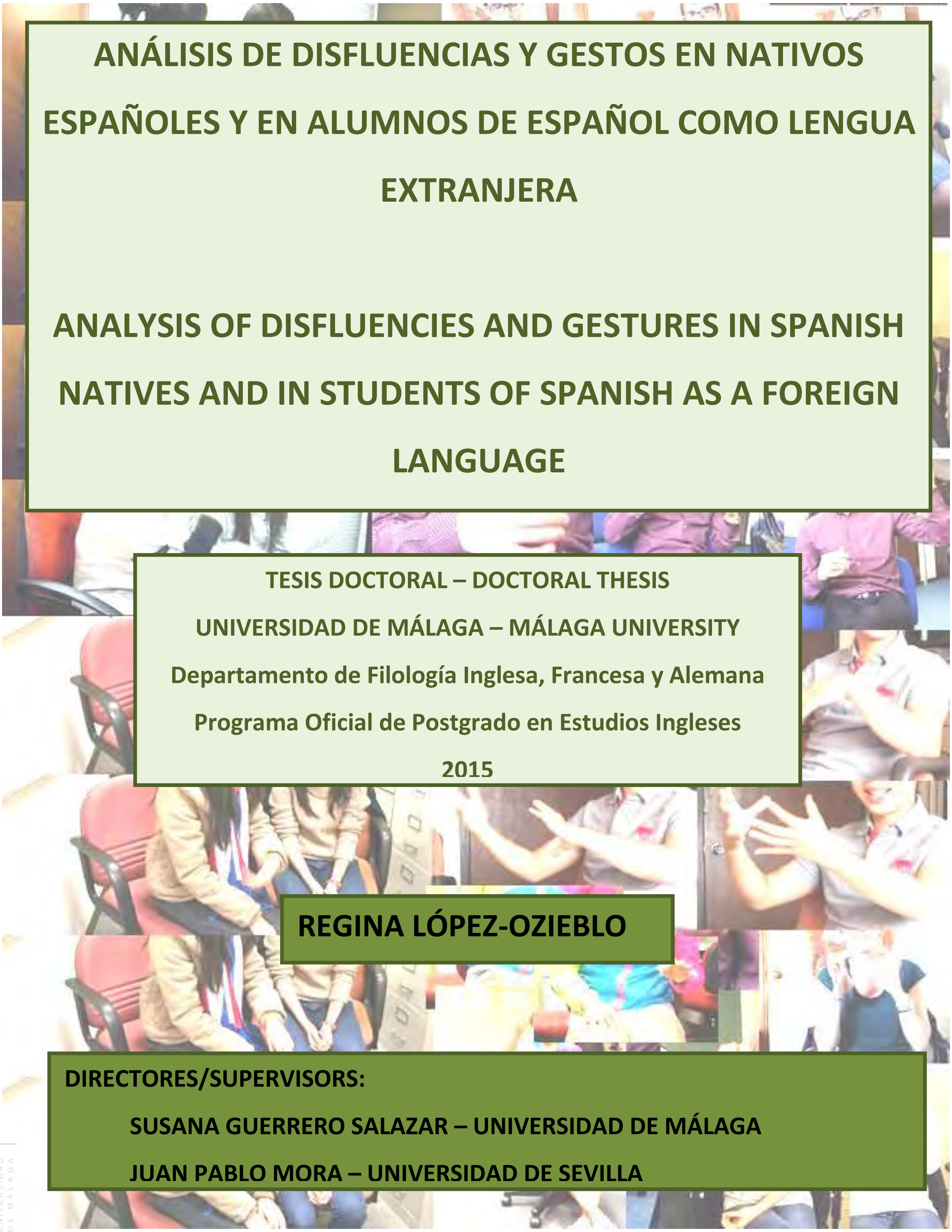
UNIVERSIDAD DE MÁLAGA



ANÁLISIS DE DISFLUENCIAS Y GESTOS EN NATIVOS ESPAÑOLES Y EN ALUMNOS DE ESPAÑOL COMO LENGUA EXTRANJERA

ANALYSIS OF DISFLUENCIES AND GESTURES IN SPANISH NATIVES AND IN STUDENTS OF SPANISH AS A FOREIGN LANGUAGE

TESIS DOCTORAL – DOCTORAL THESIS

UNIVERSIDAD DE MÁLAGA – MÁLAGA UNIVERSITY

Departamento de Filología Inglesa, Francesa y Alemana

Programa Oficial de Postgrado en Estudios Ingleses

2015

REGINA LÓPEZ-OZIEBLO

DIRECTORES/SUPERVISORS:

SUSANA GUERRERO SALAZAR – UNIVERSIDAD DE MÁLAGA

JUAN PABLO MORA – UNIVERSIDAD DE SEVILLA

Publicaciones y
Divulgación Científica
UNIVERSIDAD
DE MÁLAGA

AUTOR: Regina López Ozieblo
EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

ANÁLISIS DE DISFLUENCIAS Y GESTOS EN NATIVOS ESPAÑOLES Y EN ALUMNOS DE ESPAÑOL COMO LENGUA EXTRANJERA

ANALYSIS OF DISFLUENCIES AND GESTURES IN SPANISH NATIVES AND IN STUDENTS OF SPANISH AS A FOREIGN LANGUAGE

TESIS DOCTORAL – DOCTORAL THESIS

UNIVERSIDAD DE MÁLAGA – MÁLAGA UNIVERSITY

Departamento de Filología Inglesa, Francesa y Alemana

Programa Oficial de Postgrado en Estudios Ingleses

2015

REGINA LÓPEZ-OZIEBLO

DIRECTORES/SUPERVISORS:

SUSANA GUERRERO SALAZAR – UNIVERSIDAD DE MÁLAGA

JUAN PABLO MORA – UNIVERSIDAD DE SEVILLA

UNIVERSIDAD DE MÁLAGA

D.ª Susana Guerrero Salazar, profesora titular del departamento de Filología Española II y Teoría de la Literatura de la Universidad de Málaga, y D. Juan Pablo Mora Gutiérrez, profesor titular del departamento de Lengua Española, Lingüística y Teoría de la Literatura de la Universidad de Sevilla, como codirectores de la tesis doctoral de D.ª Regina López Ozieblo, denominada *Análisis de disfluencias y gestos en nativos españoles y en alumnos de español como lengua extranjera,* informamos de que la presente tesis doctoral cumple con los requisitos formales de calidad y originalidad y mantiene el rigor científico y académico exigible, por lo que autorizamos su presentación y defensa.

UNIVERSIDAD DE MÁLAGA
Facultad de Filosofía y Letras
Departamento de Filología Española II y Teoría de la Literatura

Málaga, 15 de enero de 2015

FACULTAD DE FILOLOGÍA
Departamento de Lengua Española, Lingüística y Teoría de la Literatura.

JPMora

Susana Guerrero Salazar

Juan Pablo Mora Gutiérrez

Quisiera agradecer a mis directores de tesis, la Dr.ª Susana Guerrero Salazar y el Dr. Juan Pablo Mora Gutiérrez, por sus consejos y ánimos. A mis compañeros de la Universidad Politécnica de Hong Kong por su ayuda, especialmente la Dr.ª Li Lan, y el Dr. Victor Ho. Así como a todos aquellos que me han apoyado, incluyendo la Dr.ª Lidia Taillefer de Haya, el Prof. David McNeill y la Prof.ª Bárbara Ozieblo Rajkowska.

CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1 BACKGROUND

The seed for this study was planted when observing the difficulties that Hong Kong students of Spanish as a foreign language had when talking to Spanish natives through *Skype*,[1] in a formal classroom context. The *Skype* exercise was created to give the Hong Kong students the chance to converse with native Spanish speakers. In Hong Kong the possibilities of speaking Spanish outside the classroom context are very limited and not all of the students can afford to travel to Spanish speaking countries. Informal observations of the *Skype* conversations showed that aside from the inherent technical difficulties, shyness and low level of language skills, there was something else at play interfering with fluency. Our belief is that the main factor was the lack of familiarity with non-verbal communication (NVC) (López-Ozieblo, 2013a) and so we[2] looked into the integration of certain NVC cues that could impact the flow of the conversations. In particular the study looked at the dynamics of turn taking and the NVC cues, or lack of them, in the dialogues of these students. One of these cues, relating to the hand gesture when the turn is being given, led to an interest in the study of hand gestures, their classification and importance within speech. There have been a number of gesture-speech studies done with students of foreign languages, but they often provide contradictory information. To gather more information on the subject we chose to focus this study on another common event during speech, that of disfluencies (less than fluent speech) – often leading to self-repair – and gestures. As far as we have been able to ascertain only a handful of studies have covered these disfluencies and gestures. Gestures is a field of growing interest, in and out of the foreign language classroom. And, as we will explain below, we believe it deserves much more attention than it has received up to now.

[1] A software programme that allows for synchronous voice and video communication over the internet.

[2] The plural 'we' will be used throughout this dissertation, as is received practice in publications.

2 PREAMBLE

Spontaneous speech will usually have disfluencies like cut-offs or interruptions, pauses, utterances like 'uh', 'um' (fillers), elongations; followed sometimes by replacements, clarifications or repetitions (Clark, 1996). These might follow obvious speech errors, but not always. Everybody, communicating through a language, be it verbal or sign language, makes errors and disfluencies, although most of the time addressees do not perceive them. The estimate is that in spontaneous speech disfluencies will occur in 6 out of every 100 words (Corley & Hartsuiker, 2011, p. 1).

It is thought that these disfluencies are the result of speakers' constant monitoring of their speech for appropriateness and phonological, lexical or grammatical errors. On identifying such an error, sometimes before it has been externalised, speakers might stop, produce a disfluency and reformulate the utterance through repetitions or other replacements for what was said originally. Thus the disfluency might be an indication of an inner error (covert). Some inner errors will be corrected before they are externalized and thus leave no external evidence at all. Another possibility exists to explain the production of disfluencies: it could be that the disfluency is not related to an error but to a disruption to the planning mechanism, from a missing word to an external disruption. A speaker will also produce audible errors that are not corrected, either because they have not been noticed or because on weighing correction versus fluidity the speaker chooses the latter. In all of these cases a monitoring process must be checking and sometimes fixing speech (Levelt, 1983, p. 45; Postma & Oomen, 2005, p. 160) but it seems that this monitoring mechanism affects more than just the speech: gesture, hand movements co-occurring with speech,[3] are also affected, to the extent that gesture is put on hold together with speech (Seyfeddinipur, Kita & Indefrey, 2008).

The study of overt (or obvious) speech errors led to a proposal of how speech is processed (Fromkin, 1971). Disfluencies, on the other hand, have not been covered by research to the same extent as overt errors, and their production is still to be fully understood. With the current study we present a closer look at disfluencies through the

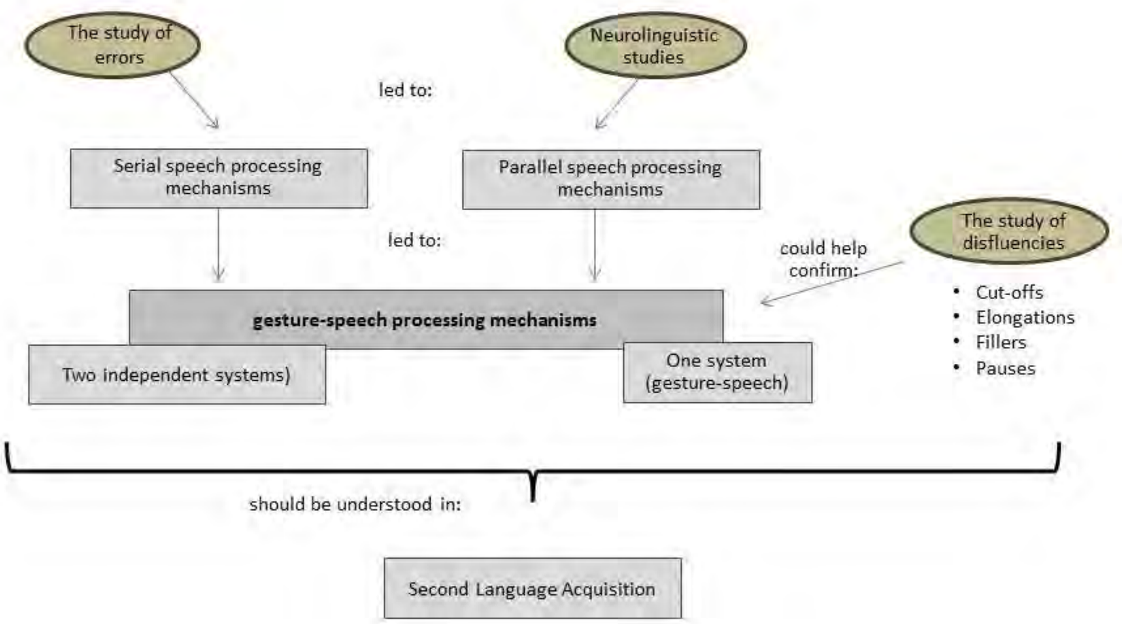
[3] This is a very simplistic definition of gestures which will be elaborated on in this study.

performance of gestures. Gesture is a growing field within psycholinguistics, led by an increasing number of scholars who defend the linguistic nature of gestures (de Ruiter, 2000; Duncan, 2006; Gullberg, 2006; Kita, 2000; Kendon, 2004, McNeill, 2012; Seyfeddinipur, 2006). We believe that understanding gesture production could help understand disfluencies.

This study took place in a foreign language teaching context. Participants were learning either English or Spanish as a foreign language. Therefore the focus of the research became to understand disfluency and gesture in both the mother tongue and the second language of these foreign language students, for the purposes of instruction. However, the hope was that our results would support the gesture-speech processing model that advocates the unity of these two modalities (gesture and speech), the Growth Point theory. This in turn would be strong evidence that speech processing models are more likely to follow a parallel framework (most linguistic studies are based on a serial processing framework).

This dissertation is divided into six main chapters starting with the objectives (Chapter three). The following chapter (Chapter four) is dedicated to a review of the literature. As it is likely that some readers will not be familiar with some of the fields we cover, this chapter is divided into three sections summarizing the most important findings in the fields of speech processing and disfluencies, gesture, and second language acquisition (in particular in relation to gestures). It also provides a brief summary of the development of the two speech processing models that are the basis for most speech and gesture studies. Figure 1 summarises the flow of the content in the *Literature Review,* Chapter four.

Figure 1. – Literature Review: Flow of content.



The fifth chapter details the research study, with sections on the methodology and description of the data and research tools. The sixth chapter details the results of the preliminary study (the original test case) and the main study. Chapter seven integrates the knowledge gathered from the literature review together with the results of the study in an in-depth discussion which is followed by chapter eight with recommendations for future research. Chapter nine summarises the main points in the conclusion. Reference material is provided at the end in the form of a bibliography and a series of appendices that include the transcription of the corpus and samples of the analysis.

3 OBJECTIVES

The main objective of this study was to find a relationship between gesture and speech during disfluency within a context of second language acquisition. However as there are only a few studies combining both disfluency and gesture, a number of secondary objectives were established in order to obtain a better understanding of these two communicative actions and their processing mechanism.

3.1 General Objective

The main objective of this work is to confirm or reject the hypothesis that there is a relationship between speech and gesture in disfluency. The hypothesis is that H_1: $r \neq 0$. Therefore, there is a relationship between speech and gesture during disfluency.

3.2 Secondary Objective

A more pragmatic secondary objective of this study is to analyse the use of verbal and non-verbal resources used in disfluencies by native Spanish speakers. These will be compared to the resources used by students of Spanish as a foreign language in Hong Kong (from now on referred to as HK students) in order to identify traits that could help students become more native in their gesture-speech production. The ultimate goal is to adapt and improve the curriculum for the program of Spanish as a foreign language currently taught at the Polytechnic University of Hong Kong, which the author coordinates.

3.3 Specific Objectives

In order to achieve the goals set out above they have been broken down into a series of specific objectives as follows:

1. Analyse disfluencies and gestures of HK students of Spanish as a foreign language and native Spanish speakers, speaking in English and Spanish.
2. Compare the type and number of disfluencies by language.
3. Compare the type and number of gesture by language.

4. Verify whether the use of disfluency is the same in English and Spanish (when used by native speakers). The hypothesis is that H_2: $r \neq 0$: there is no significant difference in the disfluencies between the two languages as spoken by the same speaker.
5. Verify whether Second Language (L2) students are more likely to gesture with a disfluency when using the L2 than when speaking in their mother tongue. The hypothesis is that H_3: $r \neq 0$: there is an inverse relationship between the level of proficiency and the number of gestures made with disfluency.
6. Verify if the gesture is synchronous with the disfluency. The hypothesis is that H_4: $r \neq 0$: the disfluency is synchronous with the gesture, with both speech and gesture put on hold during disfluency.
7. If any differences are observed between English and Spanish disfluencies (when used by native speakers), identify those that could be taught to foreign students of that language.
8. Propose to the educational institution additional activities for the Spanish as a foreign language programme that will increase student awareness of Spanish native speakers' use of disfluencies and gestures.
9. Provide support for the importance of gestures in linguistic studies.
10. Lay the foundations for future disfluency-gesture research studies.

4 LITERATURE REVIEW

The study of errors has never been as notorious as during Freud's time; for he made slips of the tongue – errors – a window into the subconscious. But this was a study for psychologists, not for linguists. Although errors have previously captured the attention of linguists like Hermann Paul (1880s) it is Meringer who is considered the father of slips of the tongue.[4] He disputed Freud's conclusions arguing that most slips have a simple phonological explanation.[5] In this category are included word substitutions (malapropisms) and reversals, consonant or vowel reversal, addition, substitution or deletion, stress errors, as well as ungrammatical utterances.[6] Contemporary neurolinguistic research concedes that Freud might have been right after all and that the subconscious impacts and interferes with speech (Dell & Oppenheim, 2012). Whether that is the case or not, the attention generated by Freudian slips in the early 20th century led to methodological collection of slips of the tongue, as well as their study and categorization. The first key conclusions came from Fromkin (1971) who, using an analysis of errors, suggested the principles of how speech could be produced. This was the basis for subsequent models that all proposed a series of modules leading from concept, to formulation to articulation of the utterance.

As well as Freud's slips of the tongue, speakers often produce sounds such as 'uh' and 'um' (fillers) and pauses, even with no errors preceding them. The study of these, although not as amusing as recording slips of the tongue, has also led to some important conclusions, the most relevant being that addressees might be paying more attention to the conversation when fillers are produced and that the speaker might be choosing among them, just like they choose a word, to signal a longer or shorter delay (Clark & Fox Tree, 2002, p. 107). Fillers are not a sign of anxiety, although other disfluencies are (Mahl, 1987) and can be eliminated with practice, suggesting they do fall under the control of the speaker, and they are part of the speech production mechanism.

[4] Cited by Fromkin, 1973, p. 13.

[5] For a non-academic description of the dispute see Erard, 2007.

[6] For a comprehensive list see Fromkin, 1971.

After a filler or a pause speakers will resume their speech either with or without a repair. The study of repairs has explored the different types and their nuances, such as the grammatical relationships between repairs and original utterances. In addition, the time lapse between the original utterance and the repair has also been analysed extensively concluding that repairs can be almost immediate, which suggests that speech production and monitoring are concurrent.

Much more might be known about errors and repairs had disfluencies been a bigger part of linguistic studies; but when Chomsky stated that “linguistic theory is concerned primarily with an ideal speaker-listener […] unaffected by such grammatical irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors” (1965, p. 3), disfluencies became the topic of psycholinguists and medicine (in the study of aphasia). The general consensus today is that disfluencies are part of the speech act and ought to be examined as part of the discourse. Although the term ‘error’ is still in use, it refers to inner or overt trouble, a deviation from an ideal utterance, at a cognitive, lexical retrieval or phonetic level (among others) that exists in the speaker’s mind. It is recognized that with disfluencies we are not necessarily dealing with errors and corrections as such, although many studies in this area still use the term ‘error’ (Hartsuiker & Kolk, 2001; Hartsuiker, Roelin, Postma & Wijnen, 2005) or ‘correction’ – suggesting errors. Some scholars have replaced the term ‘correction’ for ‘repair’ (Schegloff, Jefferson & Sacks, 1977), ‘performance additions’ (Clark & Fox Tree, 2002, p. 74) or ‘replacement’ (Clark, 1996, p. 262). Events such as pauses, fillers, cut-offs and elongations are termed ‘disruptions’ (Clark, 1996, p. 258), or just ‘disfluencies’ (Corley & Hartsuiker, 2011, p. 1). For convenience, this study will use the term disfluency[7] for all speech that is not fluent, including long pauses and repairs, and ‘repair’ for the continuation of the utterance after a disfluency.

Another point to note is that, based on the work of Schegloff, Jefferson and Sacks (1977), the literature distinguishes between ‘self-repairs’, those initiated by speakers upon

[7] In American English: dysfluency

disrupting their utterance and 'others' repairs', initiated by addressees.[8] This study focuses on disfluencies and repairs uttered just by the speaker without input from the addressee.

Now that errors and disfluencies are accepted as part of linguistic studies the next challenge is the inclusion of gesture as well. Our belief, following the views of many before us (de Ruiter, 2000; Duncan, 2006; Gullberg, 2008; Kita, 2000; Kendon, 2004, McNeill, 2012; Seyfeddinipur, 2006), is that certain gestures are linguistic and the speech act would be more accurately termed a speech-gesture act. In disfluency, gestures are also affected, the speaker stopping both speech and gesture, suggesting that they are both being monitored (Seyfeddinipur, 2006, p. 146).

If gestures are considered part of the speech-gesture act any discourse analysis should also take them into consideration. The term 'gestures' can be misleading as it can be interpreted to mean any movement with almost any part of the body, and at the same time it gives the impression of excluding gaze. Most scholars believe that the communication act involves all of the above and as such a microanalysis of discourse needs to be carried out at many levels to get the full picture of the exchange (Bavelas, n.d.). Kendon defines gestures as "visible action […] used as an utterance or as a part of an utterance" (2004, p. 7) and utterance as "any unit of activity that is treated by those co-present as a communicative 'move', 'turn' or 'contribution'" (p. 7). We understand this definition to include the speaker himself or herself as being co-present. However 'gesture' can also be used in more narrow terms to indicate movements performed with the hands (including fingers and arms very often) that are an indivisible part of the speech (López-Ozieblo & McNeill, forthcoming, 2015), which is the meaning we will use throughout this study.

Bavelas (1994, p. 202) states that gestures have various functions, including a linguistic one that helps "convey meaning to the addressee", contributing to the speech with a particular meaning in a given context. If the linguistic function is key, then an analysis of gestures ought to be made in these terms, rather than looking at the movement itself.

[8] Clark (1996, p. 285) disagrees with this distinction on the grounds that some disruptions might be due to addressee's failure to decipher the utterance. In our study all 'self-repairs' are clearly initiated by the speaker after self-disrupting the speech.

Gesture needs to be linked to the individual, the context and the speech content to understand its meaning. Often, studies focus on how the gesture is produced – what it looks like – rather than what it is achieving, studying the physical movement rather than the linguistic function. The function – the focus of this study rather than gestural form – can be multiple, a gesture occurring at a disfluency might indicate the speaker is keeping the floor, and at the same time is requesting for help from the addressee or trying to recollect a word.

4.1 Gesture-speech Link

McNeill (2012) proposes that due to the proximity of the manual and oral centres in the brain "hands are special" (p. 154). This is supported by neurological evidence by Meister et al. (2003) who suggest a strong hand-speech link in the brain. An important distinction that needs to be made is that we are focusing on gestures concurrent with speech, this excludes sign language and pantomime (a gesture specifically appearing to cover a gap in speech). Gesture-speech is thus considered as one, two modalities to formulate the concept which in itself exists thanks to them, this is the Growth Point (GP) Theory (McNeill & Duncan, 2000), the most extreme of the gesture-speech theories, supporting the linguistic, communicative and conceptual role of gestures. Other theories, also advocating a gesture-speech link, are based on serial speech production models (in particular Levelt's) but view gestures as an add-on to speech, fulfilling either a communicative role, for the addressee (Clark, 1996; Goodwin, 2000; Kendon, 1994), or a lexical, cognitive one, for speakers themselves (Iverson & Goldin-Meadow, 1997; Kita, 2000; Özyürek & Kita, 1999). Krauss, Chen & Gottesman (2000) believe, at best, in a supplementary role of gesture to aid speech processing. They claim that there is insufficient evidence to confirm that all gestures accompanying speech are communicative in nature (p. 273).

Regardless of the theory, the facts seem to indicate that gesture is part of the speech process and as such should be included in the study of discourse. The limited research into the field of disfluencies and gestures indicates that gesturing might reduce overall disfluency (Finlayson, Forrest, Lickley & Mackenzie Beck, 2003; Rauscher, Krauss & Chen, 1996); that gesture stops with certain disfluencies (Seyfeddinipur, 2006); that the type and number of gestures in disfluencies might differ depending on the level of proficiency of the foreign language speaker (Graziano & Gullberg, 2013); and also on the proficiency of children learning to speak their mother tongue (Colletta, Pellenq, & Cefidekhanie, 2014). Due to the idiosyncratic nature of gestures, studies usually concentrate on a few individuals – in some cases just one (Brown, 2010; Stam, 2014); and on specific elements, including: the type of gesture (Kita & Özyürek, 2007); the characteristics of the speakers, such as those with lesions (Duncan & Pedelty, 2007), children (So, Demir & Goldin-Meadow, 2010), the blind (Iverson & Goldin-Meadow, 1997), bilinguals (Stam, 2012), second language speakers (Gullberg, 2008); the actions accompanying the gesture (Chu & Kita, 2011), the content (Casell, 2007), or semantic elements (Stam, 2006); etc.[9] A number of studies have covered disfluency and gesture (Seyfeddinipur, 2006) and gesture and second language speakers (Gullberg, 2008; Nicoladis, 2007; Stam, 2012), but to our knowledge only Graziano and Gullberg (2014) have focused (in an on-going study) on gesture and disfluency specifically targeting second language speakers.

4.2 Verbal Communication

The processes of speech production are still being debated, not much is known about them and what has been proposed is mostly based on the study of speech errors. Two frameworks exist on which speech production models are based, a serial and a parallel processing framework. The serial processing framework suggests that there are a series of steps, following each other, that begin with an idea which is then translated, or reformulated into words and eventually articulated. On the other hand the parallel processing framework advances that speech is generated when a series of parallel nodules are activated concurrently at a semantic and phonological level. In both cases it has been proposed that there is also a monitoring function that provides feedback and detects most errors before they become overt, just as with all other bodily functions.

4.2.1 Serial processing framework

The study of error allowed Fromkin (1971) to suggest a speech processing model that explained most of the consistencies observed in error production, such as a preference to mix words of the same syntactic class, relationships between stress and the number of

[9] For a meta-study of gesture studies see Hostetter, 2011.

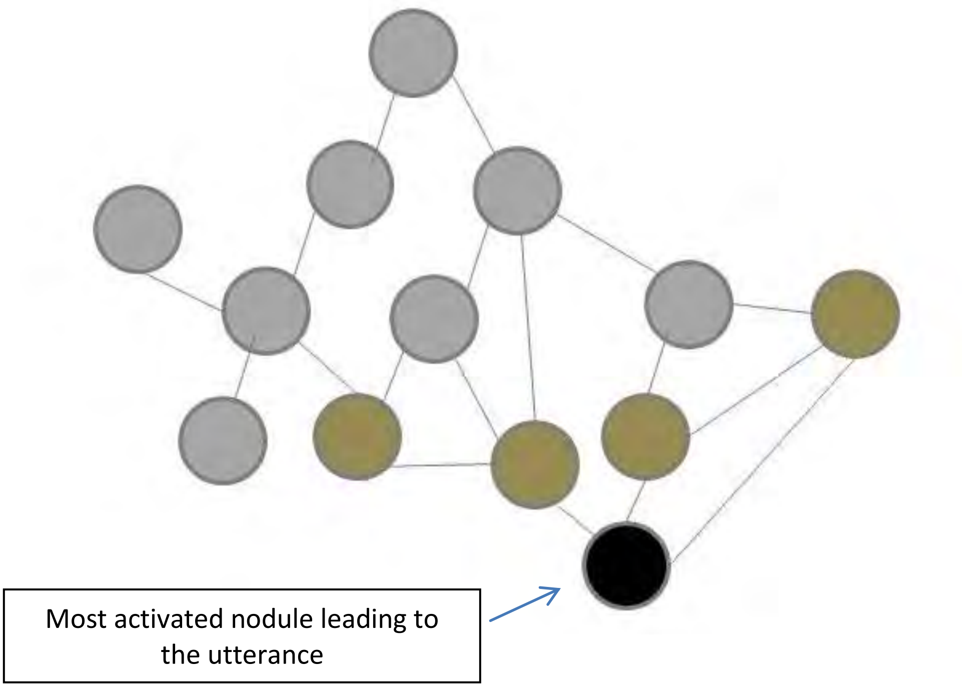
syllables or patterns in phonological slips. The model was represented by a series of modules where the process flowed from top to bottom. In this process the meaning that is to be conveyed is first generated; then "the 'idea' or 'meaning' is structured syntactically, with semantic features associated with parts of the syntactic structure" (p. 239), at this point two modules, working in parallel, generate the syntactic structure and the semantic features. This is followed by a module that adds a preliminary intonation contour. From here the structure is filled with the lexicon. The final modules specify the phonetic elements that lead to articulation through "motor commands to muscles" (p. 240).

This model is not mentioned often in the literature and yet it is clearly the foundation for all serial processing models. All the processes start with the generation of the 'intended meaning' (Fromkin's terminology. 1971), the 'concept' (Levelt, 1983) or the 'message' (Garrett, 1975, 1992). The idea is then formulated in the next module – here the above scholars vary slightly as to what element is formulated first, usually the lemas and the syntax building – and finally the phonological elements are brought together and transferred to the articulator module. Initially Garrett suggested that there were "two distinct levels of syntactic analysis" (1975, p. 175) hinting at the possibility of a monitoring system that functioned as a parallel system checking on the output of every level of the speech production process. The flow of information, originally travelling one way, was revised by Levelt (1989) to accommodate for an internal speech monitoring mechanism that allowed a two directional flow of information. Serial processing frameworks suggest that at the conceptualizer module a complete idea comes into being, not in words as these are added only later; then a grammatical structure gives it order, words are found and finally articulated.

