrows delineate pointing direction, and blue arrows delineate head movement

Ana notices that her interlocutors’ attention is now at the same object, and the same index finger used for pointing to the object is then directed towards herself while nodding, seemingly asking, albeit without vocal utterance, if she should take the flat bread (seen in Annotation 2.2), while the vendor also nods her head. However, two elements of the multimodal system are ‘missing’ - the vendor’s hands are currently busy manipulating other objects on the table and Isa is not producing any vocal utterance - leading to multiple misunderstandings.
**Example 2 (cont.)**

1) 1a refers to Ana, 2b refers to Vendor 1
2) White arrows delineate eye gaze direction, red arrows delineate pointing direction, and blue arrows delineate head movement

The vendor's gaze shifts towards the kiln and while nodding, explains, "nǐ de zài zhè biān" ("Yours is here"), though her hands are still busy with other objects, seen in Annotation 2.3.

Misunderstanding the vocal utterance, the direction of the gaze, or both, Ana reaches for and grabs the flat bread, nodding and uttering, "wǒ de, hǎo hǎo" ("Mine, ok ok"). Correcting Ana, the vendor raises her hands from the objects she is currently manipulating and points towards the flat bread in Ana's hands, explaining, "zhè gè bù shì nǐ de" ("This is not yours"), while Ana pulls her hands away from the flat bread, saying "ah bù bù" ("Ah, no no"), shown in Annotation 2.4. The vendor then, pointing with an extended index finger into the kiln while gazing in the same direction, further explains, "nǐ de zài lǐ miàn a" ("Yours is inside here"), and Ana's gaze is drawn towards the correct flat bread inside of the kiln, seen in Annotation 2.5. It is only when joint attention on the correct object is explicitly secured through deictic gesturing and gaze that mutual understanding occurs, and Ana backs away from the interaction.

In this example, we can see that hand gesturing alone could not replace spoken language and still be understood without joint attention of participants. While perhaps pointing is used to compensate for speech by the L2 learner, it is not entirely effective in communication without an understanding of the accompanying speech of the interlocutor, the direction of gaze of the interlocutor and joined attention, or a combination of these. Thusly, while these 'environmentally-coupled gestures'
are effective to some extent in communicating goals in street market interactions, they are dependent on the whole of the multimodal system including the accompanying eye gaze and speech in order to create mutual understanding between all participants involved in a given interaction.

4.1.3 The Face, the Eyes, and the Body

Furthermore, all L2 participants have noted that there are other important aspects to interactions with proficient speakers which constitute successful interaction, emphasizing the interconnectedness of the multimodal communication system. While the conclusions that can be drawn regarding facial expressions in this study are quite limited due to the factors outlined in Section 3.4.1, they seem to undoubtedly be acknowledged as communicative and utilized intentionally by L2 participants in interaction with proficient speakers. Most notable of these is the use of "smiling". While noticeable in video-recordings of the interactions in the street market, it was through interviews with the L2 participants that this became illuminated. Nick, a native German speaker and postgraduate student, explains his purpose in using this strategy:

Nick 8:20
(3) IR: ... and then what, also about body language, what about things like your face? Is there anything that you do consciously?
Nick: Smiling.
IR: Smiling, ok...
Nick: Smiling always helps, [it, it,]
IR: Ok, [why? why?]
Nick: If you approach someone with a smile, it’s a way different, like... environment or way different mood...

Ana seemed to think of this similarly when asked about smiling:

Ana 18:30
(4) IR: So what’s your reasoning? Why then do you... do you do it consciously when you do it?
Ana: I think I try to smile more than I maybe would normally to make sure that they understand that I'm trying to be really friendly and not just arrogant and I can't speak or something. So I'm trying to be more friendly...

Ana eventually also explained that there was more to utilizing forms of multimodal communication, such as smiling, than the immediate interaction in which it occurs:
Ana 25:20
(5) IR: So you're saying there may be some misunderstandings or some miscommunication, but in general, uh, all of these things working together is effective communication?
Ana: It is because if the, um, and I think even on two levels because obviously my first goal is to get the food, I succeeded at that, and I think through like things like a smile and like a general friendly body language, I can. I mean I don't have that small talk with them, but I still almost become, uh, friendly with them. And people would like recognize me and kind of, the initial smile of recognition would be bigger each time.

Mike, a native English speaker and postgraduate student, explained how posture, eye contact, and facial expression have an impact on the nature of the interaction:

Mike 9:40
(6) Mike: ...ah, but I probably mainly get by on soft body language and smiling. So just basically trying to catch people in the eye and be unintimidating, because I think as a Western male, and you know I'm a meter eighty-six or whatever, you can be overbearing perhaps on people, so I try to give a bit more distance stance, sort of off center, you know that soft body, rather than squaring your shoulders to people. You almost do it intuitively. I'm kind of aware that I'm doing it, but I purposely try to get eye contact and just smile and try to get some empathy from people I think...

He later continued, explaining how all of these elements of multimodal communication work together for him and its importance:

Mike 51:20
(7) Mike: ...whoever I do go to, I will first of all try to build up a good empathy. Eye contact, smile, a bit positive body language, and then use a lot of hand gestures with basic Chinese to try to select the food, all the time trying to take signals off them, really trying to keep the eye contract. If they don't understand me or if I have to kind of... invite them on to just skip that question.