4.2.2 Parallel processing framework

A better understanding of neurological processes suggests that a parallel processing mechanism might be a more likely framework. Dell and Reich (1980) suggested a parallel processing model where nodules with information leading to speech (lexical, phonetical) are interconnected. An idea activates certain nodules which in turn activate other related nodules (p. 277). Eventually one nodule will receive more activation than others and this is the one that will be used in the utterance. They argued that in a nodular model, editing is part of the speech process itself, and not a separate loop or step (p. 281). See Figure 2.

Figure 2. – Parallel speech processing system.



Neurological networks are based on what is perceived from the physical world, which in turn is likely to have shaped grammar. Our knowledge, based on our embodiment of the physical world, has created a series of frames to which we refer at any given situation, the context. All of these together, working in parallel would give rise to speech production. If we imagine a situation which puts the action at grandma's kitchen, at tea time, the relevant frames are ready to monitor that the language production fits within this context and describes it adequately. This means that the most likely neurological networks activated are going to be the ones relating to grandma, food (grandma always has food and it is tea time), and maybe relating to the cat (grandma has a cat who likes to sleep in the kitchen). At the same time a system of sub-frames is also at the ready. These might carry information that relates to likely actions and are activating networks that refer to eating, giving, petting, sleeping and meowing. Because the goal is to speak, the motor system will also be alert, with a frame that will dictate what sounds are acceptable to perform that function. In order to form the following sentence, Grandma gives a biscuit to Anne

a number of semantic networks are ready with the most likely possibilities of what might take place in the kitchen; syntactic, phonological and motor networks that will link to each other to give the most acceptable solution, based on what has been previously experienced. Just at the semantic level we can imagine a series of nodules activated with the options available: mum, dad, Anne, grandma, etc. As nodule 'grandma', is activated all its previous known links will also be activated, including 'biscuit', which is the strongest candidate based on previous knowledge given by the frames. The process leading to the articulation of the first word ('grandma' in this case) is probably well underway, with all the neurological-network activation this entails, before the sentence is conceptually completed, especially if the utterance is a more demanding one than the one suggested in this example.

4.2.3 Neurological description of the speech process

Over the last few decades the study of the brain has advanced our understanding of the speech process. The work of neurolinguists backs up the parallel speech process mechanism. Studies in neurolinguistics (Bergen, 2012; Kemmerer, 2010; Lakoff 2013) leave no doubt that in studying speech we need to take into account the body. Lakoff (2013) bases the speech process on the activation of superimposed imagistic frames (each carrying the experienced understanding of a certain context), as common elements in the frames are found they become more activated than others and in turn activate other related elements. If we are talking about eating in a restaurant the frames will include that of a restaurant (with all the content expected in one: food, waiters, tables, etc.); that of a service provider (including money exchange, a somewhat formal interaction, the expectation of receiving a service, different types of service in different locations, etc.); eating (including food, food utensils, drinking, etc.). As all of these frames superimpose the elements referring to being seated at a table eating the food brought to us by a waiter will give us the concept we want to express (or process). This particular network of links, reflecting a common situation, will be fairly strong and is likely to activate first when we talk about restaurants. Should we find ourselves being manicured by the waiter instead, it will be necessary to create a new network incorporating this information. It is possible to do so if there are already some existing frames and links between them to work with, but very difficult otherwise (these neuronal networks are set up by the age of 7).

Bergen (2012) confirms that we build visual simulations of objects about which we hear or read, these include characteristics such as orientation, colour and shape but also location and motion as well as how to interact with them – affordance (Ch. 3). Processing speech means building motor simulations of situations as if we were there, re-enacting physical actions (going back to the expected frame scenario); this is confirmed by brain imaging experiments (Hauk, Johnsrude & Pulvermüller, 2004; Tettamanti, Buccino, Saccuman, Gallese, Danna & Scifo, 2005) that demonstrate motor activation during speech processing and on the work done on mirror neurons.[10] Logically, then it ought to be faster to visualize nouns that refer to items with which we interact more easily (like a grape versus a pond), as is the case. As we process language (text or speech) we are making predictions as to what will come next based on what has been said. Bergen provides the following example, when reading the sentence:

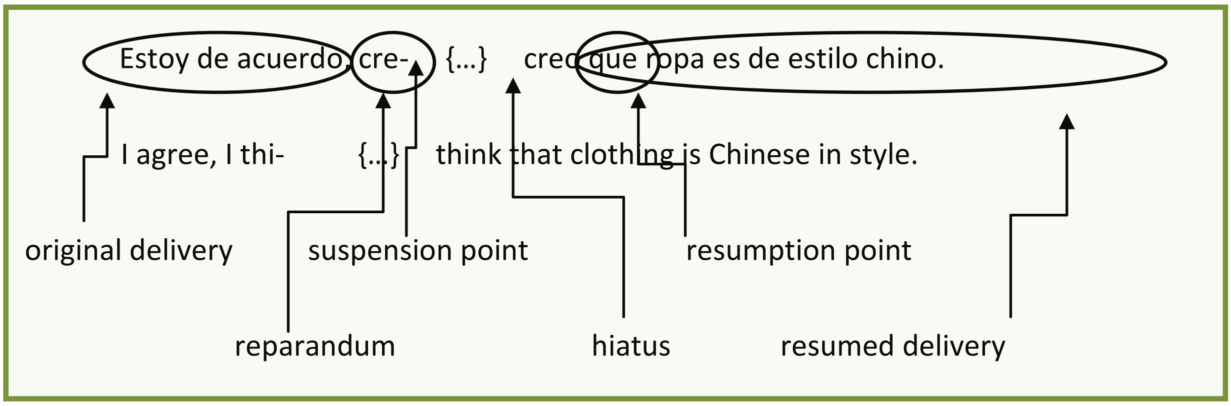
> The lawyer cross-examined by the prosecutor confessed.

The reader is likely to do a double take when reaching 'by the prosecutor' as it is not the expected ending of the sentence. Existing networks will have stronger activations between a lawyer cross-examining a witness than being cross-examined himself. This also confirms that we do process speech immediately, and not just wait for the end of the sentence. We begin to visualize the scene based on our knowledge of the world as soon as there are clues that allow us to do so, and create the scene incrementally based on the information that is given (Bergen Ch. 6). This means that if we don't share experiences (frames), be it at a professional or cultural level, internal simulations will differ and so will understanding. It also means that we are likely to process more slowly frames that are not usually found together, often externalising those conflicts, when speaking, as a disfluency. New information, or different superimpositions of frames create newsworthy content that speakers usually emphasise (McNeill, 2012, p. 33), probably recognising the higher complexity involved in its processing.

[10] See *Neurological facts supporting the gesture-speech argument* (Section 4.4.3).

4.2.4 Speech conflicts

We are able to understand speech despite the numerous instances of repetitions, pauses, clarifications, etc. that might render it disfluent. An ideal delivery, one without performance additions, is very difficult, unless it has been prepared — such as the delivery of actors, priests, lecturers, politicians, etc. Most of us are likely to produce slightly less than fluent — disfluent — speech which is marked by a disruption to the 'original delivery' and then a repair. Clark explains the process as follows: the 'original delivery' is disrupted at the 'suspension point', followed by a 'resumption point' when fluent speech, the 'resumed delivery', restarts. If the speaker recognizes an error or detects trouble, the part of the original delivery where the error/trouble is detected is termed the 'reparandum'. The reparandum might be repeated or modified after resumption. In between these two points there could be a pause termed 'hiatus' which could be filled or not (Clark, 1996, pp. 258-259), it is here that the disfluency might appear. Resumptions can be the continuation of the utterance or a repair that allows the speaker to go back to the original delivery and clarify it, making it easier for the addressee to ignore the disruption. The repair (also termed 'replacement') can affect just the reparandum or elements before it. For example in the following utterance[11] from our corpus of Hong Kong (HK) students learning Spanish:



After the hiatus speakers resume their speech and they can do so by just continuing the speech where it was left off, repeating, substituting, deleting or adding words. Clark (1996, p. 265) refers to the last three as self-repair but proposes a new term for them, 'replacement', as in all instances an element has been replaced (in the case of a deletion, by another word). He believes that the resumption of speech is just the ideal utterance the

[11] All transcriptions record the exact utterance, including grammatical and articulation errors.

speaker would have liked to have used in the first place. Speakers were not ‘adding’ an element because a new idea come to them as they were speaking, but because they found a better way of expressing that idea. It is important to remember that this is based on the serial speech production models mentioned above, where the idea is formed in the conceptualizer and ‘translated’ or formulized through words to be articulated.

Repair, a term proposed by Schegloff, Jefferson and Sacks (1977), expresses much better than ‘correction’ what takes place in spontaneous speech. A repair isn’t always the result of an error, mistake or fault – the conflict – nor is it always necessary (Schegloff, Jefferson & Sacks, 1977, p. 363). Repairs rely on the ‘error’ being noticed in the first place, which is dependent on the context, and the attention the utterance is receiving by the speaker (Levelt, 1989, p. 463). It can be self or other- initiated, although on most occasions (except in parent-child interactions) it tends to be self-initiated, as it seems that the opportunity to self-initiate comes more readily than that for other-initiation. The repair can come within the same turn, during the transition space between turns or even at the third turn (when the original trouble source speaker talks again). With self-initiated repairs, the majority are initiated during the same turn or the transition space and they are indicated by the speaker with a non-lexical marker or filler, such as ‘uh’ or a sound lengthening, followed by the repair (Goodwin & Goodwin, 1986, p. 55; Jefferson, 1974, p. 186).

The addressee doesn’t usually have a problem interpreting the self-repair correctly, at least amongst native speakers, regardless of where the interruption takes place, the length of the pause or how the repair is accomplished. One aspect helping the addressee, that won’t be discussed here, is that repairs follow syntactic rules of coordination.[12] Another aspect is that in spontaneous dialogue the speaker can also make use of non-verbal communication, Cutler (1983, p. 79) distinguished between prosodically marked and unmarked repairs. Prosodically marked repairs being those where the prosody of the trouble and the repair terms differs, in pitch, loudness or duration (Levelt & Cutler 1983; Moniz, Batista, Mata & Trancoso, 2012). Levelt and Cutler (1983) found that repairs were prosodically marked when the trouble was semantic. Therefore inappropriateness will not show many prosodically marked repairs.

[12] For a detailed explanation see Levelt, 1989, pp. 486-489.

Self-repair is considered by Levelt (1984, p. 105) as a rather complex phenomenon involving disparate phonetic processes, such as self-monitoring, production and detection of phonetics, lexical and other types of speech errors, self-interruption, prosodic marking of corrections etc. Repair then, can have a number of objectives apart from correction of a trouble source. It can also be used to attract the attention of the addressee (Jefferson, 1974, p. 186; Goodwin & Godwin, 1986, p. 55); manage a turn, taking it when there is overlap (Schegloff, 1987, p. 75), starting it when entering an ongoing discussion (Egbert, 1997, p. 632); to create an affiliation (Maheux-Pelletier & Golato, 2008, p. 689); to clarify the meaning to the addressee (no error involved); or to make the sentence more appropriate (Levelt, 1989, pp. 458-459). In general it is possible to classify repairs into those related to error (phonetic, lexical or syntactic) and those related to appropriateness, where the utterance was correct, phonetically, lexically or syntactically, but not so from a contextual or pragmatic point of view.

The following examples are adapted from our corpus[13] of HK students learning Spanish:

Syntactical:

Creo que ropa es es tiene estilo chino

I believe that clothing has has is Chinese in style[14]

Lexical:

Creo que la camisa la ropa tiene estilo chino

I believe that the shirt the clothing is Chinese in style

Phonetical:

Creo que la lopa ropa tiene estilo chino

I believe that the clothing (wrong pronunciation in Spanish) is Chinese in style

Appropriateness:

Seguro que la ropa creo que la ropa tiene estilo chino

For sure the clothing I believe that the clothing is Chinese in style

These are usually referred to as 'overt' errors and repairs, in that their nature can easily be guessed. There is also, in normal speech, a high percentage of repairs that are harder to explain as the utterance in our corpus:

cre-creo que ropa

I be-believe that clothing

The example above is a 'covert' error, often disregarded in repair analysis as they are not easy to classify, we cannot guess the source of the disfluency (Levelt, 1989, p. 478). The cause of the errors might not be clear, but the disfluency is obvious. In our analysis we have included these cases as well as overt errors that have also resulted in disfluencies (there are a few cases when errors are not followed by disfluencies or repairs of any sort, these few instances have not been included).

---

[13] This was the corpus used for the preliminary study. Participants had to describe a photograph using vocabulary relating to physical descriptions and clothing which had been covered in class.

[14] The translation inverts the use of the verbs to give the sense of the correction in English.

There have been a number of studies classifying the types of issues that lead to disfluencies and repairs (Cutler, 1983; Levelt, 1983; Nooteboom, 1980) with similar results. Lexical troubles have been observed as being one of the highest causes of speech suspension. Levelt reports that within overt repairs those due to lexical errors make 38 per cent of his corpus; phonetic errors are 1 per cent of his corpus; and other issues, we assume of a syntactic nature, add up to 2 per cent.[15] Repairs due to problems with appropriateness were 30 per cent of his corpus. In addition he also identified 25 per cent of the repairs as being covert (not clear of what the trouble source is) and 4 per cent as belonging to other categories not specified (Levelt, 1984, p. 108). The most common type of errors is the one that entails swapping of words with identical grammatical functions at the same time as keeping the phonological rules intact.

In the utterances below (taken from our corpus[16] and based on a cartoon story where Granny hits a cat with an umbrella):

> Granny gives the cat a whack.
>
> Granny gives the cat an umbrella.

notice how the phonological rule is kept in 'a whack' but 'an umbrella'.

---

[15] Although replacement repairs have been considered to solve issues related to pronunciation or caused by selectional difficulties (Nemeth, 2012, p. 2033), in reality, without context and the cooperation of the speaker, it might be very difficult to categorize the replacement. Clark (1996, p. 265) gives the following example of substitution:

> I don't think they've {.} they ever in fact embodied
>
> (the brackets signal the suspension and resumption point with the hiatus in between being a pause).

Without the original recording (and even with it) there will always be a doubt whether the original delivery was meant to be a contraction of 'they have' or whether it was a contraction of 'they ever' that was cut off. In the first case it would be a substitution and in the second a repetition with a clearer pronunciation.

[16] This was the corpus used in the main study. It is based on the retellings of three stories. The first story was based on Aesop's fable of the Lion and the Mouse. The second and third were cartoons where a cat (Sylvester) is trying to catch a bird (Tweety) and the bird's owner, Granny, always saves the bird and ends up hitting the cat.

The serial framework states that content word selection or functional assignment occurs at one level – and so it explains why words from outside that function are rarely used instead. Morpheme selection, taking place after the structure is in place, account for the change of ‘a’ to ‘an’. The parallel framework accounts for the error as both words (whack and umbrella) were acceptable options with the verb ‘gives’, one more suitable than the other as it is not often that a cat gets given an umbrella. In the activation mechanism ‘whack’ is exclusively linked to ‘a’ rather than ‘an’ and ‘umbrella’ is linked to ‘an’. Equally, both systems, the serial and the parallel, account for errors that create unintended existing words rather than non-existing ones, showing there is lexical bias (there is a preference for using existing words rather than combinations of syllables that could exist as words but do not): the parallel framework because those options are unlikely to be selected (not enough activation) and the serial framework because an internal monitoring system would catch the error before it is articulated.

Most studies focus on the errors in linguistic elements, including prosodic elements such as intonation, volume or pauses, but completely ignoring gestures. Seo and Koshik (2010), pointing out this gap in the research, studied the gestural performance of speakers and their addressees in other initiated repair situations, although focusing only on head movements and gaze. They state that some head gestures, used without speech, can be used to indicate, systematically, other-initiated repair that prompts repair by the speaker.

4.2.5 Monitoring processes

Speech production processes, serial or parallel, both have a monitoring element, that repairs ‘errors’ before they are externalised, so they are covert, or after being uttered, rendering them overt. The main issue is identifying what is a speech ‘error’, this is a highly subjective issue as an error is such just because the speaker thought he or she could do better. Any conflict that affects any element of speech production, concept, syntax, lexicon, phoneme, prosody, morpheme, appropriateness, and context could be considered an ‘error’ (Postma, 2000, p. 103). Detecting and repairing ‘errors’ might result in corrective feedback, used in inner and outer speech monitoring (Postma, 2000, pp. 101-102; Corley, Brocklehurst & Moat, 2011, p. 173).

In addition to relatively easy to identify conflicts in externalised speech, there are also disruptions which an observer cannot classify as related to any of the above categories; even the speaker would have difficulties in doing so most of the time. These externalized occurrences, often only manifested as a disfluency, hiccups in the speech process that are not linked to an apparent error, were considered by Levelt (1983) to be indications of an internal repair mechanism, assuming that they were indicators of trouble in the speech process. The likelihood is that when speakers suspend their speech it is either because they have run out of things to say, don't know how to say them — a planning issue that leads to an unintended disruption — or because they had planned to do so, intending the disruption to be a signal to the addressee: indicating a turn change, requesting help with a word or searching for it, thinking (pre-concept formation) or inviting the addressee to do so (Clark & Fox Tree, 2002, pp. 90-91). Disruptions can also be the result of other conditions including distractions, emotions, high cognitive load or brain lesions.

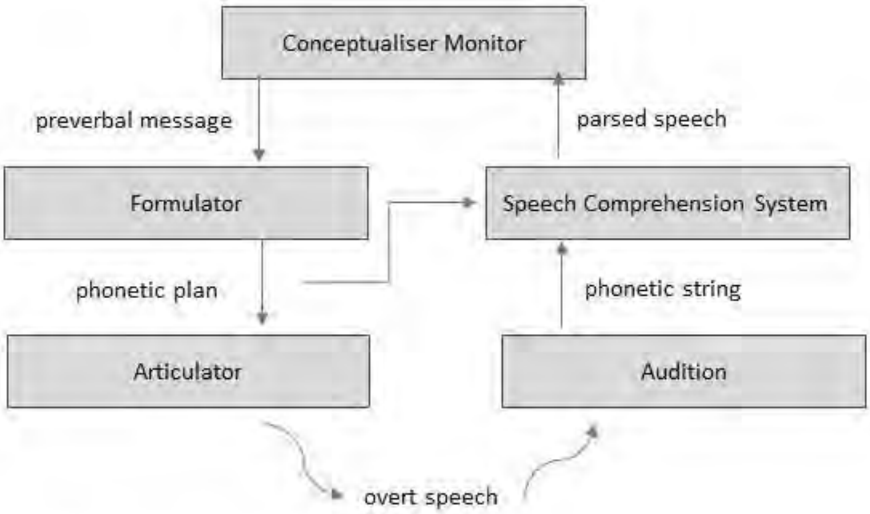
Much research has been conducted on where and how monitoring occurs; as it falls outside the scope of this study we will only expand some key ideas.[17] Subconsciously our body is constantly monitoring and adjusting what it does, at all levels, from hormonal to motor, so the existence of a speech monitoring process is more than likely. It has been proposed that overt speech, what is externalized, is the result of a monitored speech process (Hartsuiker, Roelin, Postma & Wijnen, 2005, p. 3).

Many hypotheses have been put forward as to how errors are identified. In most cases there is an internal monitoring process that provides constant feedback. Proponents of the serial processing mechanism, such as Levelt (1983, 1989) believed this to be one single loop, the perceptual loop theory, with a comprehension system monitoring internal speech production and error detection taking place at the conceptualizer level. The perceptual loop theory proposes a monitoring system that deals both with self-speech (inner and overt) and other's speech, allowing speakers to identify troubles in all. This is based on Levelt's Conceptualizer-Formulator-Articulator process (1989, p. 9). The Conceptualizer generates the message, and also monitors for errors at this stage; the Formulator carries out the grammatical and phonological encoding; and the Articulator

[17] For a detailed description see Postma, 2000.

presents it as overt speech. As the speaker listens to his or her own inert or overt speech they can analyse it with the Speech-Comprehension System and detect issues either before or after the utterance has been voiced (see Figure 3).

Figure 3. – Serial speech processing system.



Adapted from Levelt (1989, p. 470)

Other scholars (Blackmer & Mitton, 1991; Postma & Oomen, 2005) propose a similar model but with at least two systems, one self-monitoring internal speech (production) and one self-monitoring outer speech (perception). In both models it is believed that most of the time the internal monitoring process catches all falsely activated nodules (in the parallel processing framework) or choices given by the various modules (in the serial processing framework), discarding them before the motor process has kicked in, when it does not do so, errors slip out. Closer to the principles of a parallel processing mechanism, MacKay (1987) believes in a language system which is served by a series of nodes (propositional, conceptual, syllable, etc.), arranged in layers and connected among themselves. As one layer is activated it primes the relevant nodes which in turn will activate other nodes they are connected to. The node which receives the most activation will be the current one. In this model it is clear that production and comprehension are linked, although Levelt disagrees with the two systems operating so closely together (Levelt, 1989, p. 476). Supporting this theory are the results when a speaker is made to hear his own utterances with a delay of about 200 milliseconds (described in Levelt, 1989, p. 477), as might be the case when speaking through a microphone not well synchronized that produces a delay. Such a speaker will be seen to have trouble speaking; stuttering, repeating words and in general producing disfluent speech. MacKay explains that upon activation the node will deactivate almost immediately, to allow for the activation of another node. If the original node is re-activated, as it would be when processing the delayed speech, it confuses production. In our opinion these results clearly prove the link between the processing and comprehension systems.

4.2.6 Disfluencies

There are conflicts, not necessarily errors, which the monitoring mechanisms seem to identify and signal with one or more of the following five suspension devices, which we will refer to as 'disfluencies': pauses, word cut-offs, elongations, non-reductions of vowels, and fillers.

The following examples, from our corpus of HK students learning Spanish, illustrate them (the disfluencies have been underlined):

Example	Disfluency
this time <u>um</u>	Filler
Sylvester tries to get <u>uh</u>	Filler
office <u>in-</u>	After word completion word cut-off (marked with a hyphen)
<u>re-</u> in reception	Within word cut-off (marked with a hyphen -)
of the apartments <u>/</u>	Pause (marked with a slash /)
and listen at th<u>e</u>	Non-reduction (the last vowel is pronounced as a /i/ rather than an /ə/)
grandma is trying to get <u>a:</u>	Elongation (marked with a colon :)

The above are described at length by Clark and Fox Tree (2002) in a study of utterances in English made by British or American speakers[18] (p. 81). Below we detail those disfluencies which we found relevant to our study.

4.2.6.1 Pauses

Pauses are the most common disfluencies in spontaneous speech (Clark, 1996, p. 271). As happened with other disfluencies, pauses were not considered part of linguistic study until the 1950s with the studies of Maclay and Osgood (1959), and later Quinting (1971). Pauses can also contribute linguistically to the delivery in a number of ways, including the introduction of new 'thought units' that can provide information about the cognitive processes (Rühlemann, Bagoutdinov, & O'Donnell, 2011). McGregor, Corley and Donaldson found that "listeners were more likely to recognise words which had been encountered after silent pauses" (2010, p. 3983).

In North American and British speakers of English, pauses need to be less than a second long not to become a problem in the conversation, this interval is termed "the standard maximum tolerance" (Jefferson, 1989, p. 166), so longer pauses, when the speaker intends to retain the turn, tend to be filled. A long pause breaks the dynamics of the conversation and creates instability in the coordination of turns between speakers, as it is unclear whose turn it is. However, pauses can be longer or shorter depending on the speaker and the context, a fact taken into account by the addressee. Schegloff (2000) suggested that pause lengths less than 150 to 200 milliseconds are not perceived as cessation in speech by listeners.

Pauses can be filled with an utterance such as 'uh' or 'um' or they can be just pure pauses. The literature calls these 'unfilled pauses', Kjellmer (2003) calls them 'silent'. Poyatos (1993) explains how a pause cannot be unfilled as it is itself "an interactive segment with a very determined structure" (p. 136). To avoid confusion, we will use the term 'pure pauses' to refer to pauses which are not preceded or followed by another type of disfluency.

---

[18] The authors don't specify whether the speakers were native or if English might have been a foreign language to them.

4.2.6.2 Filled pauses

Some pauses are filled with a sound (a word according to some researchers,[19] such as ‘uh’ and ‘um’, the vowel pronounced as /ə/ (a schwa sound) which we will refer to as fillers. Other fillers can be glottal sounds, sounds like ‘tsz’ or editing expressions such as: ‘that is’, ‘like’, ‘I mean’,[20] etc.; as well as other elements like: throat clearing, laughter or gestures (Clark, 1996, p. 261). Monosyllabic fillers are also present in the literature as ‘ah’, ‘er’, ‘erm’, ‘mm’, ‘uhm’; we have chosen to refer to them in their American transcriptions ‘um’ and ‘uh’ within this text[21] unless citing an author that refers to them otherwise.

Lexical fillers have been studied in detail by Corley, MacGregor and Donaldson (2007); Clark and Fox Tree, (2002), showing how particular events prefer certain fillers. Specifically Levelt (1989, p. 483) differentiates between lexical fillers used to repair error troubles (mainly: ‘er’,[22] ‘or’, ‘no’), used in 62% of cases in his corpus and those used to repair appropriateness troubles, used in just 28% of cases. He singles out ‘er’ as being the most used lexical marker, a universal sound/word, employed the most. It is used in 30% of all repairs of his corpus, mostly with covert repairs.

In Tottie’s meta-study of six corpora (2011) the results indicate considerable differences in the use or ‘uh’ and ‘um’ in relation to each other (with one filler used 10% more than the other) and depending on the characteristics of the corpus analysed – such as formality (fewer fillers); speaker’s gender (men produce more fillers of the type ‘uh’); socio-economic factors (more educated speakers use more fillers) – but on average accounting for 50% of the words in each corpus (2011).

With ‘uh’ and ‘um’ the most frequently occurring preceding word is ‘and’. This suggests that the filler might be used to introduce a thought unit, new information requiring more cognitive effort, confirmed by the fact that more fillers are used before indefinite articles (Kjellmer, 2003, pp. 174-176). Kjellmer studied the use of fillers in detail proposing

[19] See Tottie, 2011 for a review.

[20] See Clark (1996, p. 262) for an interpretation of these editing expressions and others.

[21] However the transcriptions record whether participants uttered ‘uh’ or ‘eh’ sounds.

[22] Levelt seems to refer to ‘uh’ and ‘um’ as ‘er’ in Levelt, 1989.

that speakers might not be consciously aware of their use of fillers but these "facilitate the communication" (2003, p. 191) indicating hesitation, turn management, attracting attention, highlighting, correcting (Kjellmer, 2003) or what Tottie refers to as "planning" (2011, p. 193), a more positive and accurate term than filling.[23] However unconscious their use might be, the type of filler used seems to be a controlled choice, 'um' indicating a longer pause than 'uh' (55% compared to 20%) (Clark, 1996, pp. 263-264). Corley and Hartsuiker (2011) explored whether fillers facilitate addressee's word recognition, and concluded that the fillers themselves were not as relevant as the actual delay they entail, as a slower rate of words allows for better processing.

4.2.6.3 Elongation

Elongation of a consonant or a vowel, often followed by a pause is, together with pauses, the main expression of hesitations, also giving the speaker time to think. Elongations alone are more likely to be used if the disruption is going to be short and they are thought to give a greater sense of fluency (Clark, 1996, p. 268).

4.2.6.4 Reduced vowels

In English, reduced vowels are found when speaking fluently, where words like 'the', 'a', 'to' will be pronounced with vowels which are not fully formed. Reading this sentence with a metronome, a word for each beat, will produce a different 'a' than without it (Clark & Fox Tree, 2002, p. 99). If the speech is suspended at any of these words (very likely as 'the' and 'a' are frequent words) the vowel is likely to be pronounced fully, as done when saying it to the beat of a metronome.

4.2.6.5 Word cut-offs

Word cut-offs include the cutting of words before they have been completed or they can also be a sudden stop after completing a word, often indicated by closing the glottis.

---

[23] In this study we will continue to refer to them as 'fillers' for clarity, although we agree that 'planners' might be a more appropriate term.

Within-word cut-off	After (completed) word cut-off
if you please just leave me alo- eh just let me go	released the – the lion

Cut-offs are usually followed by a short pause or by any of a number of editing expressions of various lengths including: 'uh', 'um', 'that is', 'sorry', 'I mean', 'I will start again, OK?', etc.

The research done in cut-offs has led to a number of hypotheses as to when and why the cut-off occurs. The Main Interruption Rule (MIR) proposed that speakers stop as soon as they need to, which would mean that stopping could occur at any time (Nooteboom, 1980), this rule was modified by Levelt (1989) who proposed the cut-off depended on whether the word to be cut-off was erroneous or not, suggesting a strategic stop rather than an automatic one, giving a clear indication to the addressee of an error. On the other hand the Delayed Interruption for Planning Hypothesis (DIPH) (Seyfeddinipur, 2006; Seyfeddinipur & Kita, 2001; Seyfeddinipur, Kita & Indefrey, 2008) proposed that the cut-off is a controlled action and so on detecting the trouble the stop might be postponed, if necessary, to give time to the resumption process; this hypothesis is based on the work by Blackmer and Mitton (1991). Tydgat, Stevens, Hartsuiker and Pickering (2011), carried out a practical study comparing the various stopping theories and concluded that the most likely scenario was the DIHP. They add that the stopping and resumption processes are likely to occur concurrently and share the same resources, therefore the speaker has to decide whether it is more effective to stop and where, or not – this they called the Modified MIR with Strategic Allocation of Shared Limited Resources (2011, p. 379).

4.2.7 Disfluencies in Spanish

The relevant literature in disfluency and repair in Spanish is limited to a few papers (Edmunds, 2006; Gelbes, 2003; Rodríguez & Torres, 2006). Most researchers base their work on English language studies and conclude that disfluencies in Spanish are equivalent to those in English: Rodríguez and Torres (2006, p. 333) found that rates of disfluency in Spanish are comparable to rates in English; Edmunds (2006), looking at Spanish spoken in

Southern U.S. concluded that, as in English, repair is higher in function words than in content words.