In fact we see that these interactions are seen by some of the L2 learners as a means to build "empathy", and that with limited vocabulary in Mandarin Chinese, this is done through modes of communication other than speech, notably facial expression, eye contact, and "body language" or posture. Whether or not this is actually effective and perceived as "empathy building" by vendors or others they may interact with is a question that is beyond the scope of this study, perhaps falling in the realm of intercultural studies. Regardless, it seems that some of these L2 learners have placed emphasis
on a multimodal communication system as having communicative value that can transcend simply being an exchange of goods and money.

4.1.4 Discussion - Salient features of Interactions

Data from both the interviews and interactions reveals modes of communication other than speech to be intrinsic to communication between L2 learners and proficient speakers as having an *interactional* function, in that they seem to be done intentionally and for the benefit of message transmission to their interlocutor. Vital to successful communication between participants was joint attention towards objects, with this being accomplished through 'environmentally-coupled gestures' using concomitant deictic gesture and speech (e.g., Goodwin 2007). These required participants' hands, eyes, bodies, and vocalizations to be coordinated with their interlocutor's towards the same object, and subsequent interpretive efforts towards the comprehension of the interlocutor's intentions and goals.

As Mondada (2014) has argued, this is not a simple process, and requires a "complex interactional multimodal gestalt" (109). In fact, it is not one of these elements of the multimodal system that seem to be able to create meaning alone, but rather the combination. This seems to be especially apparent for L2 learners in this study. In that they are beginners, and admit through interviews that their vocabulary may be limited, the importance of other modes of communication is perhaps accentuated in order to allow for successful transmission of a given message. In this sense, the interaction as a whole may rely more heavily on the whole of the multimodal system to accomplish mutual understandings between participants. However, it is also clear that this is not always sufficient for communication. In fact, gestural deixis can refer to a number of objects in the immediate environment and is not always completely transparent (Goodwin 2014: 200), similar to McNeill's (2000) statement that, "to point at is a variable thing" (19). In both of the examples presented, it is clear that communication is limited, even when referring to objects within reach of all participants, as clarifications utilizing the whole of the multimodal system are needed for mutual comprehension of all participants.

Yet similarly, when misunderstanding does occur, it is this system which allows for eventual understanding. Thusly, joint attention and recognition of interlocutor message are established through the *progression* of these interactions, again similar to Mondada's (2014) findings of deictic reference and joint attention. Deictic reference requires first that an interlocutor's attention is drawn towards the intended direction. However, this may be misinterpreted or unclear to the interlocutor. Action is then taken by the interlocutor, either as clarifying gesture or speech (as is the case in *Example 1*) or an "agreement token" such as a head nod (Mondada 2014: 101). Yet, this still may be misunderstood and
would need to be clarified more explicitly by either participant with perhaps another mode of communication. Therefore, successful communication requires that action is taken by participants to clarify or correct a message, with these successive actions accumulating towards mutual comprehension. This requires effort from each participant, especially so when linguistic resources may be limited, such as the case with L2 learners in this study. It is therefore the accumulation of communicative actions utilizing multiple modalities, along with the interpretive efforts of participants to understand these actions, which allows for successful communication. Effort to communication, or perhaps what has been called the development of "empathy" by participants in this study, seems to be key to successful communication, as has been suggested by McNeill (2000).

It also seems that gestures, eye gaze, and facial expression often do not completely replace speech, therefore the extent to which these modes of communication can compensate for speech may be limited. In fact, the majority of interactions show speech coupled with other modes of communication. However, in interviews participants claim that gestures are intentionally used to aid a limited vocabulary during interactions, as awareness of gesturing has been argued to be the basis of understanding the degree to which they are used as compensation (Gullberg 2011). Furthermore, instances where gestures are produced without any vocalization, while limited in this study, likely still need to be accounted for. For L2 learners in this study, modalities other than speech seem to be disambiguating and supplementing communication, similar to what has been found in L1 narratives (Cristilli 2014). However, these interactions are not narratives, as have been the focus of the majority of research on gestural compensation, making further comparisons difficult. Furthermore, these are beginning learners. In fact, the combination of pointing and speech seen by participants in this study is more similar to what has been found in L1 acquisition research (e.g., Butcher & Goldin-Meadow 2000; Liszkowski 2010). However, this is not meant to claim that L1 and L2 gestures compensate for speech in the exactly the same ways or for the same purposes, but rather that there is a "gap" in the research. In fact, the use of deictic speech is one deviation from the typical L1 pattern of linguistic development. With this study having limited instances of gestures occurring without speech in interaction, it is difficult to come to conclusions as to their compensatory nature. However, in that they do occur and L2 participants note multiple modes of communication as vital to the successfulness of interactions with proficient speakers, the kind of interaction (i.e. question/answer vs. narrative), the level of the learners, and the relationship of interlocutors seem to be factors which need to be considered in determining the nature of gestural compensation for speech in a given interaction.
Regardless, multimodal communication seems to be indispensable for mutual understanding in these interactions. While deictic gesturing and speech seem to carry much of the information, they are in fact influenced by many elements of the multimodal system working together. To some of the L2 participants, this sometimes transcends a somewhat superficial exchange of money and goods towards the building of "empathy". While modes of communication other than speech may not always compensate for speech to the extent that they are able to completely replace it, they are important enough for mutual comprehension between participants to be studied. Moreover, this mutual comprehension seems to also be the basis from which linguistic development can occur in this environment.

4.2 Linguistic Development and Multimodal Communication in Interaction

4.2.1 Overview

This study has also attempted to determine whether these interactions can be conducive to linguistic development in the L2 learners. Some evidence has been found, while limited, that these interactions can allow for linguistic development and this may be aided by the use of multiple modes of communication, notably hand gesture and eye gaze. However, the extent to which this is possible is dependent on a variety of individual factors, such as learning styles and motivation.

4.2.2 Linguistic development in naturalistic settings

In fact, the common theme connecting the interviews regarding language development is variance. However, the majority of the L2 participants do note that their Chinese language learning outside of the classroom is limited. Eva describes why learning Chinese differs from learning languages more closely related to her native Dutch:

Eva 2:30
(8)

IR: Ok so then what about Chinese? Has it been similar or different to your experiences learning other languages?
Eva: Well the things is with Chinese it's so far, that I think you pick it up... like it's more difficult to just pick it up because it's so like, Spanish is a lot more familiar and especially German for example, it's a lot closer, whereas Chinese is not.
Ana 30:40
(9) Ana: I think most things that I learned is stuff that I kind of knew before and had seen it written down. Um and then I would kind of remember and know it because I would hear it. I actually personally struggle with it if I just hear stuff and I actually don’t know how you write it or what it means.

Rifkin 19:18
(10) Rifkin: I have to learn, I have to… go away work out how something’s written in pinyin and that’s how I remember the word.

Others noted that this learning was limited while the only language learner in an interaction with proficient speakers, but that their language learning in these interactions could be aided by interactions with other more experienced or proficient L2 learners:

Nick 14:35
(11) IR: Ok so anything other than numbers that you can think of specifically? Nick: Oh, to be honest like I learn the most when I travel with other people who speak Chinese. For example, other foreigners, because then you have like, the basic language, which would be English, you can ask, "ok what is this word?" And then you ask them in English the language you understand, and they will give you back the word which you didn’t understand before. So it’s always easier if you have common level of communication first, and then make the step from there.

All of these learners had mentioned being enrolled in a Chinese language course at some point in time, some of them at the time of interview still taking the courses while others were currently not attending any Chinese language course. However, Mike claimed to have never taken a Chinese course. He explained part of his experience learning Chinese when asked if he has learned anything through interactions, such as those with street vendors:

Mike 39:41
(12) Mike: Well one of them would be "liǎng gè". IR: "liǎng gè"? Mike: To start with when I came I learned the numbers, so, everybody, yī, èr is the second one. So I would say I want èr… uh dumplings or whatever. Uh, and I kept getting blank looks, you know? And they’d put two up and point to two and say liǎng, liǎng gè. Uh, ok, the penny drops…
He later explains his contact with the Chinese language:

Mike 40:34
(13) Mike: ...I've learned nothing from a book, and now both of these examples are not cause somebody's told me it either, it's not passed on, it's definitely something that I've picked up through the interaction.

Mike was asked to further explain how these interactions are facilitated and how he believes his learning of linguistic items to occur:

Mike 47:30
(14) Mike: So, but I think really like, when you say like is there one thing? There's a few of those words, there's a few gestures, so let's say the two hands, "yī gè", "liǎng gè", you know. But every time you visit them, you can pick up new things and they might just be very subtle.
IR: Mhm.
Mike: But I would say it's a very iterative process, where every time I go, I would say pretty much what I've already used, you know I will try to use the same words, and slowly you pick up the context where that select language that you have picked up doesn't work, in which case, ok I need to learn a new word.
IR: So in that sense it's trial and error?
Mike: Definitely, and that's not always a conscious trial and error as well.

He continues later, explaining how his feelings on learning Chinese changed over the course of the year and how his linguistic ability developed:

(15) Mike: ...So at the beginning I said, oh you know it doesn't really matter whatever, but slowly over the year I guess I've evolved my understanding piece by piece and built little bits of language around that.
IR: Mhm.
Mike: But again, very organically, not me looking at a book or asking people, very much through interaction.

While these claims of the "iterative process" and the language developing "organically" and "through interaction" are interesting in their own right, how this works and the facilities which allow this iterative process to operate could perhaps be better explained through micro-analysis of these interactions. During the interview, he explained and demonstrated one of his strategies in detail, positioning eye contact, hand gesture, and empathy with the interlocutor as vital to the workings of this process:
IR: So you talk about, um, you know, smiling and things, you briefly mentioned eye contact, what do you do with eye contact, for empathy, in terms of maintaining this idea of empathy?

Mike: Well certainly if you, um, if I say like, if they ask me if I'd like chili, say, and they kind of point to the chili. And I go, I want a little bit. I will go like that and then look up at the-. You know I want a little bit.