Rodríguez and Torres (2006) worked with different audio corpora, one with humans–computer speech and the other with human-human speech. It is the latter that is of interest to us, it was taken from Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC), from the Universidad Autónoma de Madrid. This corpus consists of 42 debates and interviews from Spanish TV and radio broadcasts. As the purpose of the study seems to be focusing on analogue speech (speech used in radiocommunications), Rodríguez and Torres' terminology varies from that of more linguistic studies. They divide disfluency into four types of events; spontaneous speech events (SSE) of which self-repair is just one type. The other events are: acoustic; lexical distortions; and discourse fillers. As this paper by Rodríguez and Torres is one of two on the subject of disfluency in Spanish, we have considered it necessary to match the terms used by the authors to those in other studies on self-repairs (Clark, 1996; Gullberg 1998; Levelt, 1989; Seyfeddinipur, 2006) as mentioned above. The four types mentioned by Rodríguez and Torres (2006, pp. 338-341) are explained by them as follows (text in italics as in the original).[24]

- Acoustical events refer to pauses and lengthened vowels or nasalizations that can be used to fill a pause after interruption, or to stretch a phoneme within a word – termed *prolongations*. In Spanish the most common pause fillers are of the vowel e, (/ɛ/); *a sort of nasalization sounding like /m/*, by which we believe the authors might refer to /ŋ/; *and a sound like the vowel /a/, which is close to the English uh,* which we believe might refer to / ʌʔ/. The authors do not give specific *prolongation* sounds. Although as they mention that usually no difference is made between fillers and *prolongations* it is safe to assume that they refer to the same sounds.
- Lexical distortions seem to refer to pronunciation issues: *alternative pronunciations or articulatory errors that […] do not pose a problem of understanding, so he/she leaves them uncorrected.* Here the authors include mispronunciations, such as 'pasiensia' instead of 'paciencia', any variations from what they call *proper or*

[24] We are aware that there are a number of linguistic and sociolinguistic issues related to this classification which are not relevant to this study.

canonical pronunciation […] standard Spanish used in broadcast news in Spain. Also in this category are cut off words, *which are immediately repaired either by the same word repeated correctly or by another replacing the first one*.

- Speech repairs are divided into self-repairs (including repeats […] self-corrections or reformulations with substitution, insertion or deletion of words) and abandoned phrases, that *sometimes include hesitations and self-repairs*.
- Discourse fillers used to explain or edit (such as: *perdón, quiero decir, etc.*) and to fill or pause (such as: *bueno, mire*, etc.).

One interesting finding Rodríguez and Torres (2006) recorded was that there was no obvious correlation between the utterance length and the rate of disfluency (where all four types described above plus another called disfluency[25] are considered), we believe that there might be a classification issue obscuring their results. The authors state that, as in English,

> *[…] no single cue can be reliably used to detect speech repairs, since either the frequency or the coverage (or both) are too low. Lengthenings and cut off words are found to be the strongest but not definitive* (Rodríguez and Torres, 2006, p. 363),

it is important to take into account that the focus of Rodríguez and Torres is the understanding of human speech, in order to improve human-computer dialogue in automated call answering. Nevertheless their results are relevant, and might support our theory that in spontaneous dialogue the hands might also play an important role in regulating speech and indicating events such as repairs – which is why a single cue might not be enough to reliably detect the repair, as the verbal cue is usually accompanied by a non-verbal one.

In Spanish the particle 'eh', sometimes equivalent to the English filler 'uh', was traditionally classified as an interjection and not the object of linguistic study (Gelbes, 2003,

[25] Rodriguez and Torres (2006, p. 349) mentioned that this category, which includes speech repairs and filled pauses was included to allow comparisons with previous studies. The terminology is confusing, they fail to explain what is included in this fifth category that is not in the acoustic events (which also includes filled pauses) and self-repair categories.

p. 12). In more recent grammar books its value has been reclassified and it is defined as a conversational marker (Portoles Lazaro & Martin Zorraquino, 1999) with four different functions in the discourse: to structure the conversation; to request confirmation from the addressee; to indicate the speaker's orientation in reference to a subject; and to indicate the level of certainty the speaker feels towards the content being discussed. Out of the four functions we are only interested in the first one,[26] which is recognized as used strategically in repair (Blas Arroyo, 1995, p. 95).

4.3 Non-verbal Communication

Non-verbal communication (NVC) is thought to express a large part of the message in face to face communication between adults (Mehrabian, 1972, p. 182; Burgoon, Buller & Woodall, 1989, p. 29). Often NVC is associated only with body movements, facial expressions, gaze, hands and body gestures. However, it is much more than that; NVC covers "messages sent through […] the use of time, space, artifacts, dress, and even smell" (Neuliep, 2006, p. 235); all modes of paralinguistic characteristics related to the vocal sounds (volume, intonation, speech, etc.) and even dermal and thermal body changes (Poyatos, n.d., p. 63).

It wasn't until the late 60's that Ekman and Friesen (1969) proposed that non-verbal communication was other than a secondary means of communication. Before them, the study of NVC had been dormant for centuries, although that hadn't always been the case. Confucius, in 500BC, already warned gentlemen against giving different messages with speech and facial gestures; and Greeks, like Aristotle, were well aware of the importance of NV behavior in discussions and speeches. However, it was the Romans, Quintilian in particular, who collected in a series of volumes (90AD/1920, XI, III) the actions for good rhetoric, including those related to NVC, the use of one's robe, sweat, volume of speech, body movements and, of course, hand gestures (Knapp, 2006, p. 4).

With the fall of the Roman empire came the loss of much of their culture, including the art of rhetoric which became the knowledge of clerics from the Christian church, who

[26] In Spanish: 'metadiscursivo conversacional'.

adapted its use to rituals and signs (many still in use today as emblems, like crossing the fingers to ward off evil) (López-Ozieblo, 2013a). These signs are also associated to dress and, together with gestures, were probably the two most obvious representations of affiliation, profession and education and to act in disaccord was considered a breach of etiquette (Schmidtt, 1991). This changed radically in the 70s, when – after the hippy revolution – it became acceptable to dress as one wished and to do what one wanted to do more or less wherever. Today it is much harder to guess social status or any other affiliation from dress or behaviour. One aspect of non-verbal communication that had changed much earlier was that of the use of gestures when talking. During the Renaissance (14 to 17 centuries) together with the development of the arts, an understanding of the divine nature of men (unlike that of animals) led to a dislike of non-controlled actions that showed humans' animal side. Gestures were considered non-controlled and so a movement to suppress them emerged, in part led by Erasmus's work: this fashion spread rapidly thanks to his *On Good Manners* (1530/2003). The suppression of hand gestures when speaking was consciously enforced, with Spain being the paragon of correct behaviour during the 16th and 17th centuries (Burke, 1991, p. 78).

Maybe due to the animal association, gestures laid unstudied for centuries, in addition during the 19th century France led an official ban into the study of gestures, and specifically the use of sign language was banned (Sacks, 1990, p. 122). Thus, it wasn't until the 50s, after the Second World War, that the study of nonverbal communication, and with it gestures, began to develop.[27] However, it didn't help that as the field of linguistics developed some of its greatest thinkers eschewed gestures as not being significant to communication (Chomsky, 1965).

4.3.1 Gestures

The term 'gestures' is often used to refer to movements of any part of the body: eyes, face, head, arms, hands, torso, etc. (Queck et al., 2002). These have been studied in speech, mostly independent of each other; analysing head movements (Jokinen, Nishida & Yamamoto, 2011); gaze (Bavelas & Gerwing, 2011; Bavelas, Coates & Johnson, 2002;

[27] For a more detailed explanation of the study of gestures see Kendon, 2004.

Cummings, 2011; Duncan, 1972) hand gestures (Alibali & Kita, 2010; Duncan, 2006; Kendon 2004; Seyfeddinipur, Kita & Indefrey, 2008). Although it is understood that no single element in the communication event should be taken out of context and analysed alone (Seo & Koshik, 2010, p. 2238), the reality is that it is very difficult and time consuming to look at the whole. This needs to be done through a full microanalysis (as proposed by Bavelas, n.d.) where all individual elements of the communication process are taken into account. This sort of analysis is very useful to reveal information about the psychological relationship between speakers, as shown in the study of a meeting between U.S. Air Force members where gestures singled out one of the participants as being marginalized in the group (McNeill et al., 2010). But gaze, facial expressions and body posture don't need to be associated with speech to have and/or convey meaning; moreover, this meaning often contradicts the speech in that gestures might display emotions not given by the speech. Gestures (except emblems) on the other hand, if performed without speech, would not be understood.

Rossini defines gestures as: "intentional movements of the hands, arms, shoulders and head, occurring within communicative acts, whose lexical access is shared both by the speaker and the receiver" (2012, p. 23). We believe the head and shoulders are included as they are often used instead of the hands, if these are busy, or with them, emphasizing the same message. For the purposes of this study, we would like to add to this definition the clauses that: the communicative act is a verbal one and that without speech the gesture loses meaning. The present study will limit its scope to the observation of finger, hand and arm movements; these will be referred to as 'gestures'.

It is necessary at this point to delve into the nature and the various classifications of gestures (taking Rossini's definition of gesture). Despite the fact that gestures are used throughout the world to communicate, their study in discourse analysis has only developed in the last few decades, lagging behind the study of gaze or facial expressions. No doubt the idiosyncratic properties of gestures and the lack of a visible formal code are at least in part responsible.

4.3.1.1 Types of gestures

We use the term 'gesture' for spontaneous movements made with the hands co-occurring with speech in spontaneous talk. There are also hand movements that are produced without speech and are understood by specific social groups who have codified them, thus giving them a specific meaning, these are referred to as 'emblems' (McNeill, 2005, p. 5). Gestures are used in conversation even when the addressee is not visible (Iverson & Goldin–Meadow, 1997) – although the rate of gesturing seems to be less. Gestures are, like words, not meant to last. They are difficult to repeat and are seldom remembered (Bavelas, 2007, p. 129), being kept only in short-term memory. Neither the speaker nor the addressee is metalinguistically aware of them (Bavelas, 1994, p. 210). This occurs also with words in spontaneous conversation, where the exact utterance is seldom remembered, suggesting a high level of unconsciousness. We believe there is a significant difference between consciousness and control; Lakoff (2013) mentions that 98% of brain activity is unconscious, as consciousness is a linear process and so it would be impossible to be consciously aware of all the brain activity at a single moment – there is just too much of it. However, a process such as speech is being controlled, even if we are not aware of controlling it. The same is believed to be true for gestures. This control results in a certain regularity in patterns that permits their categorization.

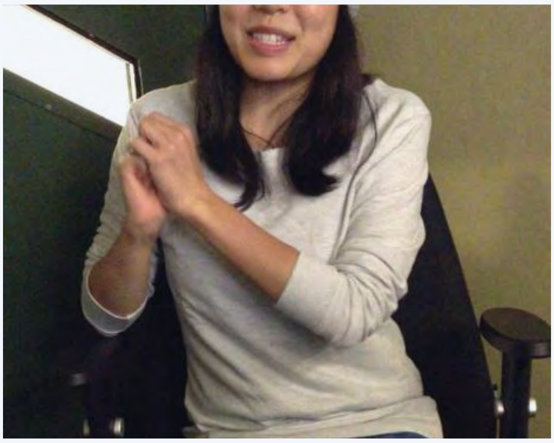
Gesture classification typically refers to the work of McNeill (1992), itself based on prior classifications suggested by Efron (1941) and by Ekman and Friesen (1969), which are based on gesture meaning and function. Gestures are categorized depending on their reference to concrete events they might resemble (iconic) or abstract concepts (metaphoric), orientation and reorientation (deictics) and discontinuities (beats). Although revised nomenclature doesn't refer to types anymore but to dimensions – as many gestures will present more than one dimension: iconicity, metaphoricity, deixis (for pointing but also any significant use of space) and beats, gestures that stress speech or keep the rhythm (McNeill, 2005, pp. 39-40). Other terms used include those proposed by Rauschner, Krauss and Chen (1996) whose research focused on what they term conversational gestures: "unplanned, fluent hand movements that often accompany spontaneous speech" (p. 226). The authors divided gestures into motor or lexical movements.

Lexical gestures, are “more complex, less repetitive, more varied, and of longer duration than motor movements and […] seem related in form to the ideational content of the accompanying speech” (Rauschner, Krauss & Chen, 1996, p.226). In conversation, these are gestures which are synchronous to speech and serve to illustrate the utterance, referring to it or adding independent information – Bavelas, who refers to them as topic gestures, observed them to be 16% of gesture events, (1994, p. 217). Sometimes it is possible to recognise the referent, such as a speaker saying:

She wacked the cat

and holding an imaginary umbrella; others are not so obvious as when the speaker moves an arm away from the body in a straight line tracing the direction of the walk. McNeill (2005) refers to them respectively as gestures with iconicity and metaphoricity.

Figure 4. – Gesture with iconicity, lexical or topic gestures.



Both hands are grabbing an imaginary umbrella which is about to come down on the cat.

Scene from Canary Row (minute 04’48”): Granny hitting Sylvester with an umbrella.

In our corpus we find the following examples (the square brackets [ ] indicate the speech co-occurring with the gesture) when one of our participants recounts the story of *the Lion and the Mouse* (where a lion shows mercy to a mouse which later saves it from hunters):

Figure 5. – Referential gestures (iconic).

C. S. 1HK, written input, line No. 11

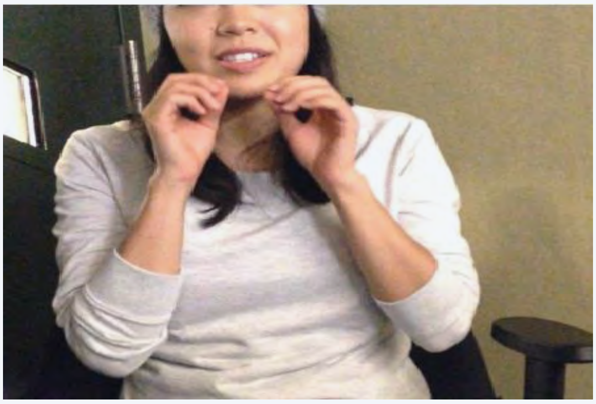
there is [a lion]	and [a mouse]
	
The left hand, and most of the forearm, is raised and dropped. This gesture is repeated to refer to the lion later in the retelling.	*The right hand, and most of the forearm, is raised and dropped. This gesture is repeated to refer to the mouse later in the retelling.*

In these utterances the gestures have taken a referent functionality, the left will represent the lion through most of the story and the right the mouse. The hands do not take on the shape of the lion or the mouse but are used to refer to them (anaphoric reference), the predominant dimension being that of indexicality, (deixis or pointing).

Figure 6. – Iconical gestures (iconic).

C. S. 2HK, written input, line No. 120

[used his small sharp teeth to grind the rope]



Both hands in loose fists are brought to the mouth and quick small up and down movements are used to represent the grinding of the rope.

In this example the gesture resembles a concrete event; that of the mouse grinding the rope with its front teeth, the main dimension seen is that of iconicity.

There are also gestures in spontaneous speech that don't look like their referent (these have often been put aside in research) such as gestures used to seemingly emphasize or stress a point of the speech – most often referred to as batons or beats (McNeill, 2005, pp. 39-40). Rauschner, Krauss and Chen refer to these motor movements as "simple, brief, repetitive, coordinated with the speech prosody, and apparently unrelated in form to the conceptual content of the speech they accompany" (1996, p. 226). See Figure 7.

Figure 7. – Baton, beat or interactive gestures (pragmatic gestures).

C.S. 4HK video L1 line No. 92

to [where] tweety bird lived]]]]]

The right hand circles above the left, no connection to the content.

Bavelas (1994, p. 211) points out that these seem to have another function, apart from giving emphasis, which could be that of involving the addressee. In 80% of the cases (Bavelas, 1994, p. 216) they do not provide any information on the topic. Bavelas calls them "interactive gestures" and has classified them as shown below in Table 1.

Table 1. - Types of interactive (or motor) gestures.

Delivery gestures	Citing gestures	Seeking gestures	Turn gestures
General delivery ("This is my point")	General citing ("As you said earlier")	Seeking help ("How do you say ...?")	Giving the turn ("Now you can speak")
Shared information ("As you know")	Acknowledgement ("I see that you understood me")	Seeking agreement ("what do you think?")	Taking the turn ("I can add to that")
Digression ("By the way")		Seeking following ("Did you get this?")	Turn open ("The floor is open")
			Keeping the turn ("I am still talking")[28]

(Adapted from Bavelas, 1994, p. 213)

Bavelas supports this classification by results from experiments where speakers did not have an addressee who was visible or were asked to produce a monologue, and she shows that in these instances, topic gestures were performed at similar rates as in face to face dyads but the rate of interaction gestures decreased significantly (Bavelas, Chovil & Lawrie, 1992) indicating that the use of interactive gestures varied with the context (whether there was an addressee or not). In a normal face to face conversation these interactive gestures will be from 10% to 20% of conversational gestures (Bavelas, 1994, p. 218).

However, we observe that there are speakers who gesture to aid themselves express the thought, so perhaps an additional category 'cognitive gestures' ought to be added to the above classification. These gestures could be considered 'seeking', with the difference that the speaker is also the addressee, and he or she is seeking not so much words but whole concepts.

[28] Added by author.

In our corpus[29] we find the following example of interactive gestures:

Line No. 131 and then Tweety [of course] saw

Small hand rotation with the right hand corresponding to 'of course.'

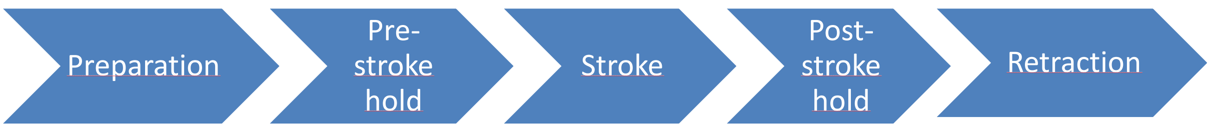
Line No. 131 (continuation) [that and knew that it was him] so

Right hand still up; repeats hand rotation with accent on 'that' then to rest position.

We chose to follow a simple categorization of gestures, based on the categorizations of Bavelas (1994) and Rauschner, Krauss and Chen (1996). Following more current research (Graziano & Gullberg, 2014) we identified just two types of gestures, lexical or topic, referring to them as iconic – as they can be linked to the content of the speech; and motor or interaction gestures, which we will refer to as pragmatic – as they seem to refer to the communicative process itself rather than the content.

4.3.1.2 Gestural phrases

A full analysis of gestures also details their structure as gestural phrases, made up by a number of phases. In many cases the gestural phrase (marked by square brackets) will start from the resting position of the hand, the hand then moves to come into position, the 'preparation phase', and might pause for a moment, the 'pre-stroke hold', before the 'stroke' (marked by an underline below) is performed, this is the nucleus of the gesture where key content, the newsworthy content, will be delivered, after the stroke the hand might pause again, the 'post-stroke hold' and finally the hand returns to the 'resting' position (Kendon, 2004, pp. 110-114; McNeill, 2005, pp. 31-33). Gestural phases might overlap or be realized in such rapid succession that not all the phases are performed or are obvious to the observer. The phases are:



[29] From the video retellings in the L1, participant C.S. 1HK.

Example based on Figure 5:

C. S. 1HK Line No. 11

there is [a lion]

the left hand, and most of the forearm, is raised and dropped

[and a mouse]

the right hand, and most of the forearm, is raised and dropped

In the above example it was possible to view two clear gestural phrases, one performed with the left hand and one with the right, however the second phrase starts before the hand has been brought to a rest in the first one. If we analyse just the right hand we find that the first gestural phrase, indicated by the first set of square brackets, correspond to the phonemes [a lion] and it lasts just 0.5 seconds. Initially the hands are resting on the speaker's lap in the resting position; as the speaker finishes the phoneme /a/ the arm begins to lift up, from the elbow, the palm faces down with the fingers loosely down, the preparation. When the hand reaches just above chest height it pauses for two tenths of a second, the pre-stroke hold, before coming back down palm sideways, thumb towards the body as if loosely holding a rectangular object, this movements corresponds with the phonemes /lī/, the stroke, and the hand is held in that semi-open position, post stroke hold, while the right hand develops its phrase. Eventually both hands return to the resting position on the lap.

McNeill (2012, p. 33) proposes that gestural phrases taken together with speech tend to show synchronicity in the newsworthy elements of the speech, the parts that the speaker considers important or wants to highlight. Newsworthy elements are understood to be those that contain new or salient information.

4.3.1.3 Gesture studies

Initially gestures were studied as tools to express emotions and personality. Birdwhistle (1952, 1970)[30] systematized and classified gestures; and Kendon (1972) deepened the studies by looking at the role of gestures in social interactions. The role of gestures with speech was limited to five functions: redundancy, substitution, to complement the speech, emphasis and contradiction (Ekman & Friesen, 1969). But as the field developed a much more interesting possibility began to emerge: gesture as part of the speech mechanism.

Modern gesture studies share the notion that gestures are part of the communication process but diverge in their consideration of its importance and relationship to speech (Gullberg, de Bot & Volterra, 2010, p. 6). The first division in the field came with studies where gesture was believed to have a communicative function versus those proposing that gesture is just an auxiliary manifestation to speech. De Ruiter (2000, p. 290) believes this distinction to be somewhat arbitrary as it is possible that the speaker intends the gesture as a communicative tool but it might not be perceived as such by the addressee, thus rendering it auxiliary at best. The second branch, believing gesture to be auxiliary, can also be split: Krauss, Chen and Gottesman (2000), advance that certain types of gestures are just an aid to lexical retrieval. Other scholars, such as Alibali, Kita and Young (2000) also believe gestures to be auxiliary to speech, not just for lexical retrieval, but as the visual interpretation of thought, what is known as the Information Package hypothesis (IPH). Nowadays there are enough studies on gesture demonstrating its use for much more than lexical retrieval. Gesture has been linked to turn management and indicating repair (Gullberg, 2008; Duncan, 2006); to refer to elements previously mentioned; to specify the manner of the verb (in languages like Spanish where manner is not inherent in the verb); to downplay unwanted information or stress other elements of the speech (Goldin-Meadow & McNeill, 1999, pp. 164-166). Other functions include the release of cognitive capacities, organize spatial information, thought development in children, or involving the addressee.[31]

[30] Cited by Kendon, 2004, p. 68.

[31] For a more comprehensive list of functions see Gullberg, 2006, p. 8.

Our belief is that the gesture is indeed much more than an aid to speech: it forms a unit with speech. The most obvious example backing up this statement is the different ways in which bilinguals gesture when speaking one language or the other. This is partly due to language differences: some languages, like English and Mandarin, encode information about the mode of the verb (how the action is taking place) within the verb but information about the trajectory, or path is given outside the verb (satellite-framed); others, Spanish, need to add information to the verb to specify the mode but the path is coded within the verb (verb-framed).

In English we can say:

The bird <u>flew</u> out of the room.

In Spanish the sentence will need a series of complements to specify how it left the room:

El pájaro salió de la habitación <u>volando.</u>

The verb 'salió' also indicates the path.[32] This difference is also observable in the gesture that accompanies the utterances, suggesting that the development of the thought is different in the two languages. In both cases we might observe the speaker mimicking the act of flying out, but this is likely to be done with the parts underlined above: the verb, in English; and with the complement in Spanish. The speaker could also give additional information through gestures, indicating a fast flying for instance, a feature not contained in the speech. Gesture, therefore, can be much more than a redundant flourish used by the speaker. It is part of the thought. McNeill and Duncan (2000, p. 142) refer to this quality as 'co-expressivity', the gesture and the speech don't always convey the same aspects of the thought.

[32] For further explanations see Slobin, 1996, Talmy, 2009 and Stam, 2006.

4.3.1.4 Gesture processing mechanisms

Just as scholars disagree on the exact nature of speech production mechanisms the same is true of gesture-speech production models. Various hypotheses have been put forward over the last few decades, the first three hypotheses detailed below follow a modular speech processing where gesture production is a process paralleling that of speech:

- The Lexical Retrieval Hypothesis (Krauss, Chen & Gottesman, 2000) which states that gestures facilitate access to imagistic representations and allow retrieval of lexical items from the mental lexicon. Gesture production is a dual process, the gesture originates outside the speech production process, in the working memory, which feeds into the conceptualizer module and aids lexical retrieval.
- The Image Activation Hypothesis, on which the sketch model is based (de Ruiter, 1998) proposes that gesture facilitates the process of speech generation by keeping an image activated while the formulator module in the speech process encodes it.
- The Information Packaging Hypothesis (Kita, 2000), leading to the interface model (Kita & Özyürek, 2003) has it that gesture facilitates the structuring of information for expression in speech.
- The Growth Point Hypothesis (McNeill & Duncan, 2000) asserts that gesture and speech are one and generated together, they are also the thought.

4.3.1.4.1 The Lexical Retrieval Hypothesis

According to Krauss, Chen and Gottesman (2000) lexical gesturing[33] helps retrieve the lexical information but not to pack it into clauses or to communicate. They also eschew the notion that gestures might signal tension, although they do believe that speech is affected if gesturing is restricted (pp. 264-265). Their gesture-speech model proposes two interactive processes, a speech processor – based on Levelt's model – and a gesture processor. According to their model, working memory holds the representation of the

[33] By lexical gesturing Krauss, Chen and Gottesman, (2000) refer to gestures which co-occur with speech but "are not deictic, symbolic or motor gestures" (p. 264). By focusing on only one type of gestures Krauss, Chen and Gottesman are likely to have discarded valuable data that might have resulted in a more complete model with various processes.

concept; this is transformed into a linguistic proposition which might not incorporate all of the features of the initial concept. Non-propositional elements, such as spatial or dynamic features, are then interpreted into gestures through a motor system.

Lexical retrieval is performed when the kinesic monitor (part of the gesture processor) feeds spatial or dynamic information to the blocked phonological encoder, priming the speech processor by activating words with those spatial or dynamic characteristics out of which the relevant one is retrieved. This model only allows dual speech flow between the elements of the formulator (lexicon, phonological and grammatical encoder). An additional link between the two systems occurs at the articulation phase when the auditory monitor signals the gesture to stop as it hears the relevant lexical element. Krauss, Chen and Gottesman (2000), describe the synchrony between gesture and speech in an experiment performed by Morrel-Samuels and Krauss (1992). The experiment showed that lexical gestures will not start once articulation has begun and that there is a high synchronicity between them. Krauss, Chen and Gottesman add that:

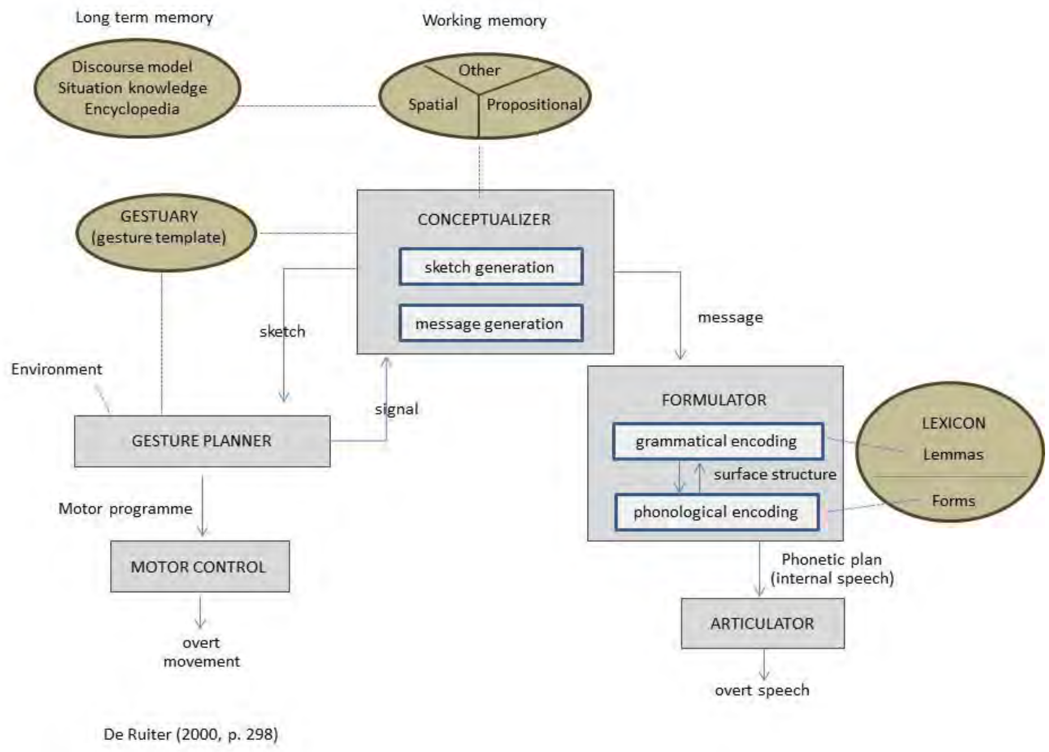
> *We can think of only two ways in which such a temporal relation could exist without interaction between the gesture and speech production systems: (a) gestures of long duration are associated with unfamiliar words; (b) speakers somehow can predict a priori how long it will take them to retrieve a particular word form, and enter this variable into the formula that determines the gesture duration.* (p. 270)

Morrel-Samuels and Krauss go on to say that they do not believe either option is plausible, but they do not spurn the possibility of alternative gesture-speech processes existing (1992, p. 621).

4.3.1.4.2 Image Activation Hypothesis

In the Image Activation Hypothesis the gesture is originated within the speech processor, in the Conceptualizer, but it then follows a parallel path to that of speech. The gesture is communicative and might carry information that the speech doesn't. The Lexical Retrieval and Image Activation Hypothesis assume that the exchange of information is one-directional always: from gestures to speech – gesture is translated into speech (see Figure 8).

Figure 8. – The *Lexical Retrieval* gesture- speech processing system.



4.3.1.4.3 The Information Packaging Hypothesis

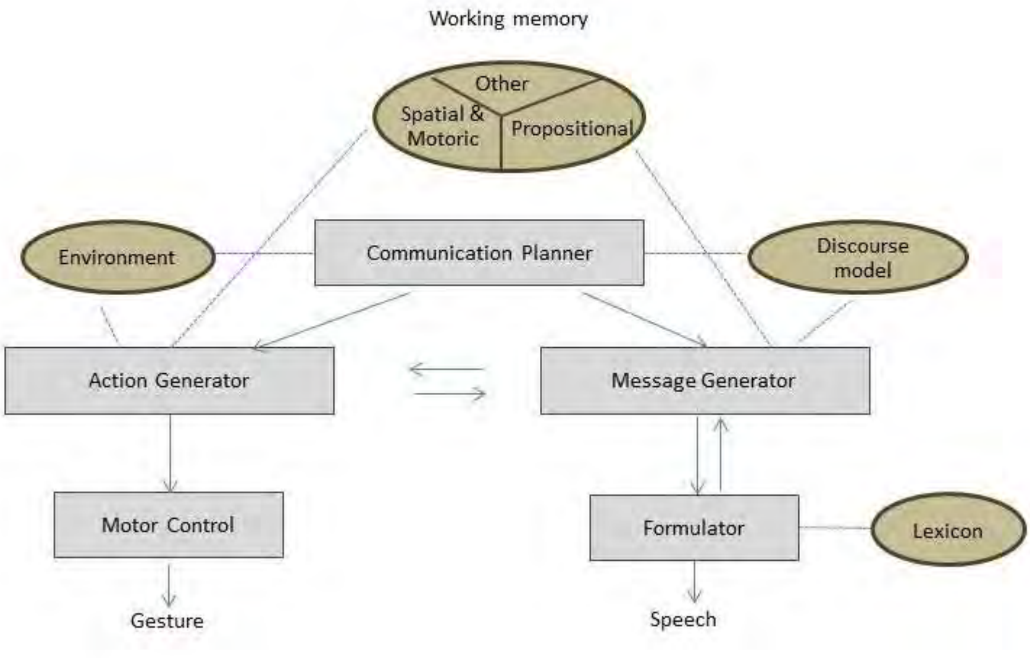
The Information Packaging Hypothesis is somewhat based on Levelt's model in that there is a conceptualizer, a formulator and an articulation module but the conceptualizer is made up of an action and a message generator, the first leading to motor control, the second to speech formulation. Here the gesture can communicate information that has not been encoded in the speech. Kita focuses on representational gestures which he defines, following McNeill's terminology (1992), as iconic and abstract deictic[34] gestures (Kita, 2000, p. 162). He focuses on these because he believes that they must have a cognitive as well as a communicative function. He bases this assertion on the fact that speakers don't suppress all gestures even if they are not being seen by an addressee. Kita's hypothesis is based on the idea that there are two types of thinking: analytical and spatio-motoric. Gestures are related to spatio-motoric thinking, while analytic thinking is the default mode when

[34] Deictic gestures will be included within the iconic category in this study.

speaking. Analytic thinking provides a structure to concepts by creating non-modality specific analytical templates, these can be linguistic or non-linguistic in nature.

Thus spatio-motoric information will be conveyed both by analytic thinking (the default) and the more specialized spatio-motoric thinking. This is possible thanks to two mechanisms, one identifying the degree of matching between the two thinking modes and the other exchanging information between the modes. Thanks to this exchange, the Information Packaging Hypothesis postulates that the gesture is created independently of the speech that accompanies it (see Figure 9).

Figure 9. – The *Information Packaging* gesture- speech processing system.



(Kita & Ozyurek, 2003, p. 30)

Kita puts forward as proof the fact that gesture is used even during silences, mismatches during phases of conceptual development and the use of similar iconic gestures when talking about the same spatio-motoric experience, as shown in McNeill's studies of various cultures (1992) (Kita, 2000, pp. 164-171).

Kita suggests that the two modes of thinking package information into chunks roughly the size of a clause (2000, p. 169). The content of the analytic thinking will be expressed as speech while spatio-motoric content will be expressed as gestures and also as prosody. The steps are as follows:

1. The speaker selects a ‘scene’ (a part of a story).
2. The two modes of thinking formulate the scene so that it can be articulated within a clause; this is done through a back-and-forth exchange of information to ensure coherence. Complex information will be subdivided into shorter ‘scenes’.
3. When a good match is reached, externalization will begin, usually in the two modalities. However, if due to external circumstances the match is not perfect then asynchrony between the modalities might appear.

The IPH suggests that speakers have the full clause ready as externalisation begins, which does not seem to be the case for all speakers in all circumstances (McNeill, personal communication, March 2013).

4.3.1.4.4 Parallel processing frameworks – The Growth Point

The Growth Point (GP) hypothesis on the other hand is not based on the traditional modular speech process, but is redolent of nodular activation theories. It believes that the seed for the idea is found within a nodule, a GP, as this GP develops it activates related GPs to which it is neurologically linked. The most activated GP is the one that will be fully developed and externalized, either as inner or overt speech, through the use of speech and gestures. The two modalities represent the same idea but not necessarily the exact same information. As mentioned above, McNeill believes that the gesture is preferred to highlight newsworthy elements in the utterance, particularly new information.

Gestures co-occurring with speech (not emblematic gestures) are idiosyncratic, global and synthetic. Idiosyncratic because they are unique to the speaker, the context and the utterance; global because they are whole in themselves, not a part of a larger gesture (unlike speech where the parts make the whole); and synthetic as one gesture can combine various meanings given in speech throughout the utterance. In the sentence:

It left the room quickly.

The speech could be accompanied by a hand gesture where the second and third fingers open from a loose fist with a fast sideways movement. This could be a gesture a particular speaker uses in this particular instant, idiosyncratic; it cannot be broken down into parts, so it is global; and it conveys the meaning of 'leaving' and 'fast', thus synthetic. Meanwhile the speech has followed a coded phonological, lexical, semantic and pragmatic system, common to a socio-cultural group, where the different parts combined to form the utterance and each can be analysed into its components.

It could be considered that these two modalities, the gesture and the speech, are opposites in their qualities, and if combined are capable of better presenting the thought. This combination is what McNeill and Duncan (2000, pp. 141-161) refer to as the Growth Point. The GP is the thought which in turn is the gesture-speech unit. The two modalities come together to stabilize the thought, as each modality, by itself can't fully contain it (López-Ozieblo & McNeill, forthcoming, 2015). As the two modalities are the same thought they have to develop together and this should be obvious throughout the utterance and in particular in those times during a conversation when speech ceases, such as in cases of turn taking and self-repair.

Synchronicity has been shown to be disrupted only in cases of meaningful repetition, where it could be argued that there is no thought, and so no GP. Stuttering, clinical or induced by delayed auditory feedback does not disrupt the gesture-speech synchrony (McNeill, 1992, p. 209). What we can conclude is that the speaking process is not just speech, gesture also plays a role.

4.3.2 Gestures and disfluency

There is a dearth of studies that look into the relationship of speech and gesture in self-repair; some exceptions are: de Ruiter, 1998; Seyfeddinipur, 2006; and Seyfeddinipur, Kita and Indefrey, 2008. Although a number of scholars are currently studying various aspects of the disfluency-gesture relationship (Graziano, Church & Brown, Colletta), the consensus of these scholars is that more research is needed in this field (personal communication with the above, July, 2014).

Most studies that look at disfluency in speech together with gesture centre on the relationship of the gesture and its synchronicity with the speech, in an attempt to find out whether the timing of a gestural phrase phase shift could shed some light onto the gesture-speech process (Chui, 2005; Seyfeddinipur, 2006; Seyfeddinipur & Kita, 2001; Seyfeddinipur, Kita & Indefrey, 2008;). Seyfeddinipur and Kita (2001) reported on the synchrony of gestures with the suspension and resumption points in overt repairs. Their findings indicate that gesture suspension precedes speech suspension (p. 463), hinting at the disfluency to come, and that gesture resumption precedes speech. This is seen as support for the Delayed Interruption for Planning Hypothesis, as the speaker chooses when to suspend speech. Seyfeddinipur and Kita suggest that the speaker knows that he or she will be disrupting speech and when, controlling it. Another explanation could be that the gesture, being faster than speech, ends earlier but if the next concept has not been developed yet there is nothing for the gesture-speech to externalize (following the GP hypothesis of gesture-speech being one and a parallel speech production process); yet another explanation would be that there are two parallel mechanisms each with slightly different information, operating at different speeds, with two stop signals, and so one concludes before the other (following the Sketch Model and speech production in series).

Seyfeddinipur, in her 2006 thesis, studied the synchronicity of gesture and speech suspension in German speaking dyads describing a room. She focused on gesture-speech timings during cut-offs. The study was divided into three sets of experiments: a control one to confirm that both gestures and words elicited a response within a similar time-frame; one measuring the time between suspension of speech and gesture; and one measuring the rate of gesture suspension in fluent and disfluent speech. Her results showed that there were no differences in gestural rate in fluent or disfluent utterances but when she refined the above results by dividing speech interruption into within word cut-offs and after word completion cut-offs, she found that speech-gesture synchronicity in after word cut-offs concurred with Seyfeddinipur and Kita’s previous results, gesture stopping before speech, but this was not the case in within word cut-offs.

From these results Seyfeddinipur concluded that there might be two stop signals being activated,[35] as suggested by the Sketch Model, and also "gesture suspension is closer in time to error detection than speech suspension" (p. 143). But she observed a "small but reliable" (p. 144) difference between the gesture-speech stopping times in cut-offs that occur within a word and those where the word was finalized before suspension (there was asynchronicity in gesture-speech in these cases). These observations, she concludes, could be explained by either the Modified Main Interruption Rule hypothesis (Hartsuiker & Kolk, 2001) or the Delayed Interruption for Planning hypothesis. The modified MIR hypothesis states that disruption is immediate upon detection of error, meaning that within word cut-offs the hiatus ought to reflect planning time, so for more complicated repairs the hiatus should be longer. The DIP hypothesis states no such constraints. There were no reliable indications to rule out either hypothesis when considering gesture-speech. The author was looking specifically for gestural actions that would support either the DIP or the MIR, see Table 2:

Table 2. - Gestural actions to support the DIP or MIR.

	Support for DIP / Against MIR
Within word cut-off	Earlier stopping for gesture (speech is not stopped immediately on error detection)
After word cut-off	Larger asynchrony gesture-speech stop (gesture stops first but the speaker goes on to finish the word)

The MIR hypothesis was not supported by results analysing within word cut-offs to repair length intervals. Seyfeddinipur's results showed comparable intervals for both simpler and complicated repairs, corroborating Blackmer and Mitton's (1991) theory and the Delayed Interruption hypothesis that planning starts before the disruption and that the disruption is controlled. This is also supported by the fact that the speaker often changes the articulation of the cut-off sound combining it with another type of disfluency and adding a repair (Fox Tree & Clark, 1997, p. 161). However, Seyfeddinipur also mentioned that the

[35] A control experiment was also carried out to ensure that motor control being faster than speech was not the reason for faster gesture stopping.

DIP predicts that "speech is interrupted as soon as repair processing has come to completion" (p. 142) which doesn't explain the pauses that often follow interruption.

Seyfeddinipur stated that error detection is closer in time in the gesture modality than it is in the speech modality. This led her to conclude that her results contradict the Growth Point Theory, as gesture and speech are not synchronized (2006, p. 143). However, this conclusion does not take into account the context in which the suspension took place, as the exact nature of the utterance might affect the suspension rate. As McNeill points out (personal communication, February 2013) there seems to be temporal segmentations of brain activity built around the lexical and morphological infrastructure of words. This means that the phonological, syntactic and lexical layers might be processed at slightly different speeds; a suspension of a syntactic or lexical nature might be faster than a phonological as lemma selection takes 200 milliseconds to process and the phonological 450 milliseconds (Hagoort & Levelt, 2009, p. 372).

Even taking context into account, the GP theory does not imply that there would be exact synchronicity word to gesture, but that there will be synchronicity of these within the thought. Seyfeddinipur seemed to be questioning the validity of the gesture-speech link in GP because they were not synchronous and were produced at different rates within the same utterance (slower speech did not result in slower gesturing). Seyfeddinipur argues that, were the two modalities integrated as one system then, gesture suspension would mirror speech disfluency, proposing de Ruiter's Image Activation Hypothesis (1998, 2000) as a more likely explanation.

If instead, we take McNeill's explanations (2012) of the package gesture and speech being the thought then we can see how as the thought is unpackaged through different components, say one more spatial, one more temporal (Kendon, 1983, p. 17), the two modalities could happen at different points in time but still be synchronous within that thought-package. They will 'appear' at any point of the thought manifestation, but not necessarily at exactly the same time.

Seyfeddinipur 2006 work was further developed in a paper, Seyfeddinipur, Kita and Indefrey (2008), where the DIP theory was confirmed by showing that that cut-off to repair times were longer in after-word suspensions, and again longer with bigger repairs. These results support the view that speakers prefer to maintain fluency by delaying repair. Also supported by the authors' observation at the end of their paper that speakers do not usually interrupt their speech with "minor phonological errors" (p. 841) or what we described above as slips of the tongue.

It is thought that if the disruption can be controlled and is related to the repair then a suspension mid-word suggests that the repair is ready and speech should resume almost immediately. On the other hand if planning is taking more time the speaker can finish the word and still delay until repair commences, therefore the more complex the repair the longer the hiatus when the reparandum ends in a full word. To keep the floor during longer hiatus the speaker can use expressions like "I mean", "that is" (p. 840). An interesting point is that Hartsuiker, Pickering and De Jong report that stopping is harder if two words are semantically related than if they are just phonetically related (2005).

4.3.2.1 Gesture in disfluency as an indication of lexical retrieval

Rauscher, Krauss and Chen (1996) hypothesized that if lexical gestures were aiding lexical retrieval then their suppression would result in speech disfluency. To test whether lexical movements aided lexical retrieval, the authors asked 41 students, native English speakers (except one), to watch six excerpts of a *Road Runner vs. Wile E Coyote* cartoon and then recount them under 6 varying conditions. Suppression of gesture was done by telling the subjects that hand skin conductance was being monitored by electrodes, which were the excuse used to immobilize their hands. Lexical difficulty was increased by instructing subjects to use obscure words and further by inhibiting words with a specific letter. The analysis of the lexical content proved that the subjects had indeed varied the lexical difficulty as instructed. The results from the six conditions were further subdivided by content, spatial or non-spatial. It seems that this variable was the one that dictated the widest differences. Spatial content speech, under all speech conditions, always had the speaker gesturing over twice as much as non-spatial content. Speakers' speech would slow down if they could not gesture if the content was spatial, but they spoke faster than under control circumstances if it was non-spatial. When the lexical content was made harder disfluency increased (both with and without gesture suppression), in particular the authors noted an increase in intraclausal filled pauses (those pauses filled with an 'eh' or 'em'[36] occurring within a grammatical clause). They put forward these results as confirmation of the link between hand gestures and speech production process and proposed two explanations: either gestures are aiding lexical search or gesture suppression is affecting the conceptualizer stage, although this latter explanation seems less likely as intraclausal pauses (those more numerous with gesture suppression and lexical difficulty) are specifically related to lexical searchers (Rauscher et al., 1996).

Rauscher and his collaborators suggested that suppressing the gesture prevented the speaker from thinking while de Ruiter (2007) and Melinger and Levelt (2004) proposed something radically different, a trade-off relationship between gesture and speech in terms of communicative content. This hypothesis states that the harder speech becomes the higher the chances of gestures taking over the communicative function. Similarly if gesturing becomes harder then there will be more reliance on speech. This hypothesis was based on the work carried out by Bangerter (2004) and Melinger and Levelt (2004) amongst others, where gesture was restricted by immobilizing the hands, or speech was made harder by restricting the use of certain words. However it is interesting to note that it contradicts the results of research done by de Ruiter himself (1998), with no difference in gesture per word rate based on the difficulty of the picture being described; or research done by Rauscher et al. (1996), where the harder the speech the less time spent gesturing.

Perhaps due to these contradictions de Ruiter, Bangerter and Dings (2012) set off to provide more evidence for or against the trade-off hypothesis. They tested the use of iconic and deictic gestures with speech between dyads (a speaker and an addressee) sharing visual input, describing an object both could see, and under no visibility conditions where the addressee had to identify the object referred to by the speaker. By providing the reference there was no question of issues with recall or speech rate, as mentioned by the authors of the study (p. 234). We would like to add that their approach also eliminated turn management issues. There were 48 dyads, all Dutch speakers, who had to identify figures in 42 trials. Iconic gestures were coded depending on their need to disambiguate information; if a gesture was necessary because the speech was unclear, that was considered an

[36] 'Uh' and 'um' respectively, Clark and Fox Tree (2002).

obligatory (or demonstrative) gesture, otherwise it would be considered non-obligatory. It was expected, to support the trade-off hypothesis, which proposes that the rate of gestures per word would be higher with the figures that were harder to describe; however, this was not the case. Overall if the interlocutors could not see each other no deictic and very few obligatory iconic gestures were used, but the rate of non-obligatory iconics remained stable. From these results de Ruiter, Bangerter and Dings (2012) concluded, supporting work by Bavelas, Gerwing, Sutton and Prevost (2008), that deictic and obligatory iconics – demonstrative gestures – are done for the interlocutor and so communicatively motivated, while beats (Alibali, Kita & Young, 2000) and non-obligatory iconics are not communicatively motivated (p. 243).

De Ruiter, Bangerter and Dings (2012) do not believe that non-obligatory iconic gestures are to reduce the cognitive load in the speaker, as then there would be higher gesture rates with higher cognitive loads, which was not the case in their study (p. 244). They do note that their study increased the formulation of cognitive load, not memory, which other authors have found to affect the rate of iconic gestures (de Ruiter, 1998, Morsella & Krauss, 2004).

4.3.3 Neurological facts supporting the gesture-speech argument

Traditionally speech and gestures were regarded respectively as the manifestations of the brain (thinking) and the body (moving), thus two separate entities, as according to Cartesian theories. This meant that their study was never done together until the cognitivist movement advocated for an integrated approach, proposing that the two entities are but one (Willems & Hagoort, 2007, pp. 278-279).

Neuroscience confirms that communication also takes place through our vision, by way of mirror neurons that trigger similar sensations in an observer as those experienced by an observed subject. Mirror neurons were discovered in primates, and are believed to help primates learn by imitation (Rizzolatti & Craighero, 2004). These are neurons that become activated when an individual observes another's actions, thus 'mirroring' the activation of the neurons in the individual performing the action. Mirror neurons "fire not only when the monkey executes a specific range of grasps but also when the monkey observes a human or other monkey execute a more-or-less similar grasp" (Arbib, 2012, p. 27). Mirror neurons are presumed to exist in humans, although their existence has not been directly confirmed. It is probably thanks to these that humans developed a feeling for others, a *Theory of Mind* that allowed us to understand others, what they feel, see or know (Corballis, 2011, p. 18), to cooperate (p. 91), and eventually to develop language.

Originally it was thought that different parts of the brain had specific functions (the field of phrenology studied this, a very popular field in the early 18th century). Similarly it was proposed that the various parts of speech production were processed by different parts of the cortex and they all came together as in an assembly line in utterances. Since the 60's, the approach to language processing research is divided into the following component parts: phonology, syntax and semantics, often overemphasizing this linguistic organisation (Banich & Compton, 2011, p. 231; Ramachandran, 2012, p. 260). However modern brain imaging technologies have allowed for the mapping of the location of these functions, showing clear overlaps. Phonology refers to the representation of sounds in speech, phonemics and phonetics (which is also related with the place of articulation and voicing – part of the motor system); syntax refers to the grammar; and semantics to the meaning. These three systems can, and are, independently affected in aphasics (persons with speech difficulties as a consequence of a lesion), proving that, although they need to work together in order to produce comprehensive speech, they are separate parts of one system. According to the level of activation detected, these three parts are usually processed in the left hemisphere – at least in the majority of right handed individuals (Armstrong, 1999, p. 116), and so it is common to refer to speech as being processed in the left hemisphere. These studies are a clear example of the lack of importance traditionally placed on the non-verbal (in the sense of "word" related) elements of speech, such as prosody.

Prosody relates to volume, rhythm, stress, pitch, intonation, timbre and tone (Poyatos, 1992, p. 4) of speech, and it is considered important, amongst other things, in the management of turns in conversations and to resolve disruptions (Benus, Gravano & Hirschberg, 2011, p. 3002) but it should also be considered as an essential part of language, as it can completely change the pragmatic meaning of the utterance.

In the following Spanish example:

Vamos a casa – with a flat intonation would mean: We go home / We are going home.

¿Vamos a casa? – with a raised intonation at the end would mean: Are we going home?

¡Vamos a casa! – with a raised intonation at the beginning would mean: Let's go home!

In addition, should we place a pause after 'vamos' stressing 'casa' it would sound like an order: 'Go home'.

The processing of prosody normally takes place in the right hemisphere. Lesions to this hemisphere, the 'minor hemisphere', used not to be regarded as important as those to the left (Sacks, 1985) and yet their disruption to communication can be as significant as that of lesions impacting the syntax or semantics processing areas. A native speaker of the language will infuse different pragmatic meanings into the above sentences not only through the use of prosody but also of other non-verbal acts. In particular there will be arm and hand movements, head tilting and gaze changes to emphasize the various meanings.[37]

It has been said that the three main components of speech comprehension and production show an increased activation of specific areas of the brain (Feyereisen, 2003, pp. 220-221), and yet a meta-study of 129 language studies (Fedorenko & Kanwisher, 2009) showed that there is much overlap in these brain parts. The phonological encoding is mostly carried out in the left inferior frontal cortex (posterior areas of the premotor region including area 44, part of Broca's area) and the superior temporal gyrus. Semantic

[37] This was shown in a small experiment during a workshop carried out by the authors at Hong Kong University, 16th January 2013. Natives of Spanish were asked to reproduce the above sentences (without priming them onto the use of NVC) giving them different pragmatic meanings. Most participants used the non-verbal gestures mentioned in the text.

processing takes place in two different places within the temporal lobe: a dorsal part (also including part of Broca's area) and a ventral (part of Wernicke's area – an area initially associated with speech comprehension, where lesions would affect semantics but not syntax). And the processing of syntax is associated with posterior frontal regions (also some parts of Broca's area, areas 44 and 45) and dorsal temporal regions (Vigneau et al., 2006). The overlaps mentioned by Fedorenko and Kanwisher (2009, p. 19) cover most of these areas, they are explained as being the result of not taking into account individual biological differences between subjects. Therefore we can assume the areas activated in each process are either much larger than specific studies would suggest or, an alternative view, that the overlap between the distinctive parts might not be so great. This view could be supported by studies that tried to activate syntactic processing areas, and instead activated semantic responses (Banich & Compton, 2011, p. 247).

Speech articulation, on the other hand, takes place not in Broca's area as was originally thought, but in the area posterior to Broca in corticol motor and premotor areas within the precentral gyrus (Josse & Tzourio-Mazoyer, 2003, p.228). These areas, motor and premotor cortex, are activated with motor tasks such as manual activity, or body gestures (Kircher et al., 2008) and also when just listening or even thinking about words (Willems & Hagoort, 2007, p. 280).

Most neurological studies of gesture with speech to date refer to gestures which are iconic. They describe some part or characteristic of the idea being uttered (Decety & Grèzes, 1999; Kircher et al., 2008; Skipper, Goldin-Meadow, Nusbaum & Small, 2007). Other research just looks at emblems, gestures which can be understood by a specific culture without speech. Some of them look into the connectivity and activation of different brain areas as the speaker produces or processes speech and iconic or emblematic gestures related (or not) to that speech (Andric, Solodkin, Buccino, Goldin-Meadow, Rizzolatti & Small, 2013). It is clear that the information given by these gestures helps to process the message, with less semantic control (less activation is shown in the relevant areas with gesture-speech compared to situations when speech occurs alone). However, emblems – gestures used with speech – seem to activate different areas of the brain than speech does (areas related to observation rather than to language processing) and when used noncongruously with speech, affect normal comprehension and processing (Willems & Hagoort, 2007, p. 285). In addition, research into users of sign language, mostly American Sign Language (ASL), indicates that similar neuronal networks are activated by processing sign language and spoken language (Willems & Hagoort, 2007, p. 283).

If a gesture is, as many scholars agree, an important element of language, then there should be a need to understand how it is neurologically processed. There is evidence suggesting a connection between the hand area of the motor cortex and the language areas, with increased activation of the hand motor area when speech occurs (Meister et al., 2003, p. 406). Furthermore, Kelly, Kravitz and Hopkins (2004) also proved that there is a strong relationship between gesture and speech processing. Weisberg and Emmorey (2014) also confirm a high level of integration between the gesture and speech, supporting the idea that the two systems work in parallel as suggested by McNeill (1992). Their study, carried out with the aid of electroencephalography (EEG), showed that speech was processed differently depending on the types of gesture (matching, complimenting the speech, referring to an object different from that of the speech, or no gesture).

Decety and Grèzes (1997) used PET scans to map what regions of the brain were associated with the observation of meaningful actions (opening a bottle, sewing a button) and meaningless ones (taken from American Sign Language, which was not known to the subjects). The results showed that there was a significant difference in the regions activated, depending on the properties of the action (meaningful or not) and also on the nature of the executive processing of the action (whether it had to be imitated or not). When imitation was required, regions in both hemispheres were used. The authors suggest that the left prefrontal region might be in charge of generating responses in relation to semantic cues, whilst the right could be related to memory. If recognition was necessary and the action was meaningful, then the left inferior frontal gyrus was mostly involved, with meaningless actions (what would be considered “gestures”) the right hemisphere was the most responsive (Decety & Grèzes, 1997, p. 171- 175).

Although not much is known about the exact neural workings of speech and gestures together, it is believed that the processing of the integration of speech and gesture might take place somewhere other than the processing of just speech or gesture, perhaps in the left posterior temporal lobe (Kircher et al., 2008, p. 175). As it has been mentioned above, there is enough evidence to believe that there are changes in the level of activation of brain areas related to speech depending on whether there are accompanying gestures or not (Kircher et al., 2008; Skipper et al., 2007; Kelly, Kravitz and Hopkins, 2004). These studies with functional magnetic resonance imaging (fMRI), also show how with congruous speech and gestures there is a decrease in the level of activation to areas activated with speech alone (the left superior temporal gyrus and the left posterior insula). This agrees with the results found of Wilson and Iacoboni (2006)[38] who found that brain activation (of premotor cortex areas) was higher for non-native phonemes, compared to native ones. This suggests that a lower cognitive load results in a weaker neurological activation. That is, easier input such as conveying the same message through speech and gesture, is easier to process.

So far, the evidence seems to go against the alternative view of the relationship between speech and gesture advocated by Krauss, Chen and Gottesman (2000) who believe, at best, in a supplementary role of gesture to aid speech processing. They claim that there is insufficient evidence to confirm that all gestures accompanying speech are communicative in nature (p. 273). This would mean that speech is processed separately from gesture, which is not always the case (Kelly, Kravitz & Hopkins, 2004).

Even if not all studies provide irrefutable evidence that speech is related to the motor system, the motor theory of speech (Galantucci, Fowler & Turbey)[39] proposes that speech is linked to the actions producing them, activating motor and premotor cortices. The speech-gesture process works by “mapping the speech input onto a motor representation” (Willems & Hagoort, 2007, p. 279), that representation is perceived as speech, which is also how the speech would be produced. This certainly works for speech describing actions where the word “kick” activates the same brain area as watching or performing the action itself; however it is still not clear whether this is related to the integration of speech and gesture or just the influence of an iconic motor representation. An alternative theory to the motor theory of speech proposed by Willems and Hagoort (2007, p. 286) is that of a balance

[38] Cited by Willems & Hagoort, 2007, p. 280.

[39] Cited by Scott, McGettigan & Eisner, 2009, pp. 295-296.

between modalities. Neural networks form as needed, with a great deal of flexibility existing thanks to functional overlaps between networks.

We take all of the above as evidence to support the idea that gesture and speech are not two separate modes that come together in production but are something that exists and develops together, a unified form, indivisible in the thought (as it would be it, if it *is* the thought) but expressed in the world through two modes.

4.4 Second Language Acquisition

The study of first (L1), second (L2) and foreign languages (FL), began with the analysis of its individual elements (lexical, syntactic) and the conclusions applied in the FL classroom with structural teaching methods, such as the grammar-translation and the audio-lingual. Eventually language study developed into a holistic study of the structures and processes at discourse and interaction level (Gullberg, de Bot & Volterra, 2010, p. 3) — applied through methods where the objective was to ensure successful social interactions, emphasizing a communicative approach. Communication, in any language, is multimodal, with verbal as well as non-verbal elements contributing to the success of the exchange. Raffler-Engel highlighted the importance of NVC by stating that "communication is multi-channelled and to reduce language to the sole channel of verbalization is not communicating in full" (1980, p. 229). And yet the non-verbal element is one often forgotten or put aside in the FL classroom.

An important consideration to take into account in linguistic research is that the communication action is affected not only by the speakers' proficiency level but also by who they are, their reasons for communicating, the topic and phase in the conversation (Bavelas & Chovil, 2006, p. 101), as well as other elements that include cultural identity, which are reflected in the interaction, and can be used by the participants to profile each other (Jefferson, 1974, p. 183). Egbert (2004, p. 1495) cautions that second language conversational analysis research should first take into account the individual's linguistic, regional or ethnic attachments and establish how they are used in the interactions, before assuming they are relevant in the analysis, as they can be projected into all aspects of the interaction, including repair. Gullberg (2010, p. 83) further recommends that when studying the behaviour of any other than native language learners the researcher is to be aware of native, target and any other languages known by the speaker as the interlanguage of the learner is going to be conditioned by all these in addition to his or her behaviour and learning mechanisms.

This contrasts with the thoughts of Schegloff (in interview with Wong & Olsher, 2000, p. 125) who advocates that too much thought is given to the non-native status of speakers in second language research, assuming before analysing the data, that this is a significant factor, instead of letting the data speak for itself. But this does not mean that these points have to be taken as relevant in the analysis (Seo & Koshik, 2010); they could be offered as explanations for the results but should not tint the methodology.