Example 3
1) White arrows delineate eye gaze direction, red arrows point towards the gesture, “a little bit”

“I will go like that and, and look up at the-”

(14) (cont.) IR: ... and look them in the eye?
Mike: And look them in the eye as well because, you get obviously, you're going all the way to universal language here almost. Uh, have they registered that? Not necessarily have they understood exactly what I mean, but have they registered my, um, have they registered my gesture? Because if I go like that, little bit, and I know they've registered it, but they give me lots, I know that this gesture doesn't work here in China. If I go, "a little bit", look up, and they're looking down at their feet or the food that they're cooking, I can't make a judgment on whether my gesture has worked or not, because I don't know that they've registered it. So I think to make sure that there's a learning process, you have to make sure that things you do try to say are, you're controlling the variables in some ways, and you're making sure that their attention is on you. You say like one word, one gesture,
did they get it? Did it work? Right, memorize that, we’re using that again tomorrow night for dinner...

(underline is the researcher’s emphasis)

It is clear from this extract that the participant believes gesture, speech, and eye gaze to be inextricably tied within this "learning process", as he states, it is a matter of "controlling the variables". For him, linguistic knowledge is "picked up" through listening in interactions and the meaning and function of this linguistic knowledge is then "tested" against the context with multiple modes of communication facilitating this process. While Mike’s iterative process clearly occurs longitudinally across time, and is therefore slightly beyond the boundaries of this study, an example of how this works can be seen in the video-recording of his interaction with a vendor.

This extract occurs over the course of twenty-seven seconds of the interaction, seen in Example 4. It begins when the vendor asks him, "zhè gè tiáo liào de yào de mā?" ("Which condiments do you want?"), an extended index finger circling above the possible items, gaze following Mike’s next actions.

**Example 4**
1) 1a refers to Mike, 2b refers to Vendor 1
2) White arrows delineate eye gaze direction, red arrows delineate pointing direction, red circles show prominent gesture

Machine’s attention is drawn towards the items, and pointing with an extended index finger at one, utters, "zhè gè", similar to what was seen in the previously outlined examples of concomitant deictic gesture and speech, seen in Annotation 4.1. He raises his hand, index finger held vertically over his
thumb (the "little bit" gesture, similar to what is seen in Example 3), into the vendor’s line-of-sight, uttering, in English, "little bit, little bit", gaze moving from his own gesture towards the vendor. At the same time, the vendor's gaze lifts towards Mike, and pointing towards Mike's gesture, utters, "bù yào" ("You don’t want"), as Annotation 4.2 reveals. She then nods, and Mike continues selecting items with deictic gesture and speech, her gaze following his hand.

Example 4 (cont.)
1) 1a refers to Mike, 2b refers to Vendor 1
2) White arrows delineate eye gaze direction, red arrows delineate pointing direction, red circles show prominent gesture

The interaction continues as other items are prepared, but perhaps because there had been a misunderstanding, the vendor asks Mike about the same item to which he had previously replied, "little bit, little bit". The vendor points towards the item, asking, "zhè gè yào yī diǎn diǎn a?" ("You want a little bit of this?") seen in Annotation 4.3. Mike repeats the same gesture used earlier, index finger held vertically above his thumb, and utters, in English again, "a bit, bit", while both participants attention is focused on Mike's gesture, shown in Annotation 4.4. The vendor utters, "oh hǎo de", ("oh, ok"), and as Mike's gesture finishes, he utters, "yī diǎn" ("a little bit"), seen in Annotation 4.5 (researcher's emphasis in both examples).

What is interesting about this extract is that Mike's native language of English seems to be used first concomitantly with gesture. While it is possible, it does not seem likely that the vendor understands this English phrase, and therefore the communicative "weight" is solely carried by the hand
gesture, eye gaze and joint attention of items and gestures. However, mutual understanding seems to be reached, and vitally, Moe ends this extract with the Mandarin Chinese phrase "yī diǎn", rather than the English translation "a little bit" that he seemed to rely on just seconds before. While it may be a stretch to call this linguistic "learning" (he mentioned, anecdotally, after the interview that it was a phrase he recognized, but did not know the exact meaning of before), it certainly seems to be accumulated through this action, and crucially, with hand gesture and eye gaze as facilitators in the transmission of the message interactionally, with the linguistic item "yī diǎn" heard, internalized, and then produced by the L2 learner.

4.2.3 Discussion - Linguistic development in naturalistic settings

What is mostly obvious regarding the nature of these interactions and language learning (or lack there-of), is that motivations and learning styles ultimately shape how these will occur. In fact, all participants noted somewhat low motivation to learn Chinese, mostly due to time constraints balancing work or school and their plans for the future. Eva echoes the sentiment felt by most in the study:

Eva 2:53
Eva: ...so I took a Chinese class when I got here.
IR: Ok.
Eva: But yeah, then you know I was thinking, I'm only staying here for a year so it's not worth the pain, because when I go back I'm going to forget about it, well it's going to happen quite fast, so um yeah you need a certain basis I guess for Chinese in order to improve outside of the classroom.

However, this should not be surprising. As Ohta's (2010) study revealed, ZPD's are not developmentally equal and are heavily dependent on an individual's own intents and motivations, the internalization of the process of language development. It also seems that most participants found their beginning level of Chinese to be insufficient to learn the language outside of the classroom in interaction. However, this should not detract from the relevance of the study, as low motivation learners should certainly be considered for obvious reasons. Furthermore, in that only one participant seemed to learn language in a "naturalistic" setting, and that this learning was clearly limited, should also not detract from the importance of the fact that language was learned in this way at all. If ZPD's do vary, then it is incumbent upon researchers to discover and describe these different learning styles.

How then was any naturalistic linguistic development able to occur? In terms of input, it is not entirely surprising that, while limited, naturalistic input could allow for linguistic development, especially
when gestures are utilized. Perhaps it is the use of gestures by vendors which allows for the comprehension of such input, similar to the findings of Gullberg et. al. (2010). While this interaction does not show the amount of "accumulation" (i.e. number of repetitions) of the phrase in conjunction with gestures that seem to be needed for learning (16), it is possible that these repetitions took place in a separate, previous interaction. In fact, this seems to mirror Mike's notion of naturalistic learning as an "iterative process", and one that can be accomplished without fully conscious learning efforts (16).

In a similar manner, given the marked use of gesturing, perhaps it is the multimodal system that facilitates this linguistic development, supporting McCafferty's (2002) argument that "the ZPD is embodied" (201). For McCafferty, gesture use promotes language learning beginning with imitation, a key component in Vygotsky's (1978) original notion of the ZPD. In Mike's interaction, we can see that the phrase "yī diǎn", used to express his intended goal, at this point a notion that was mutually understood by both participants mostly through the use of gesture and eye gaze, is uttered by the L2 participant only after the proficient speaker has used the phrase. However, in McCafferty's study imitation was of synchronous gesture and speech, while here the linguistic item is uttered after the gesture. Yet, other gestures are imitated in all of these interactions by both L2 and proficient speakers, notably deixis. Perhaps these gestures help promote the "interpersonal dynamic", and therefore interaction with the ZPD (199). When participants are pointing towards items in the surrounding space, they are engaging in a complex behavior seeking mutual understanding through "interactional synchrony" (200). It is through this activity that the ZPD could account for linguistic development, and could perhaps exist as Mike's concept of "empathy" with interlocutors. However, this data also seems to position eye gaze as important in the recognition of messages, with visual attention marking salient features of the interaction, a similar finding to Streeck's (2014) study. This "interpersonal dynamic" is then aided by more than just hand gestures. Moreover still, it seems that what the L2 participant calls "empathy" between participants is developed through modes of communication other than speech, including facial expression, eye gaze, hand gesture, and posture, and that this is necessary for the development of linguistic knowledge as an "iterative" and "interactive" process. Perhaps it is not just hand gesture that can provide scaffolding in naturalistic settings for L2 learners, but rather the whole of the multimodal system, and in some cases, this is utilized intentionally by L2 learners.

However, it is clear that if gestures do interact in the ZPD, in this study linguistic knowledge gains are still quite limited. As Mike has admitted:
Mike 49:45
(18) Mike: I'd say, but here we are at the end of the year, my Chinese is still very limited, so this is no rulebook for how to learn. You know if you want to learn a language pick up your books and memorize, study. Uh obviously pronunciation and fluency is practice, and get out there and do it, but there's no substitute for just knowing the content.

Therefore, while gestures are used quite frequently in these interactions to communicate goals and intentions, they are limited in aiding linguistic development for these L2 learners. However in the instances where it does occur, it seems that it is the motivation to "empathize" with proficient speakers which encourages this linguistic development through multiple modes of communication other than speech alone.

Chapter 5 - Conclusions and Future Directions

This study has shown that gestures, eye gaze, and other features of the multimodal communication system are utilized frequently in L2 learner interactions with proficient speakers, and in some cases, this is done intentionally and for communicative purposes. The coordinated efforts of participants trace the complex processes of interlocutor actions utilizing the whole of the multimodal system towards a mutual understanding between participants. While these modes of communication are mostly used in conjunction with speech, they seem to be used to supplement or disambiguate the message, and as admitted by participants in interviews, are used in cases where their vocabulary may be limited. While a few instances may reveal the multimodal system to compensate for speech, it is clear that the kind of interaction, the level of the L2 learners, and the relationship of interlocutors must be taken into account when determining the degree of gestural compensation in a given interaction. However, this study of the multimodal system does have limitations. Eye gaze either towards or away from an interlocutor could mostly be easily determined, but given the camera angle needed to capture the focal event, the exact location of gaze often could not. Whether a participant's gaze was directed towards their interlocutor's body, their eyes, their hands, or towards the object being referenced to with gesture may have an impact on the meaning of the gaze. In that at least one participant has noted the importance of "eye contact" in terms of message recognition, this could be a productive avenue for research in the future. Moreover, the camera angle also limited facial expressions that could be studied. It is hoped that future studies will be able to better account for these factors, perhaps using multiple cameras, as gaze and facial expression seems to undoubtedly have an impact on the interaction.
Furthermore, this study has suggested that some linguistic development can occur in beginning L2 learners in naturalistic settings and that this development may be facilitated by the use of gesture and multimodal communication. While there is precedence for linguistic development to occur with limited naturalistic input through accumulation of repetitions (e.g., Gullberg et. al. 2010), for these L2 learners, linguistic development seems somewhat limited. Furthermore, the amount that this learning can be attributed to the notion of the ZPD is limited, though if it does occur, it seems that it is the whole of the multimodal system which operates within the ZPD, not hand gesture alone. Moreover still, linguistic knowledge gains in this naturalistic environment seem to be inextricably tied to individual learner learning styles and motivations. In some cases, it seems that what some L2 learners describe as "empathy" with their interlocutors is vital to the language development process and that this is accomplished through modes of communication other than speech. Again however, the conclusions that could be reached regarding linguistic knowledge gains are still quite limited. Given the allotted time to complete the project, linguistic development over longer periods of time could not be accounted for. If, as some participants claim, naturalistic language learning is an "iterative process", long term gains would likely need to be studied. Longitudinal studies could also make comparisons of gesture use of language learners in naturalistic environments to those in the classroom, an aspect which has not been explored in this study. Furthermore, only five L2 participants were involved in this study, meaning sample size was quite limited and relied solely on qualitative analysis. Larger longitudinal studies which can analyze more participants could perhaps shed light on the naturalistic learning process by allowing for quantitative data analysis alongside a more representative qualitative sample. Furthermore, such an approach could better utilize variationist models which could better predict and explain the degrees to which this naturalistic learning is implicit or explicit (e.g., Tarone 2010), again an aspect which was beyond the scope of this study.