Ellis (2005) identified ten principles to serve as the “basis for language teacher education” (p. 292). The first one was to ensure that instruction results in learners’ rule-based competence and knowledge of a range of formulaic expressions. Learners begin by internalizing set expressions, and generally by focusing on the acquisition of vocabulary. Eventually they are able to analyse the structure of these set expressions and focus on the form as well. Norris and Ortega (2001) in a metastudy of 49 studies concluded that “explicit modes of instruction are more effective than implicit types” (p. 249). These terms usually refer to grammar teaching. Implicit instruction avoids grammar instruction, believing learners will pick it up naturally, while explicit instruction, including deductive and inductive methods, does specifically address it. If explicit grammar methods work better it is likely that this is the case in other areas of instruction. If in disfluency there exist a number of set expressions specific to native speakers, teachers should be able to transfer them to learners through explicating instruction. As gesturing like a native gives the native speaker an impression of fluency (Neu, 1990), teaching how to gesture perhaps also ought to be included in the curriculum. While there is no conclusive literature on the success of teaching gestures to L2 learners, casual personal observations in the classroom indicate that after awareness raising and explicit instruction as to the use of both gestures and emblems, students did make a conscious effort to use them more.[40]

4.4.1 Repair in second language acquisition

In Second Language Acquisition (SLA) error has been the topic of many a discussion, and it is still a subject that ignites passions.[41] Surprisingly its counterpart, correction, or repair, and its treatment, have been of little interest to second language scholars. SLA usually takes the view that any repair is going to be the consequence of linguistic difficulties (usually grammar or vocabulary). However, as among native speakers, there are more reasons for repair than just linguistic difficulties. Wong (2000, p. 264) in her studies of conversations between native and non-native (NN) speakers found that causes for repair (other initiated repair) ranged from redundancy reduction, to lack of understanding of idiomatic phrases, or of the ability to make inferences or to deal with discourse conventions. It is very likely that NNs might have characteristic interlanguage interactions which they will keep until their linguistic competence allows them to start copying native interactions. (Wong, 2000, p. 262).

4.4.2 Gesture in second language acquisition

In terms of reception, the gestures of others (including, of course, the teacher's) may help make input comprehensible by, for example, speech parsing, helping learners find the key words. They may also help link language and cognition by activating mirror neurons: seeing *you* gesture makes me feel as if *I'm* gesturing, and hence I'm connected to the thinking that motivated the gesture. But apart from the communicative element of gestures, we should also consider their impact on the psychology of the learner. New directions in

[40] In September 2013, López-Ozieblo conducted an experiment with a group of 11 students of Spanish as a FL at the Hong Kong Polytechnic University. These students were coached on the differences in non-verbal communication between natives of their culture (Hong Kong Chinese) and the language they were learning (Spanish from Spain), this included 5 hours of targeted content using videos, cartoons and photos. The results were a visible increase of gestures, both emblematic and non-emblematic, and attempts, although not as successful, to use face gestures and prosody.

[41] For a review see Rezaei, Mozaffari & Hatef, 2011.

language studies propose that language should also be evaluated in its relation to inner psychological processes (Gullberg & McCafferty, 2008). Goldin-Meadow (2009, 2010) believes that gesture can be used to influence the way people think. In a number of studies with children she has shown how teacher's gestures (on explaining mathematical equations) can help sharpen the cognitive process that will lead to the right solution. Gestures can also be used to create the right learning environment in the classroom, through what Mehrabian termed 'immediacy' (1972), referring to the perceived degree of physical or psychological distance between two (or more people). There have been a number of studies looking at teachers' gestures and their impact on students learning (in various subjects, not just SLA), such as Andersen, Andersen and Jensen (1979), McCroskey and McCroskey (2006), Tellier (2010), all showing a strong positive correlation between teachers' gestures and students performance.

However our interest is, not on the teacher, but on the student. It is often assumed that speakers of a foreign language will use gestures when not able to communicate, but the reality is that this isn't always the case. Students often forget that they can use gesture as a communicative resource.[42] Gesture in students has been shown to help recall speech learned (Cook, Yip & Goldin-Meadow, 2010); but it is used by students for much more than just lexical retrieval (Krauss, Chen & Gottesman, 2000) or to solve difficulties in production (Feyereisen, 1987). Gesture has also been linked to turn management (Duncan, 1972); to repair (Gullberg, 2008); to refer to elements previously mentioned (Gullberg, 1999, 2003); to specify the manner of the verb (in languages like Spanish where manner is not inherent in the verb) (Stam, 2006); to downplay unwanted information or stress other elements of the speech (Goldin-Meadow & McNeill, 1999, pp. 164-166); to introduce new referents (So, Demir & Goldin-Meadow, 2010); to organize spatial information (Kita, 2000); to free cognitive resources (Goldin-Meadow, 2001; Goldin-Meadow, Nusbaum, Kelly & Wagner, 2001); to involve the addressee (De Fornel, 1992).[43]

[42] From personal observations in Spanish as a Foreign Language classrooms in Hong Kong tertiary institutions.

[43] For a more comprehensive list of gesture functionality, see Gullberg (2006, p. 8).

Thus the learner's own gestures may also play an important role in language learning. It is generally accepted that any kind of learning task is aided when the learner can off-load the cognitive effort involved on to an external representation. Hence learners will gesture a lot when doing a speaking task, even when situated behind a screen and so not seen. "It is possible that L2 learners' gestures reflect their attempts to reduce the processing load of keeping words, grammar, and the relationships between entities in mind at the same time as planning what to say next. In this sense, gestures may help learners to keep talking" (Gullberg, 2008, p. 293). Moreover, gesturing while learning seems to improve recall lexis. And, very importantly, gestures help build rapport and confer on their users the status of a legitimate interlocutor. "Learners who are seen to gesture are often more positively evaluated on proficiency than those who are not" (Gullberg, 2008, p. 293). Duncan (2005) stresses the importance of re-aligning gestures to speech in the L2 in order to sound native-like. Neu (1990) points out that fluency in a non-native is perceived as being higher if the speaker is also using appropriate and synchronized gestures when talking (1990). These observations suggest that gestures should be part of second language teaching, especially when the student has not got access to a naturalistic L2 environment.

After the traditional grammar-based structural methods were expurgated the functional methods gave rise to the interactive methods, such as the popular communicative approach, incorporating a more complex level, not just vocabulary and grammar, of discourse and interaction. In interactive methods text-books and teachers strive to offer students real samples of the foreign language, with much attention being paid to context, register and paralinguistic elements (or nonverbal elements). But except for hand gestures expressing emblems, usually little thought is given to other types of gestures. Even within the multimodality field, where there is an enhanced awareness of different modalities used in communication – written text, pictures, layouts, verbal speech, lighting, films – gestures are studied by a fraction of scholars researching any of the above fields. Nevertheless we second Gullberg in that gestures ought to be seen as a resource in learning and a component of language proficiency (2010).

How students of a foreign language gesture in their second language compared to their gestures in the mother tongue seems to vary by language and proficiency. Gullberg

(1998) stated that L2 speakers tend to produce less iconic gestures in the L2 than in their mother tongue. And yet Moroccan and Japanese native speakers learning French have been reported to use more representational gestures in their L2 than their L1 (Kida, 2008). Özyürek (2000) found that Turkish natives speaking English as their L2, gestured like in Turkish with manner verbs. Choi and Lantolf (2008) and Stam (2006) found that their Korean and Spanish students (respectively) speak English using gestures as in their L1 with path verbs in the L2; findings contradicted by Kellerman and van Hoof's results (2003), and Negueruela, Lantolf, Jordan and Gelabert's (2004). Although we suspect that the variables are so many that conclusive evidence will be always hard to find (variables might include gender, age, proficiency level, manner in which the language has been learned, the teacher, mother tongue and transfer, other languages spoken, content of the discourse analysed, type of gesture, type of input leading to the discourse, etc.).

What is recognized is that even fluent speakers of a foreign language might still show traces of their L1 in their gestures (different form and timing, path and manner, Özyürek, Kita, Allen, Furman & Brown, 2005); and beginners might use different gestures, to those of their L1, to compensate for speech deficiencies (Negueruela, Lantolf, Jordan & Gelabert, 2004), showing a dissociation between gesture and speech (Stam, 2006).

4.4.3 Disfluency and gesture in second language acquisition

Not much research has been done in the combined field of disfluency and gesture in second language acquisition. The first work on gesture and second language acquisition (not even incorporating disfluency) is Gullberg's 1998 dissertation. The only current work we are aware of that covers disfluency, gesture and second language acquisition is by Graziano and Gullberg (2014). Although theirs is an on-going wider study comparing Italian and Dutch children and adult monolinguals (11 under each category looked at – including three different child age groups), it also includes eleven Dutch natives learners of A1 French. As from July 2014 they had not classified their data by type of disfluency but were able to show a significant number of findings regarding disfluency and speech. Their disfluencies categories included filled and unfilled pauses, interruptions, lengthenings and combinations of the three when narrating a video cartoon (*Pingu*) or a paper based cartoon. The gestures were analysed by function: referential versus pragmatic (as per Kendon's 2004 definitions) and were also coded for completion versus suspension with the disfluency. Suspension included interruptions, preparations and holds. Their findings indicate that less than 20% of gestures occur during the disfluency (over 80% take place with fluent speech) across all groups, except for the L2 speakers who perform about 30% of their total gestures during the disfluency. The authors indicate that this increase is mostly due to an increase of pragmatic gestures, often used to retrieve lexicon from the addressee. After the age of 9 all groups tend to suspend, rather than complete, their gestures during the disfluency, but the nature of the completed gestures varies by group (Italian adults performing more pragmatic gestures than Dutch adults or L2 learners). Their work so far supports the close link between gesture and speech and also challenges the lexical retrieval hypothesis of gestures being of a compensatory nature when speech fails, although in some cases gestures can compensate for speech conflicts. One immediate apparent drawback of the data so far, is that it does not present gesture by type of disfluency, clustering all disfluencies together, and not accounting for different types of repair. In addition there is no information as to the rate of gestures (by number of words, clauses, etc.) or rate of disfluency.[44]

Probably the work most relevant to our study done to-date is that of Stam (2012) who describes gesture-speech performance in ten adult Spanish (Mexican) speakers, learners of English as a SL in the USA. Five subjects had an intermediate level and five advanced.[45] Subjects were asked to retell extracts from the *Canary Row* cartoon, specifically from were Sylvester ascends the drain pipe to him rolling to the bowling alley.[46] Disfluencies were classified into interruption and repetition, repair, filled pause and lengthening, silent pause, pause with breath intake, as well as other sounds such as swallowing or laughter. A wide range of gestures was also classified following McNeill's categories: iconics, metaphorics, batons, deictics and Butterworths (gestures made when trying to recall a word, McNeill, 1992, p. 76). In addition the transcription included the number of lexical searches occurrences, and their outputs (whether the correct word was found or not).

[44] This point has been addressed with Graziano and it is hoped that a future collaboration might provide this data.

[45] Although exact equivalencies are difficult to make, an ESOL intermediate level corresponds to a B1 CEFR level and advanced to B2, C1 level (Consejo de Europa, 2002).

[46] See Appendix 1 for a description of the cartoon.

The results are divided by level, intermediate and advanced learners, and indicate that when there is gesture with the lexical search, be this successful or not, the gesture is synchronous with speech or a combination of speech and pause (seldom with pause alone). In addition it links the type of gesture preformed to the success of the lexical search, showing that in successful lexical retrieval the highest percentage of gestures used (during successful and unsuccessful retrieval) are of an iconic nature (50% in intermediate learners and 82% in advanced learners in the successful case and 77.8% and 88.9% respectively in the unsuccessful case) but during the lexical search no preference as to the type of gesture is obvious, suggesting that the mental iconic representation alone doesn't necessarily help with retrieval, and that it is the movement itself that helps. The data indicates a low percentage of aborted gestures 9.1% (1 case) in the intermediate level and 7.7 % (1 case) in the advanced level when the search is successful and a higher rate 25% (3 cases) in the intermediate level and 33.3 % (2 cases) in the advanced level. Stam also confirms that the more the addressee responds to the lexical retrieval gestures the more likely is the speaker to use them. She concedes that lexical retrieval is just one of the functions of gestures, but points out that not enough research is available in this field. Indeed, the instances of gestures with the disfluencies and repairs obtained in this study are relatively low, and when further categorized by proficiency level and type of gesture the significance of the analysis from a statistical point of view is reduced considerably (often with one instance in some categories). However, some of the data looks very interesting, in particular the possible link between aborted gestures and search failure. Unfortunately the data is not analysed by instances of disfluency-repair so it is not clear whether there is a link in the type of gesture with the disfluency and the one with the repair.

Another drawback is that the speech categorization includes disfluency and repair, though some disfluencies are clumped together, such as filled pauses with lengthenings. Gesture categorization is highly detailed, raising the issue of its accuracy, as gesture interpretation is highly subjective. This is a point which is often made and an on-going discussion in the field. In recent communication with Kendon (July, 2014) he raised this issue, suggesting the need for a repository of analysed data researchers could refer to when classifying gestures. The reality is that there are few purely one category gestures and their intent is highly subjective. However this work followed on from data collected for research on *Thinking for Speaking* (Stam, 2006), to identify gestures with path and manner verbs, therefore the gesture classification is likely to have been very accurate, especially for iconic gesture coding.

A number of recent studies presented at the last ISGS conference (San Diego, 2014) have reported results on disfluent speech with gesture when covering other topics, such as cognitive loads in monolingual and bilingual speakers (Brown, Quiros, Lemon, Aspuerez, Church and Mahootian, 2014) and speech development in French children (Colletta, Pellenq & Cefidekhanie, 2014).

5 DESCRIPTION OF THE RESEARCH

The objectives of the study were to confirm or reject the hypothesis of a possible relationship between gesture and speech in disfluency. As our specific interest is the acquisition of Spanish by English speakers, our participants were Spanish native speakers and HK students of Spanish as a foreign language.[47] From informal observations we had confirmed that the first group performed disfluencies and gestures in their speech but it was necessary to verify that HK students also performed them. Thus the first step was to carry out a preliminary study, using an existing corpus with data from recordings of HK students (López-Ozieblo, 2012), to confirm their production of disfluencies with gestures in Spanish. Once this was confirmed, we tried to find a relationship between the disfluency and the gesture in the corpus we had created. No patterns were found, probably due to the small size of the sample and to having too many variables. The next step was to obtain a larger corpus, reducing the number of independent variables (those we could control), and including Spanish native speakers and more HK students. The participants were given different types of inputs (stories) and were asked to retell them (hereafter 'retellings'). Both sets of participants were asked to speak in their L1 (Spanish or English) and their foreign language (English or Spanish). It is important to note that HK students' L1 is Cantonese, but all our participants spoke fluent English, having been raised in a bilingual environment.[48] We thus considered them native in their use of English and will refer to English as their L1 throughout the study.

Once the data was collected a thorough analysis was carried out in order to confirm or reject the various hypotheses we had formulated. The results were finally compared to those available from past studies to provide a meaningful discussion. This chapter begins with a description of the research paradigm and a justification for the methodology adopted. As the main study was specifically designed for this research and benefitted from

[47] For ease of reading and consistency we will refer to the two languages studied as the L1, the first language (Spanish in native Spanish speakers and English in Hong Kong students) and the L2, second language (English in native Spanish speakers and Spanish in Hong Kong students).

[48] Participants had scored 8 or 9, out of 9, in the International English Language Testing System (IELTS).

the weaknesses of the preliminary study, it was considered more logical to first provide a brief description of the preliminary study, methodology, analysis and results, and then follow it with a detailed account of the main study.[49]

5.1 Research

Magnusson (2006, 2000) believes that the confirmation of hypotheses or theories through observation of pattern forming events is key in any science. Finding these patterns (in astronomy, biology, physics, etc.) did not always help to understand them at the time, and it was only decades later that their meaning became clear, such as the relationship between the orbiting of the Earth and gravity. In the case of human communication, its study and evolution, we find theories from a number of fields (anthropology, psychology, ethnology, philosophy, linguistics, neuroscience, biology, etc.) that evolve as knowledge is gathered and understood. Theories on the use of gestures are still evolving, and there is still a need to find and classify the patterns that will help understand it (Kendon, personal communication, July, 2014).

It took at least 50,000 years from the birth of speech to formalise its grammatical rules[50] – the first Spanish grammar was not published until 1492 – and even today we find grammatical issues that have not been resolved (Hernández Alcaide, 2013). Today, grammar seems an aspect of oral language relatively easy to observe and categorize. Our proposal is that similar patterns to those in oral language also exist in gestures but as our understanding of non-verbal communication is in general very limited, those patterns are yet to be discovered. Even in the field of facial gestures, one of the first to be studied and so better known, new observations are contradicting what was recently held as true. For years it had been considered that facial expressions of certain emotions were universally recognized (surprise, sadness, joy, fear, anger, disgust) and yet studies published in 2012 by Jack, Caldara and Synchs have shown that there are significant cultural differences between Asians and Westerners, not all expressions being expressed by the same face parts.

---

[49] For more details on the preliminary study see López-Ozieblo (2012).

[50] Experts put the date between 150 - 200 thousand years ago, to 50,000 years ago, coinciding with the era when the arts and tools were developed (Corballis, 2011 pp. 194-195).

Systematic observation is the only method that will eventually allow us to find any patterns in non-verbal communication. Therefore we believe that a research based study is necessary if the objective is to find more information about the use of gesture and disfluency. Much has been said on the importance of carrying out comprehensive analysis of the communicative act, frame by frame, including the verbal and non-verbal parts of the interaction (Bavelas, 1994; Bavelas & Chovil, 2006; Gullberg & Kita, 2009; Kendon, 1967). However only observations may not be sufficient to recognize patterns that underlie communication, as these are often invisible even to trained observers. The amount of data is often such that it is necessary to use statistical tools and algorithms to discover these possible patterns. *THEME©*, used in this study, is one such tool, developed to find behaviour patterns, both human social behaviour and biological, like interactions between neurons (Magnusson, 2006, 2005, 2002; Hardway & Duncan, 2005).

This study covers in equal detail the performance of disfluencies and gestures, therefore it was necessary to apply and combine paradigms and methodologies used in both discourse and gesture analysis. Within the field of sociolinguistics it is more common to find studies that follow naturalistic methods. However, given the difficulties associated with data collection in naturalistic contexts and in order to control the many variables associated with the communicative act, we concluded that a quasi-experimental method was necessary (Madrid, Pueyo & Hockley, n.d., p. 7); given the nature of gestures and disfluency, both being idiosyncratic, the main paradigm for our research has been anti-positivist. These choices are justified below from both discourse and gesture analysis theory.

Observations, transcription and analysis of the data were systematic, rational and replicable, in as much as possible, to ensure its reliability, acceptability and validity (Tuckman, 1978).[51] To ensure their validity, transcripts of all the data used for the calculations can be accessed through *Dropbox*.[52] The data comes from real sources and it has been quantified and normalised (as a percentage) whenever possible to allow for statistical manipulation to obtain objective conclusions. To ensure its acceptability this study

[51] Cited by Brown, 1988, p. 4.

[52] Link to the Dropbox file:
https://www.dropbox.com/sh/yuljvsf40kta510/AAB_IFWFN3OgyigyNIPBYVpMa?dl=0.

is based on similar studies (mentioned in the discussion), and has used the same transcription and categorisations as previous studies. It is not possible to replicate the conversations therefore the reliability of the study is based on the quality of the transcripts and analysis. As this is an individual study, observations could only be validated through repeated listening to the recordings. All transcripts were checked at least twice with different software programs.[53] If discrepancies were found, the checks were repeated to confirm the results. Analyses were repeated after a period of at least 4 weeks to confirm the results. Throughout the analysis automatic checks have been programmed to ensure the reliability of the calculations. However, we strongly believe that studies of this nature benefit from at least two observers and researchers to confirm observations and discuss the results – which in this case was not possible.

5.1.1 Justification of the methodology adopted

The methodology followed previous gesture and disfluency research studies, usually based on the methodology adopted by Conversation Analysis.

5.1.1.1 Based on discourse analysis

Within Discourse Analysis, Conversation Analysis stipulates that a conversation is interactive, involving at least two individuals; speaker and addressee manage the turns and their content; and this content is usually practical. Something is said (locutionary act) in order to do something to an addressee, the intended meaning of the act: to impress, request, etc. (illocutionary act); and to have an effect on the addressee (perlocutionary act) (Austin, 1962/1975). Usually the speaker's main intention in performing the utterance is to ensure the addressee recognizes the purpose of the act (Grice 1968). In order for the utterance to be successful (to be recognized for what it is) speaker and addressee need to share the knowledge that will allow them to codify and interpret the utterance act correctly. The closer the relationship between the participants the more likely it is that the purpose will be communicated between them, but not necessarily to a third party observer (Albright, Cohen, Malloy, Christ & Bromgard, 2004).

[53] Software programmes used included: Praat, *Video-Pad Editor*, *THEME transcriber* and *Express-Scribe*.

The communicative intent of narrating a story shared by the speaker and the addressee, under experimental conditions, is clearly different from a spontaneous dyadic conversation in a natural setting. It could be argued that in the experimental condition the speaker's main objective is perlocutionary, mainly to impress the addressee with their memory recall; or just to complete the requested task, while the addressee's interest might lie elsewhere. In analysing speaker's disfluencies, the locutionary act itself is most important (the articulated utterances). The communicate act and its purposes vary with content, context and addressee but it is also likely to vary with language spoken. At a very basic locutionary act level, research suggests that even the number of words used to retell a story vary from monolinguals to bilinguals speaking the same language (Brown et al., 2014). If significant differences can be found at that basic level (number of words used), it would be adequate to expect many other differences between the L1 and L2 of the same speakers.

We ask whether a native addressee of the foreign language used by the speaker will be able to interpret the communicative act correctly. In second language and foreign language acquisition the focus is often on the locutionary act, in particular the syntactical and phonological components; educators are quick to recognize these and match them to either the speaker's L1 or L2. Not much research has been done in the differences of disfluencies by language, not only from a locutionary point of view, but also for illocutionary or perlocutionary purposes. Often it is assumed that the Gricean Maxims and Cooperative Principle apply to all speakers, forgetting that intercultural differences also play a role in determining the rules of conversations. A successful conversation depends on understanding these rules that help the interlocutors to manage the 'infrastructure' elements such as the turns, who has the floor when, how to use disfluencies and repairs, etc. Considerable work has been devoted to the study of turn taking and the non-verbal behaviour that accompanies it. In particular Duncan (1972) highlighted a number of cues that are used by interlocutors to manage the turn. He included: intonation, lengthening of the last syllable, end of the gesture, lowering of the volume or the tone of voice, relaxation of the foot or the vertebral column (López-Ozieblo, 2012).

Disfluencies and repairs, other important features of conversations, have received much less attention apart from purely linguistic features; most of these studies are of disfluencies in English (Blackmer & Mitton, 1991; Clark & Fox Tree, 2002; Corley & Hartsuiker, 2011; Kjellmer, 2003). The disfluency of a foreign language speaker might have a very different illocutionary purpose (consciously or not) than that of a native speaker. Graziano and Gullberg (2014) suggest that this is often to elicit help from the addressee, this purpose becomes obvious when the analysis of the conversation includes the gestures (one certainty regarding speaker's disfluencies, regardless of language, content and context, is that they indicate a conflict – internal or external).

It is very difficult to teach the infrastructure bounding any interaction in a formal environment, as it does not exist as a set of rules but gets built alongside the interaction (Schegloff, 1987; Maheux-Pelletier & Golato, 2008). However, it should be possible to observe whether there are any particular cultural traits, such as longer pauses or interruptions that might indicate a preference for correction over fluency and teach these.

Therefore we considered it essential to reduce the infrastructure elements participants would need to manage in the conversation. The most obvious one being turn management; in order to do this, the main study elicited monologues from our participants. The second conclusion was that it would be beneficial to be able to compare the production of the same speaker in their L1 and L2, maintaining as many variables as possible constant (such as the type of input), thus all speakers were asked to retell a story based on video input in their L1 and their L2.

5.1.1.2 Based on gesture analysis

Gesture research and analysis is anything but consistent, from the definition of a gesture to the conditions in which the gestures are elicited. It is thus very difficult to compare results or use existing data to extract new information. A recent meta-study by Bavelas and Healing (2013) points out these differences and concludes that often the conditions or variables of the experiments carried out are not even adequate to test the hypothesis proposed by the studies.

The conditions and variables can be somewhat controlled and so the challenge is to choose those which will give the best results. The purpose of an experiment is to extract spontaneous speech and gesture, akin to those one might observe in real daily life that can be analysed. Kuhlen and Brennan (2013) state that: "The challenge of studying dialogue experimentally is to create situations in which language processes can be observed in a controlled fashion while preserving the natural development of the phenomena of interest" (n.p.). To obtain comparable gestures it is necessary to somewhat control the content of the speech and to record the performance of the speaker; it is not practicable to just listen to everyday talk (although this has been done by following people with a camera and recording whole days, Mondada, 2012; or recording interactions in specific situations like meals, Goodwin, 1979). Most gesture studies control speakers' content by providing them with a theme: retelling a cartoon (Kimbara, 2007; Kita and Özyürek, 2003; McNeill, 2000); giving directions (Cassell, Kopp, Tepper, Ferriman, & Striegnitz, 2007); or describing something such as: a dress (Bavelas et al., 2008); Tangram figures (de Ruiter, Bangerter & Dings, 2012); a space (Bavelas, Gerwing, Allison & Sutton, 2011; Seyfeddinipur, 2006); a task by imagining interaction with objects (Chu & Kita, 2011). In addition they might also control the "sociability" of the exchange, by allowing it to be an unconstrained dialogue between two volunteers (where neither have been prepped as to what they might or not do or say); a volunteer and a confederate (who might also be a volunteer but most often is someone working with the researcher and has been briefed as to how to behave); or in some cases just a volunteer who delivers a monologue (although usually there is an addressee, the researcher). The conditions of visibility also vary, with dyads communicating face to face, through intercoms, or separated by a partial or full screen.[54]

Providing a topic is necessary to ensure the speakers have something to talk about, the topic might or might not be familiar to the speakers (depending on the objective of the experiment, if difficulties in lexical retrieval are being observed it might be best to use a less common topic); it might entice virtual object manipulation or spatial description to induce more iconical or deictical gestures; it could be contentious, to involucrate the speakers and generate more interaction gestures.

Whether speakers or addresses should be briefed (the use of confederates) has been a recent topic of discussion. Discourse analysis advocates for a natural setting where

[54] For a detailed description see Bavelas and Healing, 2013.

speakers are free to interact; Clark (1996) points out that speaker and addressee need each other, the discourse act being a joint feat (p. 29). We are aware that by using a confederate, and controlling the topic, the discourse is less than free; if the confederate is restricted in his or her actions this might affect the speaker and result in a less than natural performance. On the other hand the confederate might help reduce stress, motivate, and collaborate with the speaker. Kuhlen and Brennan (2013) in an article summarizing confederate use and views, suggest that there is no categorical 'yes or no' answer as to the use of confederates, but that the objectives of the research should dictate this need. In particular they warn against using a confederate in cases where they are used as addressee to different speakers telling the same story, as addressees who know too much might affect the way the speakers tell the story. They suggest instead, that the confederate should be allowed to work together with the speaker to complete a specific task or that their confederate role is made obvious to the speaker. An additional issue Kuhlen and Brennan highlight is that confederates are rarely properly briefed and their restrictions are not made clear in the description of the experiments.

It has been pointed out by Bavelas (1994, p. 208) that a speaker directs his or her speech to the addressee and not to a third observer who might not interpret the message the same way as the addressee; in cases where the addressee and the observer were the same person this was not such an issue, but when the observer was a third person a certain amount of inferring was necessary. It seems that knowing the content of the stories also helped the researcher in the interpretation of the gestures.

Based on the above, the methodology followed in the main study provided a topic in order to control possible gestures performed, help the researcher in the interpretation phase and to follow the methodologies of previous studies. The topics were cartoons, two based on the *Tweety and Sylvester* television series (as used by McNeill's lab)[55] and a third one based on a well-known fable. Participants retold the stories to a confederate, the author of this study, they were well aware of her role and, except one occasion, they did not try to elicit information from her. The confederate ensured that her response to participants was consistent, with the same sitting position, smiling, nodding and maintaining eye contact

[55] For more information on McNeill's lab go to: http://mcneilllab.uchicago.edu/.

in all cases. The preliminary study followed a different methodology as it was based on pre-existing data from analysis of dialogues between two HK students speaking in Spanish.

5.2 Preliminary Study

For the preliminary study data was collected to confirm or reject the hypothesis that there was a relationship between gesture and disfluency and a causal link between events was sought. It was hoped that the results would allow us to formulate some general conclusions that could be further tested in the main study.

5.2.1 Research paradigm

The preliminary study followed a positivist paradigm in that it tried to explore social reality through observation of individuals, distilling their actions into quantifiable comparable events. The corpus was obtained from existing research consisting of over 15 hours of recordings of dialogues between HK students of Spanish (López-Ozieblo, 2012). The study had followed a quasi-experimental design with participants allowed to speak freely but about a given topic within the constraints of a formal setting. The researcher had been an observer (present during the interaction) who spoke little or nothing, only to suggest possible topics of conversation, and control the time.

5.2.2 Corpus and research tools

The research corpus used in the preliminary study was the recorded dialogues in Spanish between students of Spanish as a foreign language from the Hong Kong Polytechnic University (HKPU), and one dialogue between two Spanish native speakers. All dialogues were dyads – that is, between two students – and were part of the on-going activities of the programme of Spanish as a foreign language offered at the HKPU. This was a formal context, in some cases the dialogues were part of assessment exercises. It was also an artificial context as students were asked to talk about issues they might not have been very interested in. [56]

There were a total of 20 conversations, dialogues 10 to 30 minutes long, during which students asked each other questions or just talked, in Spanish, about topics covered by the curriculum. The dialogues were based on a photo of a person or persons, or on specific topics given in writing, depending on the Spanish proficiency level of the students. Extracts from these were taken and any events with a disfluency and a gesture were tabulated regardless of the type of input used in the dialogue or the speaker.

All the participants were students of Spanish at the HKPU who had had either 80 or 120 face-to-face hours with a Spanish instructor within a classroom setting, not in a naturalistic context, in Hong Kong. All had been briefed on the tasks in advance and were asked for consent to be recorded and for the data to be analysed for research purposes. The observer was the author of this study, also their teacher. In addition a dialogue between two native Spanish speakers was also recorded to confirm the production of disfluency and gesture in spontaneous speech.