Moreover still, the role of culture was not taken into account in this study (aside from language) in that it was thought this would allow for a better focus on the L2 learner and the processes which may or may not facilitate linguistic development. In examining interactions which are quite obviously molded by the cultural context (e.g., Hymes 1962, 1974; Goodwin & Duranti 1997), a better understanding of the role of culture in such interactions would change the conclusions that could be drawn. A study which utilizes interviews with proficient speaking participants is an obvious starting point.

Despite these limitations, this study is perhaps best viewed as exploratory, likely raising more questions than providing definitive answers. However, in studying naturalistic language learning (as it
has been defined here), there are a plethora of factors which need to be taken into account to provide a holistic picture of the event, and likely many which have not been mentioned in this work. However, the use of interviews alongside micro-analyses of interactions were vital in understanding some of these factors and processes by allowing for 'triangulation' of data. The Chinese street market is an exciting context for research and it is hoped that future studies can better examine multimodal features of L2 interactions, provide larger and more comprehensive data sets to better understand L2 learning processes, and consider the important role of culture in interaction within this vibrant natural setting.
References


Gullberg, M. (2011) 'Thinking, Speaking and Gesturing about Motion in more than One'. Thinking and speaking in two languages, 77, pp. 143-169.


Mondada, L. (2007a) 'Multimodal resources for turn-taking pointing and the emergence of possible next speaker's'. *Discourse studies*, 9, pp. 194-225.


Appendices

Appendix 1

Participant Information Sheet: Multimodal Communication, Compensation, and Context: Chinese language learner communication in the marketplace

Dear Participant,

Thank you for agreeing to participate in this research in connection with my Master's dissertation coursework at the University of Nottingham Ningbo. The project is a study of individuals who are learning Chinese. We aim to find out how beginning Chinese speakers use language to interact with Chinese vendors and shopkeepers.

Your participation in the survey is voluntary. You are able to withdraw from the survey at any time and to request that the information you have provided is not used in the project. Any information provided will be confidential.

For this study, we will use both audio and video-recording equipment to collect your language use. All of this information will be kept on the researcher's computer (Jake Wild) and an external memory drive. Those scoring the project and UNNC tutors will have access to this information.

Furthermore, this research may be used at future conferences or publications where still images or examples of expressions you use may be seen by other researchers, though data will not be shared. This research may be used in academic contexts as examples, such as in classroom courses. In all of these cases, you may be identifiable by your image and voice, but no personal information would be shared.

Participation in this research is completely voluntary. You will be given a consent form to sign where you can decide to what extent we may use your image and voice. You are able to withdraw from the research at any time and request all of your information to be removed from the project. Any personal information provided will be confidential. Your identity will not be disclosed in any use of the information you provide. Once the project is finished, all information will be kept by the researcher (Jake Wild) and access will be granted only to those scoring or UNNC tutors assisting with the project (Simon Harrison; Yu-Hua Chen - UNNC).

The research project has been reviewed according to the ethical review processes in place in the University of Nottingham Ningbo. These processes are governed by the University's Code of Research Conduct and Research Ethics. Should you have any question now or in the future, please contact me or my supervisor. Should you have concerns related to my conduct of the survey or research ethics, please contact my supervisor or the University’s Ethics Committee.

Yours truly,

Jake Wild

Contact details:
Student Researcher: Jake Wild, zx17278@nottingham.edu.cn
Supervisor: Simon Harrison, Simon.Harrison@nottingham.edu.cn
University Research Ethics Committee Coordinator, Ms Joanna Huang
(Joanna.Huang@nottingham.edu.cn)
PARTICIPANT CONSENT FORM

Project title: Multimodal Communication, Compensation, and Context: Chinese language learner communication in the market place

Researcher's name: Jake Wild

Supervisor's name: Simon Harrison

- I have read the Participant Information Sheet and the nature and purpose of the research project has been explained to me. I understand and agree to take part.
- I understand the purpose of the research project and my involvement in it.
- I understand that I may withdraw from the research project at any stage and that this will not affect my status now or in the future.
- I understand that personal information will be confidential.
- I understand that I will be audio and video-recorded, and that these recordings will be seen by the researcher (Jake Wild) and those scoring and assisting with the project (Simon Harrison; Yu-Hua Chen - UNNC).
- I understand that recordings in which I may be identifiable could later be used by the researcher in various academic contexts, such as conference presentations, academic publications, and university lectures where this research is described. In that case:
  - I agree to full use of video recordings of me [ ] Yes [ ] No
  - I agree to full use of audio recordings of me [ ]
- If you answered no, please select the extent of use below:
  - Please blur out my face (so that the face is not recognizable). [ ]
  - Please distort my voice (so that my voice is not recognizable). [ ]
  - Please do not use any of the audio or video recording other than for private analysis on this research project. [ ]
- I understand that data will be stored in accordance with data protection laws.
- I understand that I may contact the researcher or supervisor if I require more information about the research, and that I may contact the Research Ethics Sub-Committee of the University of Nottingham, Ningbo if I wish to make a complaint related to my involvement in the research.

Signed: [Signature] (participant)

Print name: Benjamin Schmoll Date: 03/06/2015

Contact details
Student Researcher: Jake Wild, zx17278@nottingham.edu.cn
Supervisor: Simon Harrison, Simon.Harrison@nottingham.edu.cn
UNNC Research Ethics Sub-Committee Coordinator: Ms. Joanna Huang
Joanna.Huang@nottingham.edu.cn
PARTICIPANT CONSENT FORM

Project title  Multimodal Communication, Compensation, and Context: Chinese language learner communication the market place

Researcher's name  Jake Wild

Supervisor's name  Simon Harrison

- I have read the Participant Information Sheet and the nature and purpose of the research project has been explained to me. I understand and agree to take part.
- I understand the purpose of the research project and my involvement in it.
- I understand that I may withdraw from the research project at any stage and that this will not affect my status now or in the future.
- I understand that personal information will be confidential.
- I understand that I will be audio and video-recorded, and that these recording will be seen by the researcher (Jake Wild) and those scoring and assisting with the project (Simon Harrison; Yu-Hua Chen - UNNC).
- I understand that recordings in which I may be identifiable could later be used by the researcher in various academic contexts, such as conference presentations, academic publications, and university lectures where this research is described. In that case:
  - I agree to full use of video recordings of me  Yes  No
  - I agree to full use of audio recordings of me
- If you answered no, please select the extent of use below:
  - Please blur out my face (so that the face is not recognizable).
  - Please distort my voice (so that my voice is not recognizable).
  - Please do not use any of the audio or video recording other than for private analysis on this research project.
- I understand that data will be stored in accordance with data protection laws.
- I understand that I may contact the researcher or supervisor if I require more information about the research, and that I may contact the Research Ethics Sub-Committee of the University of Nottingham, Ningbo if I wish to make a complaint related to my involvement in the research.

Signed  (participant)

Print name  Thomas Wildson  Date  29/5/15

Contact details
Student Researcher: Jake Wild, zx17278@nottingham.edu.cn
Supervisor: Simon Harrison, Simon.Harrison@nottingham.edu.cn
UNNC Research Ethics Sub-Committee Coordinator: Ms. Joanna Huang
Joanna.Huang@nottingham.edu.cn
PARTICIPANT CONSENT FORM

Project title: Multimodal Communication, Compensation, and Context: Chinese language learner communication the market place

Researcher's name: Jake Wild

Supervisor's name: Simon Harrison

- I have read the Participant Information Sheet and the nature and purpose of the research project has been explained to me. I understand and agree to take part.

- I understand the purpose of the research project and my involvement in it.

- I understand that I may withdraw from the research project at any stage and that this will not affect my status now or in the future.

- I understand that personal information will be confidential.

- I understand that I will be audio and video-recorded, and that these recording will be seen by the researcher (Jake Wild) and those scoring and assisting with the project (Simon Harrison; Yu-Hua Chen - UNNC).

- I understand that recordings in which I may be identifiable could later be used by the researcher in various academic contexts, such as conference presentations, academic publications, and university lectures where this research is described. In that case:
  - I agree to full use of video recordings of me
  - I agree to full use of audio recordings of me

- If you answered no, please select the extent of use below:
  - Please blur out my face (so that the face is not recognizable).
  - Please distort my voice (so that my voice is not recognizable).
  - Please do not use any of the audio or video recording other than for private analysis on this research project.

- I understand that data will be stored in accordance with data protection laws.

- I understand that I may contact the researcher or supervisor if I require more information about the research, and that I may contact the Research Ethics Sub-Committee of the University of Nottingham, Ningbo if I wish to make a complaint related to my involvement in the research.

Signed: ____________________________ (participant)

Print name: Elaine Maierand Date: 11-06-2015

Contact details
Student Researcher: Jake Wild, zx17278@nottingham.edu.cn
Supervisor: Simon Harrison, Simon.Harrison@nottingham.edu.cn
UNNC Research Ethics Sub-Committee Coordinator: Ms. Joanna Huang
Joanna.Huang@nottingham.edu.cn
Appendix 5

PARTICIPANT CONSENT FORM

Project title Multimodal Communication, Compensation, and Context: Chinese language learner communication the market place

Researcher's name Jake Wild

Supervisor's name Simon Harrison

- I have read the Participant Information Sheet and the nature and purpose of the research project has been explained to me. I understand and agree to take part.
- I understand the purpose of the research project and my involvement in it.
- I understand that I may withdraw from the research project at any stage and that this will not affect my status now or in the future.
- I understand that personal information will be confidential.
- I understand that I will be audio and video-recorded, and that these recordings will be seen by the researcher (Jake Wild) and those scoring and assisting with the project (Simon Harrison; Yu-Hua Chen - UNNC).
- I understand that recordings in which I may be identifiable could later be used by the researcher in various academic contexts, such as conference presentations, academic publications, and university lectures where this research is described. In that case:
  - I agree to full use of video recordings of me
  - I agree to full use of audio recordings of me

If you answered no, please select the extent of use below:
- Please blur out my face (so that the face is not recognizable).
- Please distort my voice (so that my voice is not recognizable).
- Please do not use any of the audio or video recording other than for private analysis on this research project.