5.2.3 Data collection

The sessions within which the dialogues took place were video-recorded in their entirety, including any preparation time students had. To reduce the initial stress all tasks were preceded by a series of informal questions so that the students would grow accustomed to the camera. The camera was mounted on a tripod in front or slightly to the side of the participants to record the movements of the hands, shoulders and head (although only hand gestures were analysed).

A total of 15 hours were recorded, this included the informal questions and preparation time. The level of speech fluidity varied by dyad and task so 17 fluid dialogues were pre-selected (three dialogues on different topics were recorded for each pair of

[56] To avoid this artificial context and still obtain spontaneous dialogues, the alternatives were either secret recordings, which were not allowed within the institutional context, or constant recordings of students' interactions, which was not feasible.

students). This pre-selection was necessary to ensure there would be content to analyse, since in dialogues between students with lower levels of proficiency there were often long and frequent pauses or single utterances that would not have led to any meaningful conclusions. Out of the 17 dialogues pre-selected, one to three (depending on the total length of the dialogue) random extracts were taken from each ensuring there was at least one disfluency in each. A total of 25 extracts were selected (the list of extracts is given in Appendix 2). The average extract length was 16.08 seconds, the longest being 40 seconds long. Sixty four per cent of the students whose extracts were selected had a proficiency level A1 in Spanish as a FL and the rest a level A2 (the levels were set according to the guidelines of the *Plan Curricular del Instituto Cervantes* (Caffarel, 2008). In addition two extracts from the dialogue between the two Spanish natives were also selected. The selected extracts were transcribed, first the speech and then the gesture, with the help of *VideoPad Editor* (a video software program). They were checked a few weeks later using *Express Scribe* (a video transcription software). The linguistic transcription was carried out according to an adaptation of the Jeffersonian transcription system that included annotations for disfluencies and gestures (for details of the annotations used see Appendix 3). Once the accuracy of the transcripts was confirmed gestural phrases, as described by Kendon (2004), were identified, and gestures categorized following an adaption of Kendon's descriptions (2004, pp. 100), detailed in Appendix 4.

The speech transcription was marked by speech disfluencies, a row always ending in a disfluency (except when the disfluency was a pause or a non-reduction). The following events were annotated:[57]

Disruption point:

- Due to an error.
- Due to appropriateness.
- Due to other (unknown reasons).

Disfluencies:

- Fillers (or markers).
- Elongations (or lengthenings).

[57] The annotation symbols used are listed in Appendix 5.

- Cut-offs (or interruptions) and whether these were after word completion or during the word.
- Repetition[58]

Repairs:

- Repetition
- Substitution
- Deletion
- Addition
- Continuations

Type of Gesture:

- Iconic and metaphoric (labelled iconic).
- Deictic.
- Lexical (batons).
- Emblematic.
- Adaptors.
- No gesture, gesture on hold before the disfluency, gesture on hold at disfluency.

Gesture information recorded whether the disfluency fell within a gestural phrase or not and what phase of the phrase it was:

- Preparation.
- Pre-stroke hold.
- Stroke.
- Post-stroke hold.
- Retraction.
- No gesture.

Synchronicity of speech and gesture at the beginning of disfluency, during the disfluency and at the onset of the repair.

- Synchrony gesture-word.
- Gesture stops before the word.

[58] Repetitions were included here originally following the practice of other researchers in disfluency-gesture (Stam, 2006). The main study recognizes the repetition as a repair and classifies it as such.

- Gesture starts before the word.
- No gesture or no speech.

5.2.4 Data analysis

Out of the 25 extracts analysed from the dialogues of HK students, 160 disfluencies were observed. Out of these 110 were occurred together with a gesture. All cases of disfluencies were tabulated with all the information about the disfluency, the repair and the gesture. Appendix 6 gives an example of the transcription from conversation extract no. 1, together with an in-depth description and analysis of the gesture phrases explaining their integration with the speech. Such detailed descriptions were noted only in a handful of cases as such in-depth analysis was not the objective of this study.[59]

As the main objective was to find whether there was a relationship between the disfluency and the gesture a statistical analysis was necessary to review all the data and find whether any such relationships would be significant. It was decided to use a significance of 90%, ($p < 0.01$), as this is the confidence level used in similar studies carried out in the field of Language Studies (Brown, 1988, p. 116), which means that we are 90% certain that relationships observed are not the result of chance. To identify the patterns we used a tool called *THEME ©,* a software that analyses strings of data through various algorithms to identify possible patterns not readily observable. These patterns are in the form of repeated strings of events (two or more). The algorithm seeks patterns between the data treating each data point as an independent value, which means that we could have more than one data event from the same participant (dependent events) and this should not interfere in the analysis. The following variables were identified:

1. Dependent variable: gesture-speech disfluency.
2. Independent variables (as detailed above): reason for the disruption; type of gesture, repair and disfluency, synchrony gesture-speech; phase of gestural phrase.

---

[59] For the results of such a study to be meaningful it would have been necessary to analyze more than a handful of gesture-speech performance of one individual, unfortunately we did not have enough data to do that.

3. Controlled variables:
   - Manipulated (in order to have meaningful data for the analysis): fluidity of speech.
   - Fixed (to facilitate analysis): gender (female students only); age (between 20 and 25 years); knowledge of Spanish (A1 and A2); nationality (China, Hong Kong); sociolinguistic aspects (all students at HKPU, Cantonese as their mother tongue and English as a second language, similar education and socioeconomic status).

The data table was modified according to the guidelines of THEME© and fed to the software.[60]

5.2.5 Results

The results from the data confirmed that HK students also produce disfluencies when speaking Spanish and that these did co-occur with gestures in some cases. We also observed that there were significant individual differences in disfluency and gesture and so more data for each individual would be necessary to carry out statistical research. However, when it came to using THEME © to find a pattern only two strings of patterns emerged as being significant. These results were confirmed by Magnusson (author of the software, personal communication, May 2013). The reasons for the lack of patterns, which included having too many variables and not enough data, are detailed in the discussion below. As a result, every attempt to control such fluctuations was made in the main study.

5.3 Main Study

For the main study we took the view that speech and gesture are idiosyncratic actions, each individual being different, and so each participant's performance was considered a case study, analysed separately, and then the results cross-examined and compared.

[60] The data used can be found at https://www.dropbox.com/sh/yuljvsf40kta510/AAB_IFWFN3OgyigyNIPBYVpMa?dl=0 .

5.3.1 Research paradigm

The main study followed an anti-positivism paradigm in the research development and execution and a Critical Theory paradigm in the analysis of the results. Anti-positivism focuses on the individual, through case studies in as natural a setting as possible. The methodology employed to carry out the research made the context as natural as possible (within the constraints of the research itself). It focused on a number of individual cases, encouraging participants to provide reasons for their actions, which in our analysis were then combined with their performance, to develop the discussion and our conclusions.

5.3.2 Corpus and research tools

This study took place over a period of 12 months during which 28 participants were recorded: 14 were Spanish natives, all except two, students of English at the University of Málaga; and 12 were Hong Kong natives, as mentioned before although Cantonese was their mother tongue they were considered to be fluent in English, having been raised in a bilingual environment; and two were neither but spoke Spanish as a foreign or second language. Not all of the HK participants were students of Spanish as a FL or were confident enough to do the retelling in Spanish. Furthermore some participants held their hands between their legs or in their pockets for most of the recordings not producing any gestures. One participant had been diagnosed with mild autism. Each individual's profile and participation is detailed in Appendix 7.

The subjects were all volunteers who received no compensation for their time, they were not informed about the specific nature of the research but all understood its general nature and agreed to be part of it. Most of the recordings took place within an institutional setting. Most participants were students, either of Spanish as a foreign language studying under the researcher at the Hong Kong Polytechnic University or Spanish native speakers, living in Spain, studying English at the University of Málaga (Spain). This latter group of participants was not known to the researcher but had volunteered to participate in the study. The study was approved by the ethics committee of the HKPU, number HSEARS20131213001. Appendix 8 includes a copy of the form volunteers were asked to sign as well as the description of the project they read before giving their consent to participate.

Not all of the participants were ideal subjects for the study so some were recorded to test the methodology and ensure the smooth running of the recordings to be used for the analysis.

Many gesture studies are based on visual inputs: cartoons, or computer images that subjects have to describe, recall or recount in the case of stories. We wondered whether the fact that the input was a visual one might taint the speech-gesture performance process – leading to more gestures if these are direct representations of the visual input – therefore it was decided to use three different types of input with each speaker: a written one, an aural one and a visual one. In all three cases the speaker had to read, hear or watch up to three minutes of the input and then recount the story with as much detail as possible. Initially this was done in their L1 and then, those who felt comfortable doing so, watched the last three minutes of the video and retold it in their L2. In all cases one type of input was given and the story retold immediately after.

For this experiment, subjects were not told the specific objective of the study, although they knew it was to study verbal and non-verbal communication on retelling stories coming from different inputs and whether recollection depended on the type of input. After the last retelling speakers were asked a number of questions in their L1 to further understand their memories and speech processes. HK students who participated were also asked to fill in a questionnaire that categorized them as a visual, aural, kinaesthetic or read/write type of learner (*VARK*). This helped to maintain the cover of the research and gave something back – hopefully useful – to our volunteers.

In order to relate this study to existing gesture studies, for the audio and video we chose to use *Tweety and Sylvester* stories (a television cartoon). Based on the gestures work pioneered by McNeill (2000), we selected the *Canary Row* (Warner Brothers, 1950) episode for the video input. The aural story was based on a made-up episode of *Tweety and Sylvester*. The written texts were versions of the fable of T*he Lion and the Mouse* by Aesop downloaded from the internet (different versions, one in English and one in Spanish). This was the first story given to the subjects as many of them confessed to being nervous. It was thought that a somewhat familiar story might help them relax. In most cases subjects were also familiar with *Tweety and Sylvester*, but we don't believe this affected our results as our objective was trying to understand the gesture-speech unity. Appendix 1 provides a summary of the video, a transcript of the aural story and the text used for the fable.

The three inputs were designed to be as similar as possible in terms of length; an average of three minutes of input time was given to participants for each exercise. The input was also consistent in its content level (although the video was split into two parts, each with a series of similar events and the conclusion was given in the second half).

There were 381 words in the English text (340 in the Spanish version), read at a speed of 250 to 300 words per minute, this should take an average of 1.2 minutes to read, although most students took longer, suggesting they were re-reading the text. The aural story contained 225 words in English and it lasted 1.5 minutes, a number of students requested to hear it twice (others declined the offer to hear it again). The Spanish version told the same story but was slightly shorter, lasting 1.4 minutes, it contained 218 words. The video was silent and 3.1 minutes were watched to be retold in the mother tongue and another 3.3 in the foreign language.[61]

After some testing, it was decided that the inputs were to follow this order: text, aural, video. This ensured that images from the video would not cause any interference to the aural story and for those speaking a foreign language it gave them the confidence of a known story (as the first part had been retold in the L1).

5.3.3 Data collection

Recordings were conducted in various locations, although mostly in a private office or in a sound-proof room with only the subject and the researcher present, both were seated facing each other, and the subject was talking to the researcher. Each session lasted about 40 minutes, this included briefing and debriefing time. Each session was recorded with a video camera and also with an i-pad, or with two i-pads if a camera was not available. Each device was placed slightly off-centre from the subject giving a clear view of his or her central body, arms and hands. Most of the session was recorded, except when subjects were reading, hearing or watching the video (to minimize video length times which might

[61] Appendix 1 provides the content used for each type of input.

have hindered the transcription process). The researcher did not interrupt, help verbally or reply at any point, however she did provide non-verbal cues (such as gaze and smiles) to indicate understanding and agreement. In many instances the researcher was known to the subject and this contributed to a relaxed atmosphere during the experiment. An effort was made to keep the temperature of the room comfortable to suppress a normal reaction to keep hands between the legs or in pockets. As no specific mention was made of recording hand gestures it was considered essential not to draw attention to them in any way, for instance by asking the subject to place his or her hands on their lap. Whenever possible a chair with arm rests was provided to discourage participants from hiding their hands.

5.3.3.1 Recording process

The sessions were recorded on video at 24 frames per second in frames of 1080 x 1920 pixels. Most recordings lasted around three minutes, which was considered an ideal length for handling the files (bigger files often crashed). When necessary the aspect ratio was manipulated (to change to a portrait from a landscape layout).

5.3.3.2 Data selection

Although the focus of this study was not to find gender differences, we were aware of the need for gender equality and so the contribution of the participations analysed were specifically chosen to ensure data from both genders. Out of the 28 participants 11 were male, five being Spanish native speakers with English as a FL. We only had two male HK students with Spanish as their FL willing to participate, so to ensure an equal number of participants from each gender and language it was necessary to select two male and two female Spanish natives and two female HK students. In total we analysed eight case studies, four Spanish native speakers and four HK students of Spanish as a FL. These provided us with four retellings each, plus recordings of their impressions of the exercise and which input had been the hardest to retell and why (these were not used in the analysis). In total 32 retellings of various lengths were analysed. In some cases, when the retelling exceeded 170 seconds the length of the script was cut to ensure some consistency in the lengths of the transcripts.

Spanish natives and HK female students' transcriptions were chosen by asking a panel of independent volunteers to rank performances of each group (divided by gender and language) by perceived number of gestures performed overall in all the recordings. We selected the highest and lowest gesture performers of each group. With this selection we hoped to obtain an average sample when all the data was combined for each sub-group. The reality is that gesture performance perception is very subjective and the number of gestures varies by participant but also by input, as we found during the analysis.

The data was divided by group: HK students and Spanish natives and gender (numbers 1 and 2 in each group being females). Each participant's data was labelled accordingly, thus Case study 1HK (C.S. 1hk) refers to a HK female student. Her participation was in English (L1) or Spanish (L2) and it was a retell of either the written input, the aural recording (audio) or the video.

5.3.3.3 Transcription process

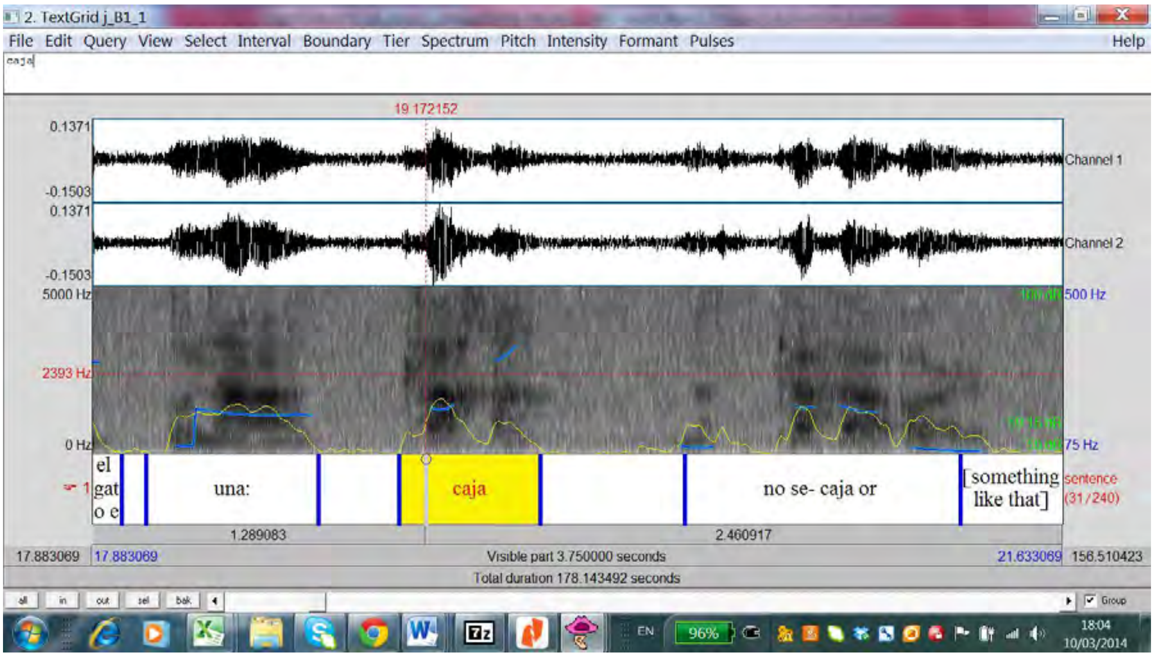
The speech in the recordings was transcribed, using the same annotation as in the preliminary study. *Express-Scribe* and *Video-Pad Editor*[62] (software tools for transcribing and video manipulation respectively) were used to do the initial speech transcription (see Appendix 3 for details of the transcription annotation). For the transcription, the videos were slowed down to different speeds (usually to 30% of the original speed was best as it was still possible to understand the speech); frame by frame analysis was carried out when looking at the gestures (*Video-Pad Editor* allows analysis down to an average of 40 frames per second).

The transcriptions selected for analysis were checked a second time, to confirm the original results. During the transcription process for the preliminary study, pauses had been initially identified manually with the help of a spectrogram available with the software *Video-Pad Editor*. However it was considered that a more accurate result might be obtained by using an automated programme, such as that available through the software *Praat*. *Praat* can be programmed to predefine pauses depending on their length and intensity and to

[62] Elan was used initially, however it was found that the video files were too large for the software and frequent crashes prevented smooth analysis.

mark the boundaries of the pause, giving its duration. The pause is marked by boundary lines and the duration of the pause is then automatically calculated. The programme analyses speech providing a waveform, a spectrogram, pitch and intensity graphs (as shown in Figure 10 below). The minimum pause duration to be recognized was set at 0.01 seconds but it was often necessary to revise the pause boundaries manually based on the graphs. From these three graphs it is possible to pinpoint quite accurately the beginning and end of a pause. It is possible to zoom up to 0.00001 seconds, thus providing very accurate pause starts.

Figure 10. - *Praat* screen-shot.



Note: pitch is shown in blue; intensity in yellow over the spectrogram (darker shades indicate speech) and the waveforms are shown in the two windows at the top (stereo sound). Speech–pause boundaries are indicated by the vertical blue lines below the graphs and the relevant speech transcription has been added.

Unfortunately the automatic function is not totally reliable, external noises, very low pitches or intensities interfere with automatic pause and speech detection and it is necessary to double check the results manually. There were instances where manual adjustments were necessary as *Praat* does not always record certain consonants, in particular /k/ and /g/ which, although easily heard, need to be added manually and thus add error to the data.

5.3.3.4 Description of annotations

Praat records pause duration up to 6 decimal points, but three decimal points was considered enough for this study. The results are divided into *Praat* sentences (in the example above there were 240 sentences as marked on the bottom left of the graph, sentences are bound by pauses or significant pitch/intensity changes), it also allows the transcription to be added to each detected sentence. The information was saved into a text file which was copied onto excel where it could be manipulated. Initially all pauses were marked with their duration. Then they were looked at in relation to the types of disfluency, fillers ('uh' and 'um'), cut-offs and elongations.

Using *Praat* sentences separated by pauses gave us the data divided into phonic groups (groups of phonemes separated by pauses longer than 0.01 seconds, in our case). Each phonic group was transcribed on a separate line, unless the utterance was too long, in that case it was divided at the point of a disfluency (if there was no disfluency the utterance was noted in as many lines as necessary to fit the format of the table). The analysis of the transcriptions noted the following:

- Number of words and phonic groups for the whole retelling, total length of the retelling.
- Number of gestural phrases (marked by [ ]).
- The stroke within the gestural phrase (marked by an underline); only in some cases as it is not the focus of this study.[63]
- Errors not leading to disfluencies; only in some cases as it is not the focus of this study.
- Type of hiatus (pure pause, pause with a disfluency).
- Type of disfluency (elongation, pause, cut-off, filler).

[63] A number of notes were made in the data to aid future research that might be based on this corpus.

- Type of repair (addition, substitution, deletion, repetition) or continuation of speech.
- Type of gesture at the beginning of the disfluency (suspension point), if any (pragmatic or iconic).
- Gesture phase at the beginning of the disfluency (suspension point).
- Gesture during hiatus.

The disfluencies recorded are described in Table 3:

Table 3. – Annotation of disfluencies in speech.

Type of disfluency	**Annotation in transcript**	**Description and annotation used for analysis purposes**
Pause	/	(0.000) Pauses refer to silences; these were automatically identified in *Praat*, although in some cases manual adjustments were necessary (as indicated above). The length of the pause is shown in seconds. Pauses are also marked in the transcript in a separate row from speech. Most disfluencies are followed by a pause (in a following row in the transcript). If there is no pause this is marked in the transcript with an ‘x’ after the disfluency, as also before in the case of fillers.
cut-off	-	(‘i’) An interruption of the word, the analysis recorded if the cut-off occurred after the word had been completed (‘i’) or if the cut-off occurred within a word (‘wi’). It also recorded whether there had been an interruption of the word but the last sound was elongated (‘wi:’). In some cases there was an after word completion cut-off, marked by a glottal sound or a sudden stop of the elongation (‘i:’). It was also recorded whether there was a pause after the cut-off or if there was speech, this was marked with an ‘x’ (‘i’ x’; ‘wi x’; ‘i: x’; ‘wi: x’)

Type of disfluency	Annotation in transcript	Description and annotation used for analysis purposes
Elongation	:	('l') Elongations were originally identified manually, bearing in mind that each speaker is different and some words in English have elongated vowels. Confirmation of elongation was done with the help of *Praat. A*s explained below this software gives speech information in graph form that identifies pitch, intensity and waveforms. A combination of these making it possible to confirm elongations. The length of the elongation was not recorded. In some cases the elongation was followed by speech ('l x').
Filler	Uh, um or equivalent	('f') Fillers included instances of 'uh' and 'um' and their variations as well as other fillers such as 'and then', or 'bueno' (in Spanish). The pauses before and after the elongations were also recorded, if there was no pause an 'x' marked this ('x uh x' indicates no pause before or after the filler). If the filler was elongated this was recorded ('u:m').

It was decided not to analyse non-reductions as subjects were speaking in a second language and it was deemed that identifying non-reductions might not be possible, as there could be issues in the pronunciation.[64] The resumption points were analysed by type of repair, if any. The repairs recorded are listed below in Table 4:

[64] This was indeed the case, to the extent that a truly phonetic transcript was not feasible, as it would have hindered understanding of the utterance when read by a third person not familiar with Cantonese English or with Spanish as a Foreign Language. As the study was not focusing on the phonetic elements it was considered appropriate not to do a phonetic transcription.

Table 4. – Annotation of repairs in speech.

Type of resumption	Annotation
Addition	a
substitution	s
Deletion	d
Repetition	r
Continuation of speech	c

The nature of all gestures, concurrent with the disfluency or not, were annotated recording whether they were pragmatic (p) or iconic (i) gestures. Iconic gestures are those that refer to the content of the utterance in any way (metaphorical, iconical or indexical). Pragmatic gestures were those that did not seem to refer to the content of the speech itself but were used to stress parts of the utterance or manage it, such as quick flips of the hand that can be interpreted to mean 'you know what I mean'. This category included lexical retrieval movements seemingly used to recall a word, such as a circling of the index finger in a horizontal plane. Adaptors, such as head scratching, were not included. In cases where the gestures were unclear this was marked in the transcript and it was not included in the analysis.

At the point of suspension as well as during the hiatus the gesture (if any) was noted to see if this stopped with the speech and whether it remained frozen or continued during the hiatus. Gesture annotation was based on the annotations used by Seyfeddinipur and Kita (2001, pp. 460-461), although simplified, and added to the transcriptions with the aid of *Video-Pad Editor*. Table 5 describes the annotations used:

Table 5. – Annotation of gesture in disfluency.

Start of disfluency	At resumption
(s) Stop: an on-going gestural movement is suspended, gesture is put on hold.	(c) Continuation: The gesture continues.
(c) Continuation: The gesture continues.	(n) A new gesture phrase begins.
(e) Start: a new gesture phrase commences.	(r) The previous gesture phase is repeated (within the same phrase).
	(0) The gesture is still on hold or has been suspended

Although not the focus of this study, but considering the possibility of using the data for future research, the gestural phrase during the original delivery, the reparandum, the hiatus and the resumed delivery were described in some cases, including their likely interpretation.

We decided not to make pure pauses the focus of this research, despite the fact that pure (unfilled) pauses are obviously a disfluency. There are a number of reasons for this: the main one is that most of the recordings were not made in sound-proof environments and as such there are a number of external noises which might interfere with a fully accurate annotation of pauses, also accurate pause recording and analysis require specific tools which were unavailable to us. An additional reason for not analysing pauses is that they are highly variable (at least in spontaneous speech), dependent on the individual, the type of interaction (Goldman-Eisler, 1968, p. 15) as well as the language (Campione & Veronis, 2002). The latters' study on short (<200 milliseconds), medium and long (>1000 milliseconds) pauses produced in readings in European languages have shown that Spanish pause duration is on average 100 milliseconds (ms.) longer than in English, French or Italian (p. 202). As our study is already dealing with a number of variables dependent on individuals it was decided to exclude a comprehensive pure pause analysis. However pauses preceding and following other disfluencies were noted as well as their approximate length.

5.3.4 Data analysis

We have chosen to use a contrastive analysis approach to try to predict problems learners might have in the L2 basing ourselves on native use of that language. Although this approach, influenced by structuralism and behaviourism, was very popular half a century ago, it has since been discredited as not always predicting learners' errors, or predicting errors that never take place. However, we find that it is still being used in gesture studies in which L1 and L2 utterances are compared, and indeed in a number of other research studies looking at cross-cultural comparisons and at positive transfers (Saville-Troike, 2012, pp. 39-40). In gesture studies it has been suggested that there is transfer both from the L1 into the L2 and from the L2 to the L1 (Pika, Nicoladis & Marentette, 2006). The tendency is to look at negative transfers, or interferences (where the use is inappropriate in the receiving language). This study looks for patterns that might indicate either type of transfer, through the analysis of comparable uses of the language in Spanish and English.

In addition, as we are also focusing on disfluencies, there was an expectation that we would find a transitional state between the L1 and the L2, the interlanguage, where speakers would be following a set of rules that would not belong to either the L1 or the L2, and leading to 'errors'. This stage would be idiosyncratic to each learner as it changes with the learner's position in the acquisition continuum. We resist labelling this study an error analysis, as we do not believe disfluencies are necessarily indications of errors. The analysis identified instances of disfluency but not of actual errors where there was no apparent disfluency, see example 1 below. The disfluency might have occurred later on in the utterance but if it wasn't repaired immediately it would not have been linked directly to that error (see example 2). In the following sentence from the corpus:

tweety bird[65] saw outside the window there were big fat worms

the syntax error was not identified and the speaker carried on without any disruption. So this error was considered an uncorrected error. A correct option could have been (correction in italics):

[65] No capital letters were used in the transcription except to signify emphasis.

> tweety bird saw *that* outside the window there were big fat worms, or
> tweety bird saw *big fat worms* outside the window

In this other instance:

> sylvester fall off the grass and it was rolling back uh /

the most correct option would have been:

> sylvester fall*s on* the grass and *rolls down* uh /

the marker and the pause indicate a disfluency but it is not possible to state whether it is directly linked to the first ('fall') or the second error ('rolling back') or whether these are syntaxical or phonetical, conceptual or semantical.

The number of each type of event, as detailed above, was calculated for each participant by type of input (text, aural, visual) and language (Spanish, English) and compared manually. As the main objective was to record disfluencies, the initial analysis used only the speech transcript. The analysis recorded the average length of the pauses, the total number of each type of disfluency, whether there was a pause before or/and after it and their average length. The length of the various pauses and types of disfluencies and how they were repaired were noted. In some cases the possible meaning of the gesture was described in detail. The following variables were identified:

1. Dependent variable: gesture-speech disfluency.
2. Independent variables:
   - the type of disfluency and sub type (in the case of fillers and cut-offs);
   - the type of repair;
   - whether there was a pause preceding or following the disfluency;
   - length of the pause (if it existed);
   - type of gesture;
   - synchrony gesture-speech at point of disruption
   - synchrony gesture-speech during the hiatus or disfluency;
3. Controlled variables:

- Manipulated (in order to have meaningful data for the analysis):
  - gender;
  - mother tongue;
  - type of input for retell.
- Fixed:
  - age (between 20 and 25 years);
  - knowledge of Spanish (A1 and A2)for HK students or English (level B1 and B2 for Spanish natives);
  - nationality (China, Hong Kong) or Spanish;
  - sociolinguistic aspects (all university students).

Simple frequency graphs were produced to provide easy to read visual input as to the frequency of each event recorded. Data from the participants was compared, by type of input and by language. This data was again compared by speaker, and language. Disfluencies with and without gesture were studied and gestures with and without disfluencies were also analysed. To carry out meaningful comparisons of the data the data was normalised and converted into rates.

5.3.4.1 Rates

In order to normalize the number of disfluencies and gestures per speaker for comparison purposes it was necessary to provide it as a ratio. The dearth of studies on gesture-speech disfluency has meant that there are few precedents in the literature. Gullberg, de Bot and Volterra (2010) warned about asserting that L2 speakers gesture more than in the L1, without taking into account how the comparisons were made, by fluency or processing units (p. 19). The options for gesture-speech ratios include number of gestures per number of words and number of phonic groups or length. More recently, gesture per number of clauses (Colletta, Pellenq & Cefidekhanie, 2014; Graziano & Gullberg, 2014;) has replaced gesture per number of words (Stam, 2006) but the latest work by Colletta (personal communication, July 2014) suggests that phonic groups might be more relevant. In his study of speech and gesture in narrations carried out by French children aged 4 to 11, he found that gesture-speech ratios are significant at Phonic Group (PG) level but not necessarily at word or clause level. A PG is defined as a "continuous string between two inspiration breaths or silent pauses longer than 200 ms.” (Colletta, Pellenq & Cefidekhanie, 2014, n.p.).

The performance of language learners, at least at basic levels, might be comparable to that of younger native speaking children. Children 3~4 years of age point and utter a word (McNeill, 2012, p.53) referring to two different things. Language learners in the initial levels seem to follow this pattern as well.[66] Therefore using the PG to establish a gesture-speech ratio might be more meaningful than a gesture-words or clauses ratio. This conclusion has been supported by Stam (personal communication, July, 2014) who believes the PG might be a more appropriate measure to use with foreign speakers.