- I understand that data will be stored in accordance with data protection laws.
- I understand that I may contact the researcher or supervisor if I require more information about the research, and that I may contact the Research Ethics Sub-Committee of the University of Nottingham, Ningbo if I wish to make a complaint related to my involvement in the research.

Signed ........................................... (participant)

Print name .................................. Date ....................................

Contact details

Student Researcher: Jake Wild, zx17278@nottingham.edu.cn
Supervisor: Simon Harrison, Simon.Harrison@nottingham.edu.cn
UNNC Research Ethics Sub-Committee Coordinator: Ms. Joanna Huang

Joanna.Huang@nottingham.edu.cn
APPENDIX 6

PARTICIPANT CONSENT FORM

Project title Multimodal Communication, Compensation, and Context: Chinese
language learner communication in the marketplace

Researcher's name Jake Wild

Supervisor's name Simon Harrison

- I have read the Participant Information Sheet and the nature and purpose of the
research project has been explained to me. I understand and agree to take part.
- I understand the purpose of the research project and my involvement in it.
- I understand that I may withdraw from the research project at any stage and that this
will not affect my status now or in the future.
- I understand that personal information will be confidential.
- I understand that I will be audio and video-recorded, and that these recordings will be
seen by the researcher (Jake Wild) and those scoring and assisting with the project
(Simon Harrison; Yu-Hsin Chen - UNNC).
- I understand that recordings in which I may be identifiable could later be used by the
researcher in various academic contexts, such as conference presentations, academic
publications, and university lectures where this research is described. In that case:
  - I agree to full use of video recordings of me
  - I agree to full use of audio recordings of me

- If you answered no, please select the extent of use below:
  - Please blur out my face (so that the face is not recognizable).
  - Please distort my voice (so that my voice is not recognizable).
  - Please do not use any of the audio or video recording other
then for private analysis on this research project.
- I understand that data will be stored in accordance with data protection laws.
- I understand that I may contact the researcher or supervisor if I require more
information about the research, and that I may contact the Research Ethics Sub-
Committee of the University of Nottingham, Ningbo if I wish to make a complaint
related to my involvement in the research.

Signed .................................................. (participant)

Print name ........................................... Date 07/11/2014

Contact details
Student Researcher: Jake Wild, zx17278@nottingham.edu.cn
Supervisor: Simon Harrison, Simon.Harrison@nottingham.edu.cn
UNNC Research Ethics Sub-Committee Coordinator: Ms. Joanna Huang
Joanna.Huang@nottingham.edu.cn

Page 67 of 69
Translator Verification Form

Researcher's Name: Jacob Wild

- I understand the intent of this study and received verbal consent for audio- and visual recordings from vendors present in the research.

- I understand that I may contact the researcher or supervisor at UNNC for more information, and that I may contact the Research Ethics Sub-Committee of the University of Nottingham, Ningbo if I wish to make a complaint related to my involvement in the research.

I, ____________________________ am a fluent speaker of Mandarin Chinese and assisted the researcher in data collection for this dissertation project by translating between him and vendors present in the study.

Contact: 186 6742 7371

Signed ____________________________ (translator)

Contact details
Student Researcher: Jake Wild, zx17278@nottingham.edu.cn
Supervisor: Simon Harrison, Simon.Harrison@nottingham.edu.cn
UNNC Research Ethics Sub-Committee Coordinator: Ms. Joanna Huang
Joanna.Huang@nottingham.edu.cn
Appendix 8 (see accompanying USB flash drives)

Folder 1 - Clips and interview audio (dissertation relevant data)

**Video Samples**

1. *Example 1* (Rifkin street market interaction)
2. *Example 2* (Ana street market interaction)
3. *Example 3* (Mike interview example)
4. *Example 4* (Mike street market interaction)

**Audio Samples**

5. Rifkin interview
   Interview examples (1) - 8:40
   (10) 19:18

6. Ana interview
   Interview examples (2) - 34:45
   (4) - 18:30
   (5) - 25:20
   (9) - 30:40

7. Nick interview
   Interview examples (3) - 8:20
   (11) - 14:35

8. Eva interview
   Interview examples (8) - 2:30
   (17) - 2:53

9. Mike interview
   Interview examples (6) - 9:40
   (7) - 51:20
   (12) - 39:41
   (13) - 40:34
   (14) - 47:30
   (15) - 47:50
   (16) - 51:58
   (18) - 49:45

Folder 2 - Full data sets (for use in demonstrations and instruction)

Full street market interaction videos/audio

*Please contact Jacob Wild at jlwild2@hotmail.com for interview videos