The exact length, 200 ms., has been taken from previous studies (Hagoort & Levelt, 2009) although Colletta himself (personal communication, July 2014) believes that for his younger subjects this might be too short. A search of the literature does not shed much light on the issue, except for the work done by Kendall (2013) whose studies of pauses and speech rates suggest that pause length might vary by pause function, with shorter pauses (<220 ms.) perhaps indicating cognitive processes, such as word searching, while longer pauses might be of a pragmatic nature, such as providing emphasis. He concludes that the 200 ms. seems to be arbitrary and up to each researcher to adjust. Kendall also points out that more research is needed in this area (p. 138). Campione and Veronis (2002) also warn against using limiting pause thresholds and pause lengths, and advise adjusting them according to the speaker’s language. Based on this advice and the work by Colletta, Pellenq, and Cefidekhanie (2014), who use the 200 ms. cut off to divide PG, we will also use this length to calculate the number of phonic groups.

Rates of disfluency and of gestures are calculated using number of words, length and number of phonic groups but in most cases number of gestures or disfluencies per number of phonic groups per second has been used for comparison purposes.

[66] Informal classroom observations.

5.3.4.2 Significance

In order to calculate the reliability of the data a number of tests were performed on any calculations that might provide meaningful results. These tests, the statistical significance, show whether the results could have been the result of chance or whether there is a pattern that could be connected to a specific event. When comparing percentages a one-sample or two-sample t-test was used. If the percentages being compared were calculated based on the same total population, a one-sample t-test was used, if the total populations were different (e.g. when comparing L1 and L2 data) then a two-sample t-test was performed. In all cases we used a 95% confidence followed by a 90% confidence if no significance was found at 95% (a 95% confidence indicates that we are 95% sure that the results are not happening by chance, i.e. there is a reason for them occurring). Unless stated otherwise, tests were two-tailed probability t-tests, as our hypothesis did not indicate whether one event would have a higher percentage than the other. In all cases we stated the null hypothesis as: "There is no significant difference between the percentage of event A and event B". If the results were found not significant, we accepted the null hypothesis and concluded that there was no significant difference between the events and that the difference could be attributed to chance and should not be considered a pattern.[67] In some cases there was enough data to compare means (averages), to identify whether the difference between them was significant. In these cases we also carried out a t-test, assuming unequal variances in all cases. A third type of test checked whether there was a correlation between pairs of data points, two variables that could be dependent or independent. As the data was either intervals or ratios we used Pearson's coefficient. The value of the coefficient can be negative (the closer to -1 being the stronger the negative correlation) or positive (the closer to +1 the stronger the positive correlation). A t-test was used to calculate the significance of the correlation (as mentioned above). To carry out the significance analysis we used, *Statistics Calculator*, an online software tool that performs the same type of analysis as license software like SPSS.

[67] If samples do not satisfy the requirement that $n*p$ and $n(1-p)$ must both be equal to or greater than 5, where n=samples size and p=percentage, a note has been added.

5.3.4.3 Calculations

The calculations were developed based on the results that were being obtained, in some cases we believed in depth analysis of a specific event might shed some light into its production and so we extended the analysis. The calculations, their results presented below, began with the analysis of disfluency and then of gestures. Under the results, below, each section gives a preliminary overview of the total number of events observed and then normalizes the data of each participant as rates (number of disfluencies or gestures per number of words, per length and per number of PG). The number of events per type of input and language is analysed next, followed by a breakdown by type (in the case of disfluencies into elongations, fillers, cut-offs and pauses; in the case of gestures into pragmatic and iconic gestures). Further calculations focus on disfluencies occurring during gesture phrases. In all cases the data was divided by participants L1 (Spanish natives and HK students).

The second part focused only on the data from the video corpus in the L1 and L2 repeating the above calculations. In addition each type of disfluency with gesture is further analysed to include the repair (in case of elongations), and to detail the type of cut-off and filler and their relation with the pause following them. The calculations reported below include basic data on the length of the retellings, number of words, and phonic groups as well as:

- Number of pauses with and without disfluencies and the length of the latter before and after the disfluency (by type of disfluency).
- Type and sub-type of disfluency in each retelling and whether they occurred with or without a gesture.
- Type of repair by type of disfluency.
- Frequency of each type of disfluency with and without pauses.
- Gesture phase (whether it continued, started or was put on hold) at the start of the disfluency
- Type of gesture (pragmatic or iconic) by type of disfluency and repair.

6 RESULTS

The results are divided into two sections: results for the entire corpus and for the video retellings, both in the L1 and the L2. An introduction to basic data is provided initially and then each section details information on disfluencies and then gestures. From this data it seems that there might be differences in the L1 data that could be related to the type of retelling. Therefore a more scientific approach was followed, comparing only the retellings that were based on the same type of input (the videos), for which there was data in both the L1 and the L2 for both groups. This approach takes out the possible variable related to differences in the type of input. The results of the various analyses are presented in the form of graphs or tables, with the data given as either percentages, frequencies (number of cases) or rates. Whenever useful, the data has been combined to present meaningful comparisons, with statistical significances calculated when relevant. All the data is available through *Dropbox*[68] and an example of the analysis carried out is detailed in Appendix 9.

6.1 Basic information

The following tables provide information about the length of the retellings, number of words and phonic groups used, as well as rates of word per second and phonic group per second.

6.1.1 Lengths of retellings

The following tables detail the lengths of the various retellings carried out in the L1 and the L2 by the participants.

[68] Link to the Dropbox file:

https://www.dropbox.com/sh/yuljvsf40kta510/AAB_IFWFN3OgyigyNIPBYVpMa?dl=0.

6.1.1.1 Spanish natives

Table 6. – Lengths of retellings, Spanish natives.

Length (seconds)	C.S. 1sp	C.S. 2sp	C.S. 3sp	C.S. 4sp
Written L1	89	90	100	73
Aural L1	52	48	65	45
Video L1	76	68	123	112
Video L2	113	52	95	66
Average all	82.50	64.50	95.75	74.00
Average L1	80.33	56.00	94.33	74.33
Standard deviation all	25.46	19.07	23.85	27.99
Chi-square for L1 data = 15.258		Probability of chance = 0.0183		Degrees of freedom = 6

6.1.1.2 Hong Kong students

Table 7. – Lengths of retellings, HK students

Length (seconds)	C.S. 1hk	C.S. 2hk	C.S. 3hk	C.S. 4hk
Written L1	170	63	106	119
Aural L1	115	76	82	63
Video L1	130	162	105	125
Video L2	161	180	169	154
Average all	144.00	120.25	115.50	115.25
Average L1	138.33	100.33	97.67	102.33
Standard deviation all	25.83	59.30	37.35	38.04
Chi-square for L1 data = 50.85		Probability of chance = 0.00		Degrees of freedom = 6

6.1.1.3 Observations

For the HK students, for all participants, the video L2 retellings extended to over 250 seconds, in order to keep the data manageable and for the lengths not to be too different from those recorded in the L1, these were cut (usually after the second event in the video was retold). Thus lengths themselves cannot be considered variables when comparing L1 and L2 differences. Spanish natives' retellings were shorter, for the written and the aural retellings. This could be because the input was also slightly shorter than the English versions. However, the video shown was the same for all participants, suggesting there were others factors making Spanish natives more succinct. There were also significant differences in the lengths of retellings by type of input.

6.1.2 Number of words used in the retellings

The following tables detail the number of words recorded in the various retellings carried out in the L1 and the L2 by the participants.

6.1.2.1 Spanish natives

Table 8. – Number of words used in the retellings, Spanish natives.

Number of words	C.S. 1sp	C.S. 2sp	C.S. 3sp	C.S. 4sp
Written L1	224	265	232	237
Aural L1	122	136	131	132
Video L1	169	190	285	387
Video L2	246	94	153	158
Average all	190.25	171.25	200.25	228.50
Average L1	171.67	197.00	216.00	252.00
Standard deviation all	55.85	73.83	71.22	114.71
Chi-square for L1 data = 68.7		Probability of chance = 0.00		Degrees of freedom = 6

6.1.2.2 Hong Kong students

Table 9. – Number of words used in the retellings, HK students.

Number of words	C.S. 1hk	C.S. 2hk	C.S. 3hk	C.S. 4hk
Written L1	397	168	290	288
Aural L1	221	165	206	156
Video L1	250	380	277	282
Video L2	163	215	199	187
Average all	257.75	232	243	228.25
Average L1	289.33	237.67	257.67	242.00
Standard deviation all	99.63	101.29	47.15	66.79
Chi-square for L1 data = 122		Probability of chance = 0.00		Degrees of freedom = 6

6.1.2.3 Observations

As the number of words will be dependent on the length, which was controlled, this information is given for record purposes. All participants attempted the retellings giving complete and usually detailed descriptions of the stories. The count of words included all fillers and incomplete words that had been interrupted. There were also significant differences in the number of words used in the retellings by type of input.

6.1.3 Number of phonic groups used in the retellings

The following tables detail the number of phonic groups recorded in the various retellings carried out in the L1 and the L2 by the participants.

6.1.3.1 Spanish natives

Table 10. – Number of phonic groups used in the retellings, Spanish natives.

Number of Phonic groups	C.S. 1sp	C.S. 2sp	C.S. 3sp	C.S. 4sp
Written L1	26	21	29	24
Aural L1	15	11	22	16
Video L1	31	17	40	28
Video L2	45	16	46	22
Average all	29.25	16.25	34.25	22.5
Average L1	24.00	16.33	30.33	22.67
Standard deviation all	12.45	4.11	10.78	5.00
Chi-square for L1 data = 2.03		Probability of chance = 0.92		Degrees of freedom = 6

6.1.3.2 Hong Kong students

Table 11. – Number of phonic groups used in the retellings, HK students

Number of phonic groups	C.S. 1hk	C.S. 2hk	C.S. 3hk	C.S. 4hk
Written L1	82	26	39	67
Aural L1	56	29	43	38
Video L1	58	48	51	67
Video L2	87	89	91	83
Average all	70.75	48	56	63.75
Average L1	65.33	34.33	44.33	57.33
Standard deviation all	16.03	29.02	23.86	18.75
Chi-square for L1 data = 16.049		Probability of chance = 0.013		Degrees of freedom = 6

6.1.3.3 Observations

The number of phonic groups will also be dependent on the length, which was controlled. Phonic groups were calculated as the number of utterances between pauses of more than 0.199 seconds (200 milliseconds and above). In Spanish natives there was not a significant difference in the number of phonic groups by retellings by type of input.

6.1.4 Rates: number of words by length (seconds)

To normalise the data and provide meaningful comparisons that would take into account speed of speech, number of pauses, phonic groups or words the data was recalculated as rates, either words per second or phonic group per second.

6.1.4.1 Spanish natives

Table 12. – Number of words by length of retellings Spanish natives.

Words/Length (seconds)	C.S. 1sp	C.S. 2sp	C.S. 3sp	C.S. 4sp
Written L1	2.52	2.94	2.32	3.25
Aural L1	2.35	2.83	2.02	2.93
Video L1	2.22	2.79	2.32	3.46
Video L2	2.18	1.81	1.61	2.39
Average all	2.32	2.59	2.07	3.01
Average L1	2.36	2.86	2.22	3.21
Standard deviation all	0.15	0.53	0.34	0.46

6.1.4.2 Hong Kong students

Table 13. – Number of words by length of retellings HK students.

Words/ Length (seconds)	C.S. 1hk	C.S. 2hk	C.S. 3hk	C.S. 4hk
Written L1	2.34	2.67	2.74	2.42
Aural L1	1.92	2.17	2.51	2.48
Video L1	1.92	2.35	2.64	2.26
Video L2	1.01	1.19	1.18	1.21
Average all	1.80	2.09	2.27	2.09
Average L1	2.06	2.39	2.63	2.38
Standard deviation all	0.56	0.63	0.73	0.59

6.1.4.3 Observations

In all cases slightly higher word per second rates are observed in the L1 retellings in both groups, suggesting faster speech. It is also interesting to note that in all cases the rate for retellings based on the written input is higher than for the aural. The significance of these differences is calculated (see Table 16).

6.1.5 Number of phonic groups per length (seconds) used in the retellings

The following tables detail the number of phonic groups per second calculated for the various retellings carried out by the participants in the L1 and the L2.

6.1.5.1 Spanish natives

Table 14. – Number of phonic groups per length (seconds) used in the retelling, Spanish natives.

Number PG/length	C.S. 1sp	C.S. 2sp	C.S. 3sp	C.S. 4sp
Written L1	0.29	0.23	0.29	0.33
Aural L1	0.29	0.23	0.34	0.36
Video L1	0.41	0.25	0.33	0.25
Video L2	0.40	0.31	0.48	0.33
Average all	0.35	0.26	0.36	0.32
Average L1	0.33	0.24	0.32	0.31
Standard deviation all	0.07	0.04	0.09	0.05

6.1.5.2 Hong Kong students

Table 15. – Number of phonic groups per length (seconds) used in the retelling, HK students.

Number PG/length	C.S. 1hk	C.S. 2hk	C.S. 3hk	C.S. 4hk
Written L1	0.48	0.41	0.37	0.56
Aural L1	0.49	0.38	0.52	0.60
Video L1	0.45	0.30	0.49	0.54
Video L2	0.54	0.49	0.54	0.54
Average all	0.49	0.40	0.48	0.56
Average L1	0.47	0.36	0.46	0.57
Standard deviation all	0.04	0.08	0.08	0.03

6.1.5.3 Observations

The number of phonic groups per second shows a higher rate for the L2 than for any of the retellings in the L1, in all cases but two, reflecting a higher use of pauses. It is observed that within individuals from the same group there seems to be more consistency (smaller standard deviations as a relation to average) than for rates of words per second.

6.1.5.4 Significance test: Pearson correlation for L1 retellings.

Tables 16 a & b. – Pearson Correlation for rates for L1 retellings.

Correlation	Spanish PG/second			HK PG/ second		
	Written vs. Aural	Written vs. Video	Aural vs. Video	Written vs. Aural	Written vs. Video	Aural vs. Video
Coefficient	0.93	0.03	0.13	0.61	0.98	0.5
Standard error	0.06	0.5	0.49	0.31	0.018	0.34
***t*(2)**	3.68	0.04	0.18	1.09	7.36	0.8
p	0.06*	0.97	0.9	0.39	0.02*	0.5

*reliable at the 0.05 critical alpha level **reliable at the 0.1 critical alpha level

Correlation	Spanish words/ second			HK word/ second		
	Written vs. Aural	Written vs. Video	Aural vs. Video	Written vs. Aural	Written vs. Video	Aural vs. Video
Coefficient	0.97	0.85	0.95	0.5	0.8	0.9
Standard error	0.03	0.14	0.05	0.32	0.17	0.09
***t*(2)**	5.76	2.25	4.13	0.8	1.95	3.04
p	0.03*	0.15	0.05**	0.5	0.19	0.09**

*reliable at the 0.05 critical alpha level **reliable at the 0.1 critical alpha level

The Pearson correlation tests whether there is a correlation in the data being compared. From the above results it seems that depending on what rate is used a correlation might be found or not. The only comparison where a correlation is found using both rates is in Spanish natives' written and aural data. In all cases where there was a strong positive correlation (a coefficient of 0.9 or above) this was also shown to be significant at either the 0.05 critical alpha level (indicated as *) or at the 0.1 critical alpha level (indicated **), meaning that the probability of the correlation occurring by change is either 5% or 10% respectively.

6.1.6 Pauses

Some basic information was calculated for pauses, average length (for all pauses) and length of pause by type of disfluency. These results are presented below.

6.1.6.1 Spanish natives

Table 17. – Average length of pauses (seconds), Spanish natives.

Average length (seconds)	C.S. 1sp	C.S. 2sp	C.S. 3sp	C.S. 4sp	Average	Standard deviation.
Written L1	0.61	0.27	0.91	0.47	0.56	
Aural L1	0.59	0.28	0.80	0.55	0.55	
Video L1	0.39	0.38	0.57	0.31	0.41	0.11
Video L2	0.38	0.37	0.75	0.45	0.49	0.18
Average L1	0.53	0.31	0.76	0.44		
Standard deviation L1	0.12	0.06	0.17	0.12		
Standard deviation average L1 vs. L2	0.10	0.04	0.01	0.01		

6.1.6.2 Hong Kong students

Table 18. – Average length of pauses (seconds), HK students.

Average length (seconds)	C.S. 1hk	C.S. 2hk	C.S. 3hk	C.S. 4hk	Average	Standard deviation
Written L1	0.44	0.37	0.42	0.49	0.43	
Aural L1	0.46	0.32	0.34	0.33	0.36	
Video L1	0.59	0.28	0.44	0.48	0.45	0.13
Video L2	0.71	0.52	0.71	0.78	0.68	0.11
Average L1	0.50	0.32	0.40	0.43		
Standard deviation L1	0.08	0.04	0.05	0.09		
Standard deviation average L1 vs. L2	0.15	0.14	0.22	0.25		

6.1.6.3 Observations

In Spanish natives we observe mostly shorter pauses in the retellings of the L2 than in each retelling in the L1. While in HK students we observe longer pauses in the L2 than in any of the L1 retellings, where the aural retelling shows the shortest pauses.

6.1.6.4 Significance tests for pauses video L1 and L2– paired t-test

Table 19. – Paired t-test for rates for video L1 and L2.

	Spanish natives	HK students
***t*(3)**	1.508	5.894
probability (two-tailed)	0.2288	0.0098**

*reliable at the 0.05 critical alpha level **reliable at the 0.1 critical alpha level

6.1.6.5 Observations

The difference between pause length in the L1 and L2 video retellings seems to be significant among HK students. As the expectation is that pauses in the L1 will be shorter, then the two-tailed test ought to be a one-tailed test (we divide the result by two), obtaining a probability of p = 0.0049. This is significant at the 0.05 critical alpha level. The difference in the pauses among Spanish natives does not seem to be significant.

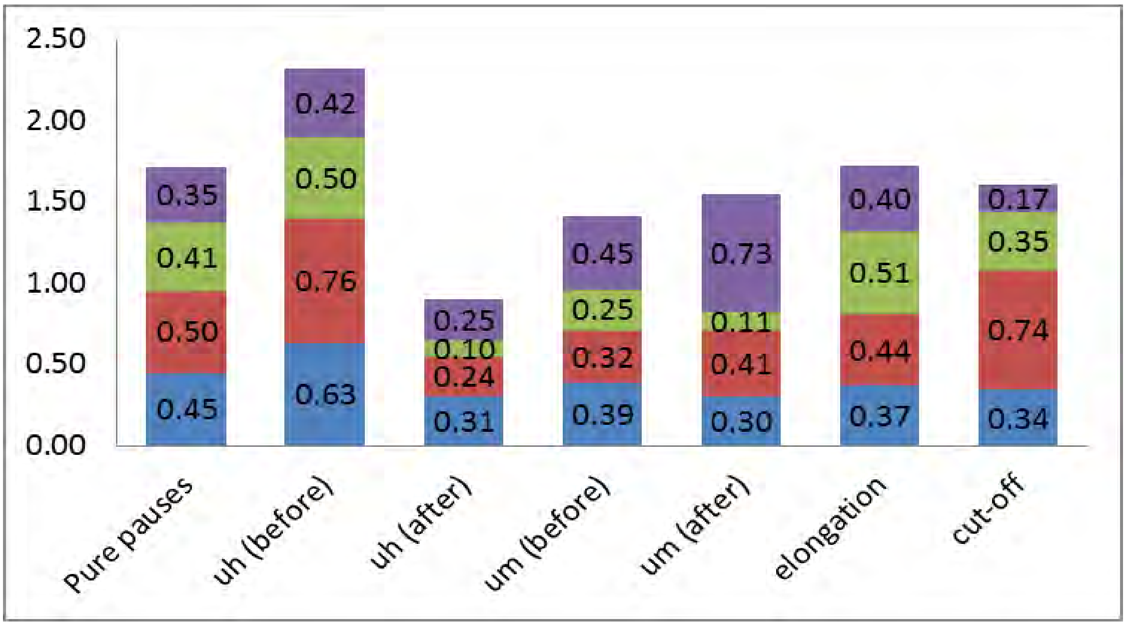
There seem to be differences in the pause length by type of retelling, although we do not find any significant differences (in paired two-tailed t-tests) among HK students in the pauses between written and aural retellings (p = 0.166 $t(3)$ = 1.811); between aural and video L1 retellings (p = 0.142 t = 1.981) or written and video L1 (p = 0.327 t = 1.168). The probabilities are close to the 0.1 critical alpha level but they are slightly above it, suggesting the results have over 10% probability of occurring by chance. Results for Spanish natives give even stronger possibilities of chance occurrences.

6.1.7 Pauses by type of disfluency

The following graphs depict the average pause length (in seconds) associated with a disfluency, for each type of retelling carried out in the L1 and the L2.

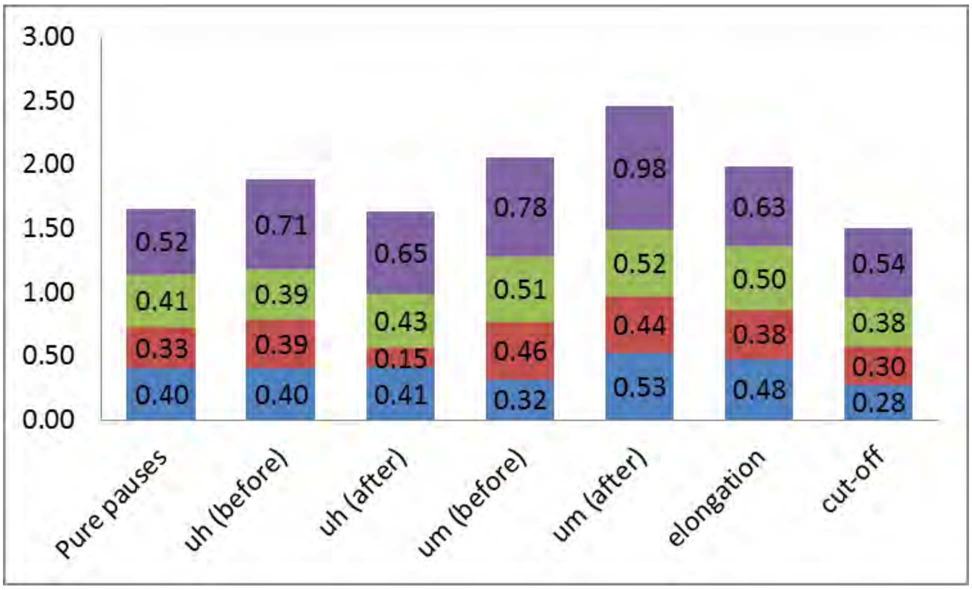
6.1.7.1 Spanish natives

Figure 11. – Length of pause by type of disfluency, Hong Kong students.



6.1.7.2 Hong Kong students

Figure 12. – Length of pause by type of disfluency, Spanish natives.



6.1.7.3 Observations

There seem to be differences in the pause length by type of disfluency, these will be analysed in more detail in the sections below (under each type of disfluency). In HK students in almost all cases the pause before ‘uh’ seems to be the longest, while in Spanish natives the pause length after the filler ‘um’ seems to be the longest. As there might be differences in the use of pauses (and disfluencies) per type of retelling, further analysis of pauses by type of disfluency will only include disfluencies from the video L1 and L2 retellings.

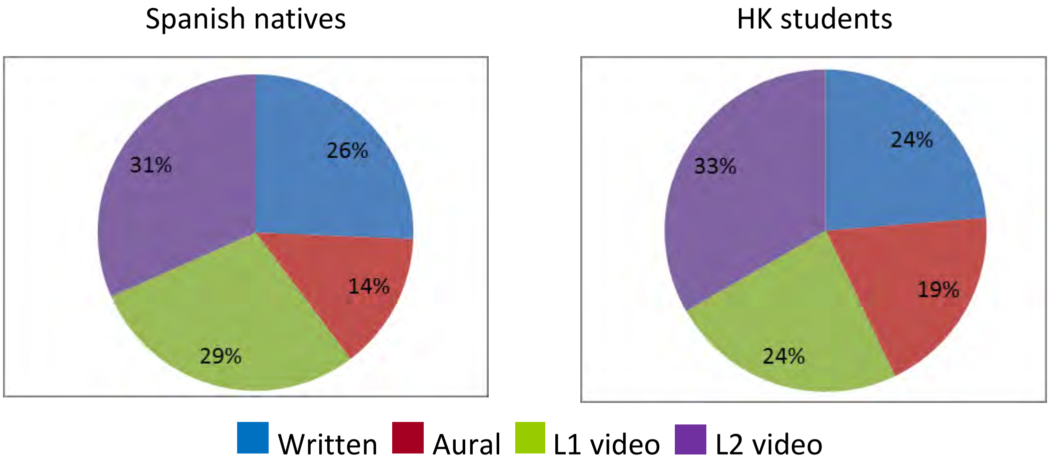
6.2 Entire Corpus – Disfluencies

This section analyses the disfluencies recorded in all the retellings, both in the L1 and the L2.

6.2.1 Disfluency numbers and rates

The following figures detail the results from the analysis of all disfluencies (occurring with and without gestures in all retellings).

Figure 13. – Percentage of total disfluencies per input.



6.2.1.1 Spanish natives

The total number of disfluencies analysed between the four participants was 555, with an additional 249 pure pauses. Out of the total 804, 31% were produced in the L2 and the rest in the L1. Figure 13 shows the percentage of total disfluencies per retelling, (percentage calculated over the total number of disfluencies for all participants).

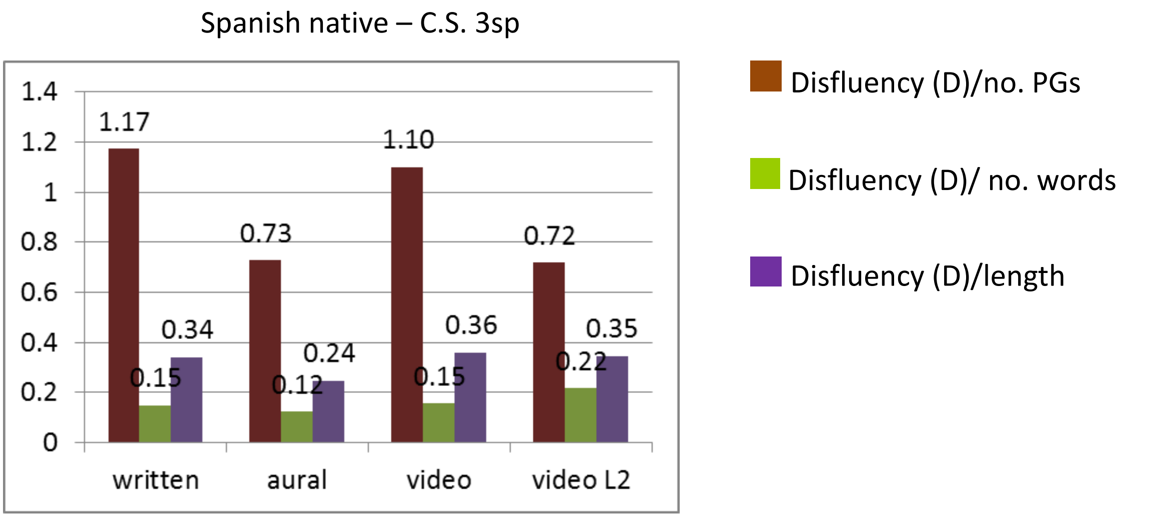
6.2.1.2 Hong Kong students

The total number of disfluencies analysed between the four participants was 1030, and an additional 327 pure pauses were also recorded. Out of the total 1357, 33% were produced in the L2 and the rest in the L1, see Figure 13.

6.2.1.3 Observations

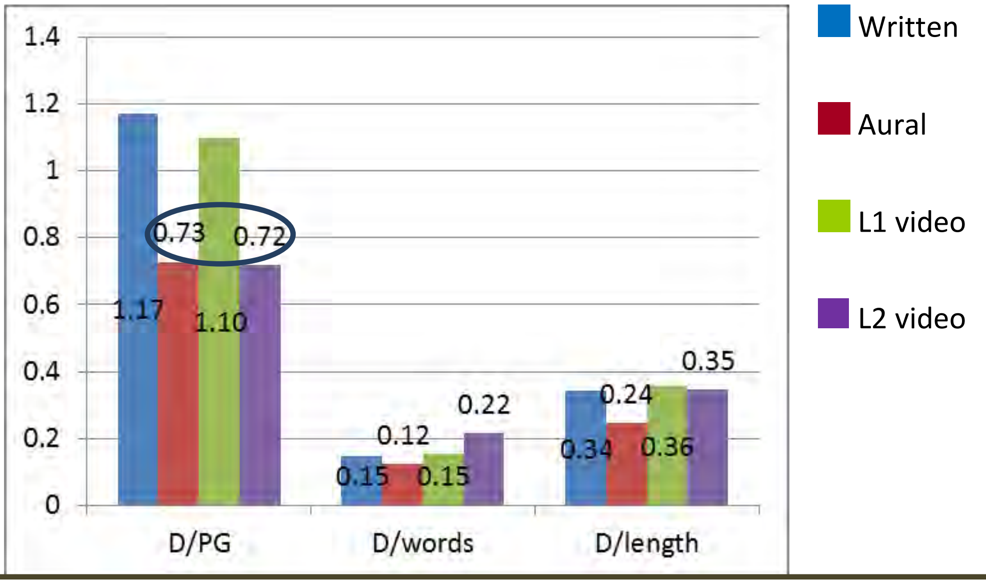
As lengths of retellings varied, the objective of presenting this information is to show that in both groups of participants there were disfluencies. In most cases HK students spoke for longer periods of time, using more words than Spanish natives, but their level of fluency was lower than Spanish natives' (more pauses). In addition Hong Kong students' proficiency level in the L2 was also lower, both conditions contributing to more disfluencies being recorded in the HK students' corpus. The number of disfluencies seems to vary by type of input, but the distribution as a percentage of the total number of disfluencies is relatively consistent in all participants with the aural retell showing the least number of disfluencies in all participants, probably because the retellings were the shortest. However, when the data was normalised by providing it as a rate to account for differences in retelling length, number of words or phonic groups it was found that patterns varied by rate. Figure 14 presents an example, where it is possible to observe how the increasing-decreasing trend is not the same in the three types of rates: using phonic group or length as the denominator shows a decrease in disfluency rates in the L2 exercise compared to the L1 video exercise, but an increase if we use number of words as the denominator.

Figure 14. – Examples of patterns in the data by rate.



If we further analyse Case Study 3sp, presenting the data by type of rate:

Figure 15. – Case study 3sp - Differences in rates by measurement.



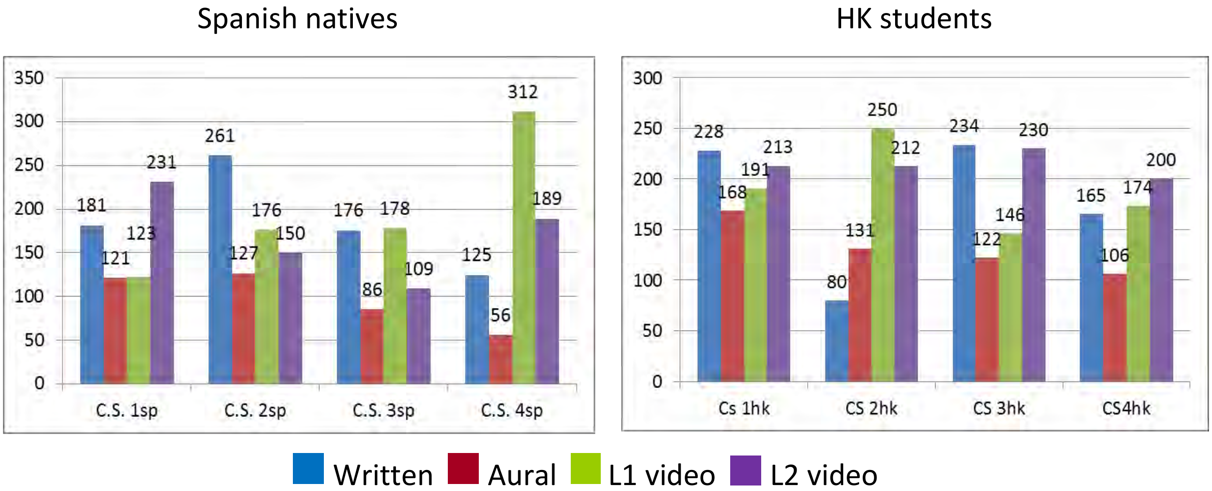
it is clear that while in the aural and L2 cases the relationship between the disfluency per phonic group and per number of words or length is different. When measured by phonic group the difference between the two retelling rates is minimal, a ratio of almost 1 to 1 – exactly 1.03 – but the difference between them as a ratio is 0.57 per number of words and 0.71 by length.

It was recognised that as the number of phonic groups will vary by length these two variables needed to be normalised, resulting in a denominator of phonic group per second.

The same is true with number of words, which is dependent on the length of the retellings. From the above results and from results of similar studies (explained in the discussion) it was decided to use phonic group per second as the normalising denominator hereafter.

6.2.2 Disfluency in the L1 vs. L2

Figure 16. – Disfluencies per phonic group/second for all participants and all input types (including pauses).



6.2.2.1 Spanish natives

Figure 16 shows the rate of disfluency (including pure pauses) per phonic group per second for all participants. The expectation was for a significant increase in the number of disfluencies when the participants were speaking in the L2, but the results do not confirm this. Only in C.S. 1sp is the rate of disfluency greater when speaking in the L2 than in all other cases of L1 speech.

6.2.2.2 Hong Kong students

Among Hong Kong students we also failed to observe a consistent increase in disfluency in the L2, only in C.S. 4hk is the rate of disfluency greater when speaking in the L2 than in all other retellings using the L1.

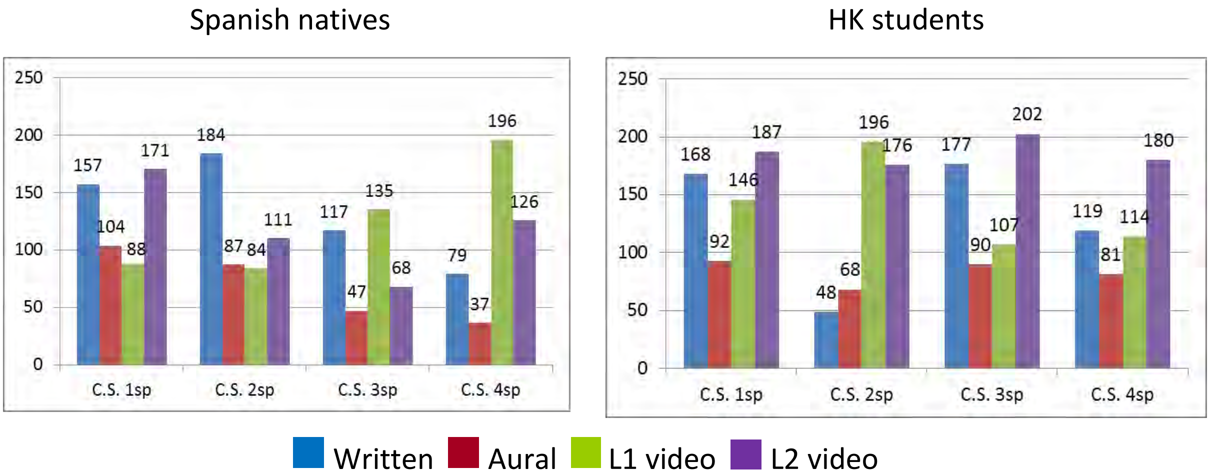
6.2.2.3 Observations

As the above data included pure pauses[69] (and the phonic groups are dependent on the pauses) it was decided to eliminate the pure pauses from the overall disfluency count and calculate the disfluency rates again.

6.2.3 Disfluency in the L1 vs. L2 excluding pauses

The data was recalculated taking out the unfilled pauses, see Figure 17.

Figure 17. – All disfluencies, excluding pauses, per phonic group per second for all participants and all input types.



6.2.3.1 Spanish natives

Once pauses have been excluded C.S. 1sp is still the only case showing more disfluencies per phonic group when using the L2. The patterns are very similar with or without pauses.

[69] Pure pauses are those not followed or preceded by a disfluency. These might not be disfluencies at all (just pauses to breath) and cannot be classified with any degree of confidence as such unless followed by a repair or preceded by an error. Therefore they are not included in further analysis of disfluencies.

6.2.3.2 Hong Kong students

The same is true for the patterns for HK students, except that the differences by type of retelling in C.S. 1HK and 2HK seem greater. Once again we cannot categorically state that there are more disfluencies in the L2 once the pauses have been excluded.

6.2.3.3 Observations

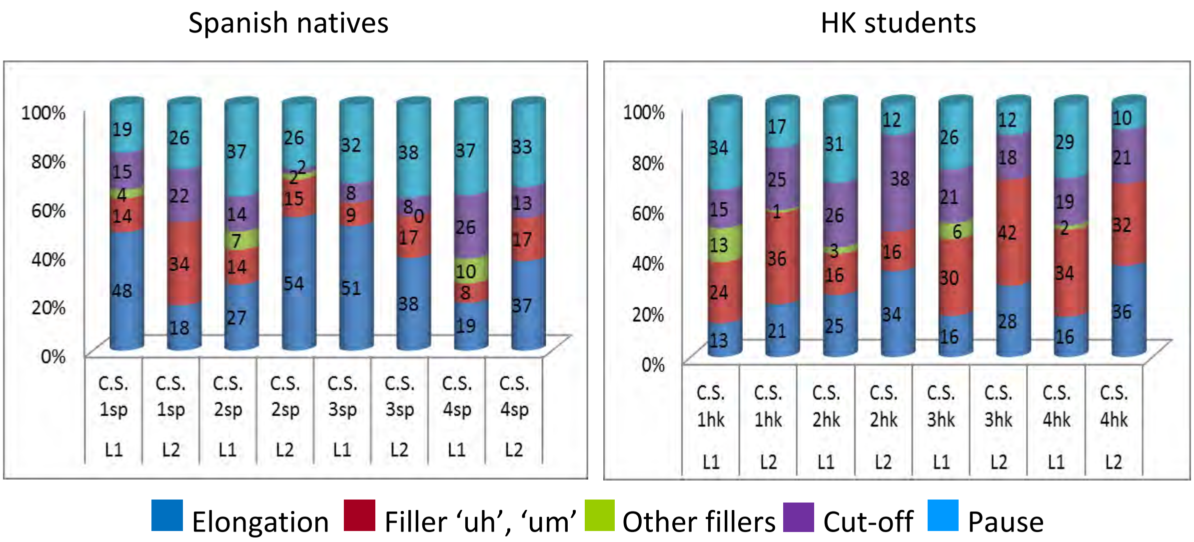
In neither group can we observe a pattern for rate of disfluency. It was speculated that there might be differences by gender as we observed (see Figures 16 and 17) that the rate of disfluency seems to be greater for the first two Spanish native participants (C.S. 1sp and 2sp), who happen to be the female participants. However in the Hong Kong students we cannot see a noticeable difference in the number of disfluencies by gender (C.S. 1hk and 2hk were the female participants).

In all participants we observe different rates for each type of retelling, even within the L1. The significance or the reason for these differences falls outside the scope of this study (but it will be noted for future research).

6.2.4 Type of disfluency

To test whether any participant showed a preference for a particular type of disfluency the data was combined to compare the type of disfluency by participant. All disfluencies from all L1 and L2 retellings were totalled and no obvious differences were found by participant or by gender, see Figure 18.

Figure 18. – Number (on bar) and percentage (y axis) of disfluency by type by subject.



6.2.4.1 Spanish natives

Comparing the instances of disfluencies by language, L1 vs. L2 the only reliable pattern is that in all cases there is an increase in the percentage of fillers used in the L2 (English).

6.2.4.2 Hong Kong students

We observe fewer pure pauses in the L2. This is normal as other types of disfluencies tend to occur with a pause, but there seem to be more elongations in the L2 than in the L1.

6.2.4.3 Observations

We believe there might be some significant differences in the type of disfluency used by language and by group of participants. This is further explored in the second part, where the data for the video retellings L1 and L2 is compared (thus excluding the variable related to the type of input, which might be having a significant effect).

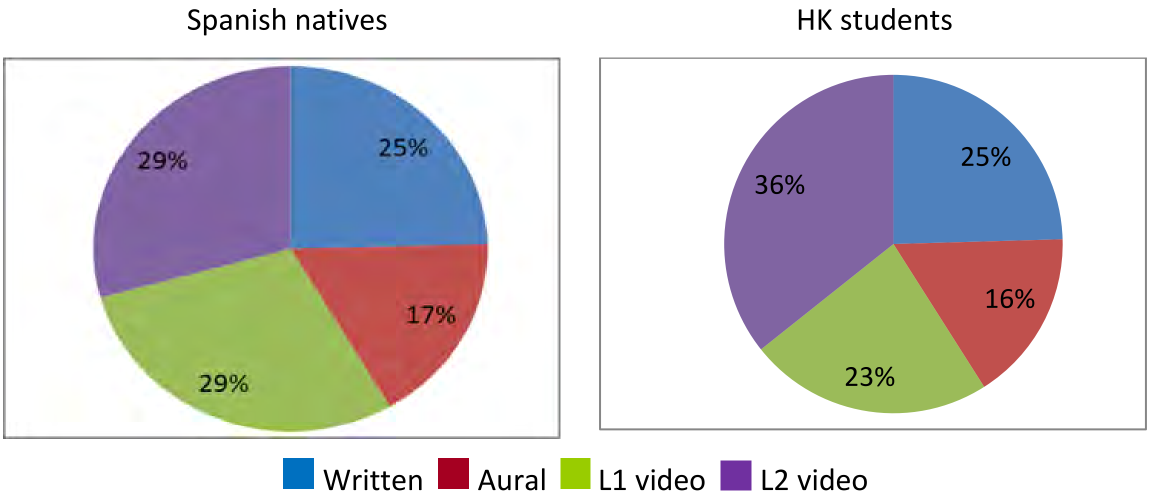
6.3 Entire Corpus – Gestures

This section analyses the gestures recorded in all the retellings, both in the L1 and the L2, with and without disfluencies.

6.3.1 Gesture numbers and rates

The analysis of gestures initially included all gestures in all retellings (occurring with and without disfluencies). Gestures were counted by gesture phrase (including preparation, stroke and retrieval phases).

Figure 19. – Percentage of total gestures by type of input.



6.3.1.1 Spanish natives

The total number of gestures recorded was 594. As with the disfluencies, the only noticeable difference between the percentages observed per type of input is a lower percentage rate of gestures during the aural exercise; the video exercises in the L1 and L2 both generated 29% of all the gestures.

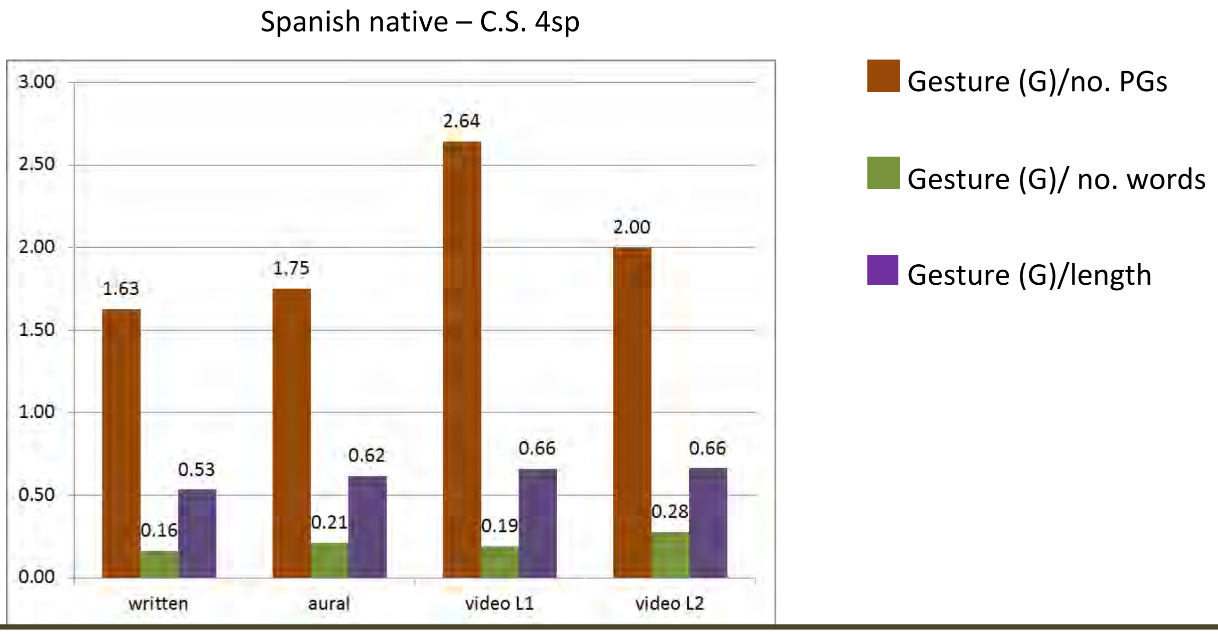
6.3.1.2 Hong Kong students

The total number of gestures recorded was 888, more gestures were recorded during the L2 retellings than in any other type of retelling (36% of all recorded gestures), and fewer for the aural (shorter retellings).

6.3.1.3 Observations

As some of the differences could be related to the length of the retelling, the aural retellings being shorter than the others, it was again necessary to normalize the data by providing it as a rate. As with the disfluency rate calculations, the first challenge was to determine a rate that would allow us to compare gesture rates among the various participants by type of input. Gesture by phonic group, number of words and length was calculated. As observed with disfluency rates, different patterns are seen across the various retellings per participant. We took an example to illustrate the point, C.S. 4sp, see Figure 20.

Figure 20. – Examples of patterns in the gesture data by rate.

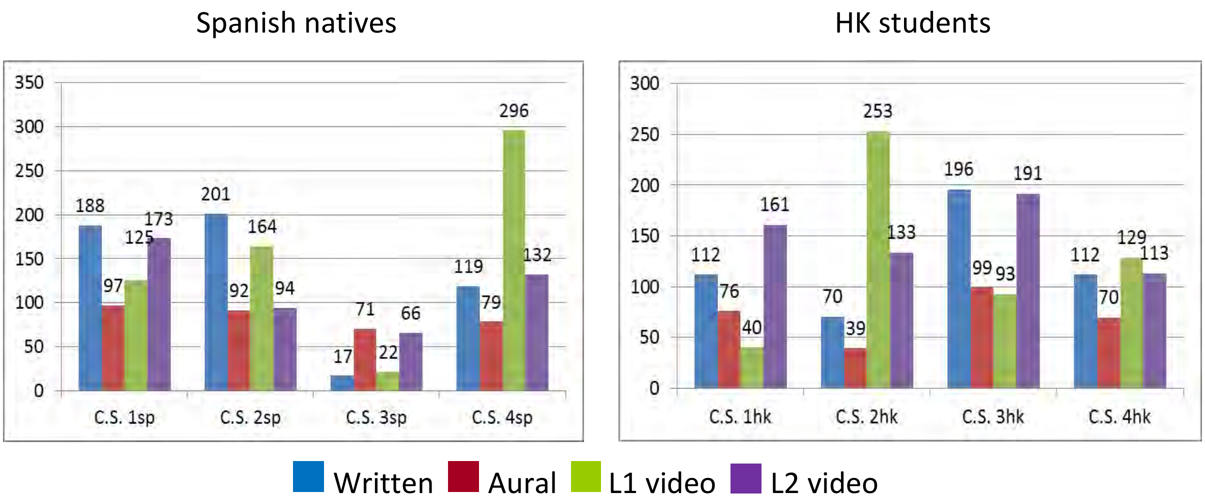


In C.S. 4sp, we can observe that the pattern in the gesture per PG and word increases from written to aural, then increases to video L1 and decreases to video in L2. The pattern in the gesture rate per length is increase, increase, stays the same. In gesture per word it is increase (written to aural), decrease (to L1), and increase (to L2). Thus leading to different conclusions depending on the rate used.

6.3.2 Gesture rate per phonic group per second

To neutralise two of these variables when normalising the data we used gesture rate per phonic group per second, to be consistent with the analysis done on disfluencies. Also – as mentioned in the discussion – taking into account that we were comparing L1 and L2 utterances, phonic groups might be a more relevant denominator to normalize the data (as the number or words and length of retellings is more likely to vary between the L1 and the L2 than number of phonic groups).

Figure 21. – Number of gestures per phonic group per second.



6.3.2.1 Spanish natives

There appear to be no obvious patterns in the data of the Spanish natives.

6.3.2.2 Hong Kong students

There appear to be no obvious patterns in the data of the HK students.

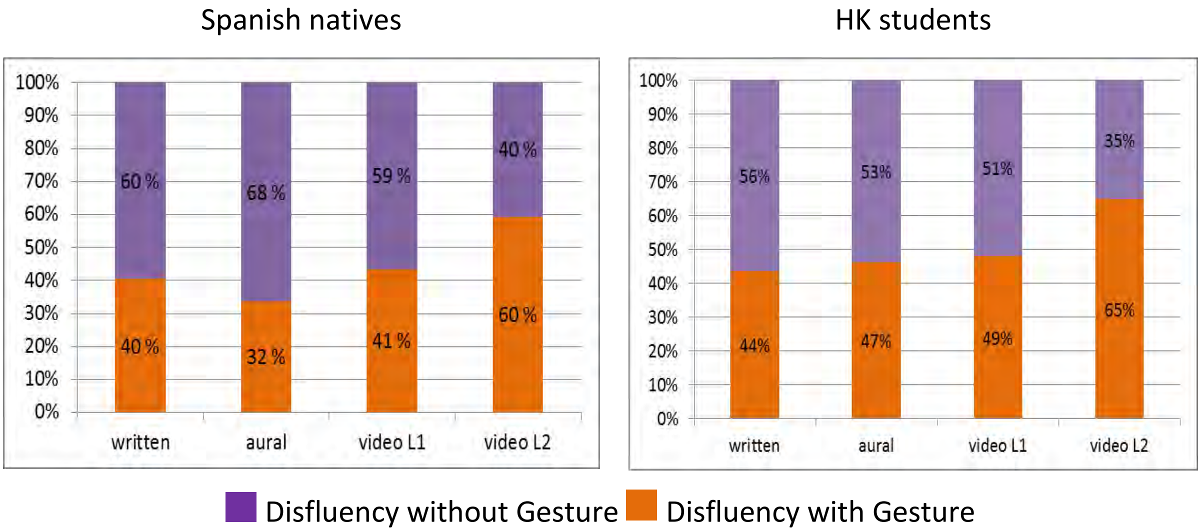
6.3.2.3 Observations

It is not possible to conclude that there are higher rates of gesturing with any particular type of retelling.

6.3.3 Gestures during fluent and disfluent speech

The lexical Retrieval hypothesis suggests that gesturing might be a compensatory strategy when speech fails, thus suggesting a causal link between gesture and speech. If this is the case we ought to observe more disfluencies with gestures than without them. Figure 22 gives the percentage of disfluencies that occurred (in all the retellings) with and without gestures.

Figure 22. – Percentage of disfluency with and without gesture.



6.3.3.1 Spanish natives

Spanish natives show a tendency not to use gesture during disfluency (62% on average) when using the L1, but in the L2 there are more disfluencies accompanied by gestures (60%).

6.3.3.2 Hong Kong students

There is also a prevalence for disfluencies without gestures in the L1 (53% on average) and a higher number of disfluencies accompanied by gestures in the L2 (65%).

6.3.3.3 Significance: One sample t-test between percentages

Table 20. – One sample t-test between percentages disfluency with and without gestures.

	Spanish natives			HK students		
	t statistic	**Degrees of freedom**	**probability**	**t statistic**	**Degrees of freedom**	**probability**
Written L1	2.371	178	0.069*	2.065	291	0.0398*
Aural L1	4.192	117	0.001**	0.831	190	0.4072
Video L1	2.062	126	0.04*	0.331	272	0.7413
Video L2	2.327	129	0.0215*	5.206	273	0.000**

*reliable at the 0.05 critical alpha level **reliable at the 0.1 critical alpha level

6.3.3.4 Observations

Some gesture phrases had more than one disfluency, although the incidence is not high among Spanish natives, with 16 cases of two disfluencies in one gesture phrase (most, except for two cases, recorded in the L2) and 4 cases of three disfluencies in one gesture phrase (all in the L2). Among HK students is even rarer.

The results of the significance test suggest that the differences between percentage of disfluencies with gestures and without is significant (in almost all cases) and unlikely to have occurred by chance (except for aural and video L1 retellings in HK students, in all other cases the probability of a chance result is less than 10%).

6.3.4 Overall types of gestures used

There seems to be a difference in the use of disfluency with and without gestures depending on the language being used. To further test these differences gestures were subdivided depending on their function into either gestures of a pragmatic nature, used to keep the rhythm or stress particular utterances, and all others which could be considered iconic in some aspect (we would like to stress again the subjectivity of this part of the analysis, as mentioned in the section describing the analysis). Gestures which were related to personal grooming (scratching, touching the hair, etc.) were also annotated but not included in the analysis. Figure 23 gives the total instances of gestures by function (all gestures in all retellings), pragmatic (p) and iconic (i). To clarify the data the actual numbers were also converted into percentages (based on the total number of gestures recorded in the same type of retelling). Tables 21 and 22 provide the data by participant while Table 23 summarises the results as percentages by group for all retellings.

Figure 23. – Average percentage (y axis) and average number (on the bar) of gestures per type (all retellings and gestures).

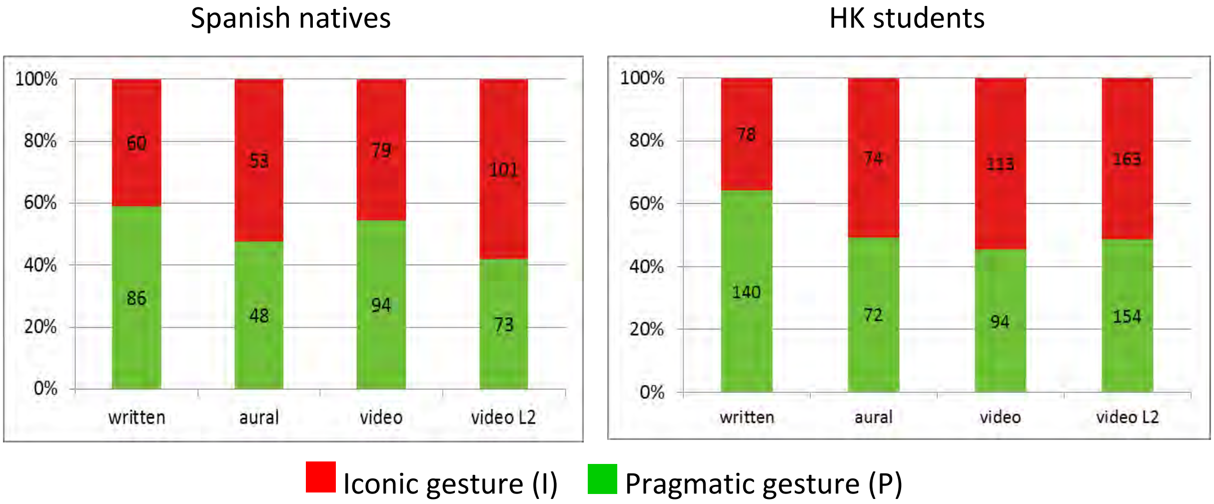


Table 21. – Number of gesture per type (all gestures), Spanish natives.

	C.S. 1sp		C.S. 2sp		C.S. 3sp		C.S. 4sp		
	P	**I**	**P**	**I**	**P**	**I**	**P**	**I**	**total**
Written L1	37	18	21	26	5	0	23	16	146
Aural L1	14	14	5	16	15	9	14	14	101
Video	24	27	18	23	6	1	46	28	173
Video L2	34	35	11	18	13	19	15	29	174
Total L1	75	59	44	65	26	10	83	58	420
Average. L1	25	20	15	22	9	3	28	19	140

Table 22. – Number of gesture per type (all gestures), HK speakers.

	C.S. 1hk		C.S. 2hk		C.S. 3hk		C.S. 4hk		
	P	I	P	I	P	I	P	I	total
Written L1	48	6	15	14	46	26	31	32	218
Aural L1	32	5	11	4	18	34	11	31	146
Video L1	12	6	41	34	15	30	26	43	207
Video L2	45	42	21	45	59	44	29	32	317
Total L1	92	17	67	52	79	90	68	106	571
Average. L1	31	6	22	17	26	30	23	35	

Table 23. – Type of gesture as a percentage by group.

%	Spanish students		HK students	
	P	I	P	I
Written L1	59%	41%	64%	36%
Aural L1	48%	52%	49%	51%
Video L1	54%	46%	45%	55%
Video L2	42%	58%	49%	51%
Total L1	54%	46%	53%	47%

Note: Distribution of types of gesture within each retelling (total sum P and I for each = 100)

6.3.4.1 Spanish natives

There seems to be a prevalence of iconic gestures during the L2 retellings, making up 58% of the total number of gestures performed in the L2, but during the L1 retellings we observe a slight prevalence of gestures of a pragmatic nature (an average of 54% for the three types of retellings). We observe that during the aural retelling there is a prevalence of iconic gestures.

6.3.4.2 Hong Kong students

There is almost the same percentage of iconic and pragmatic gestures used during the L2, and a preference (53%) for pragmatic gestures during the combined L1 retellings. Only in the written retellings is there a clearer tendency to use pragmatic gestures (64% of all gestures).

6.3.4.3 Observations

These results were the average results for all gestures, but the question remains whether during a disfluency there were more pragmatic gestures, the type used when a speaker does not recall a word and waves the hand, or twitches the fingers, in an attempt to recall the word. To answer this question the next analysis included only gestures that had taken place during a disfluency, excluding gestures that fell solely during pauses (as these could be interpreted as emblems).

6.3.5 Types of gestures used during disfluencies

This part of the analysis included 265 cases of disfluencies with gestures in the Spanish natives' corpus and 572 from the HK students' corpus, excluding pauses.

Figure 24. – Average percentage (y axis) and average number (on the bar) of gestures per type (all retellings gestures during disfluencies).

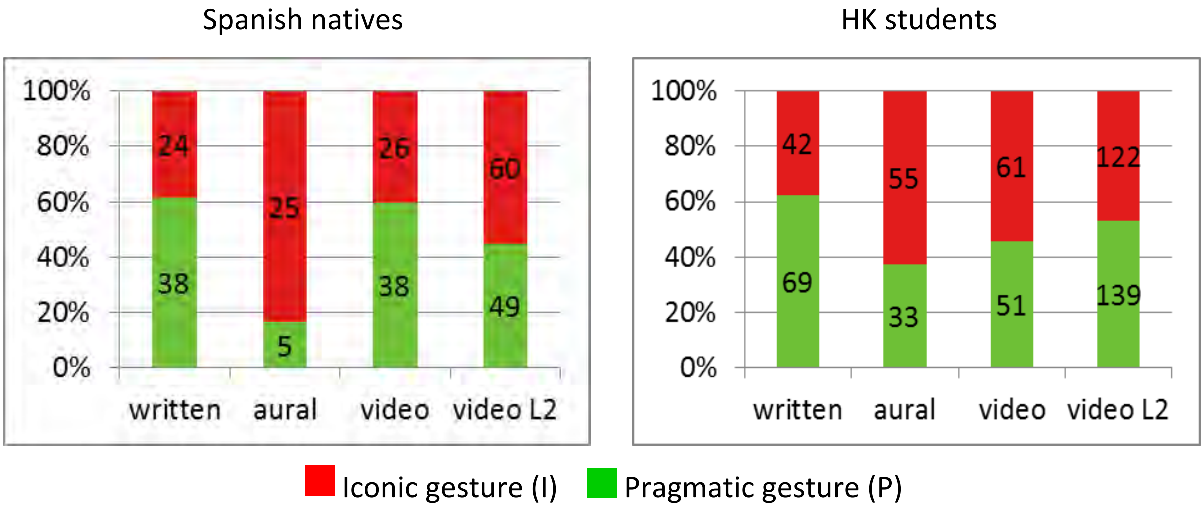
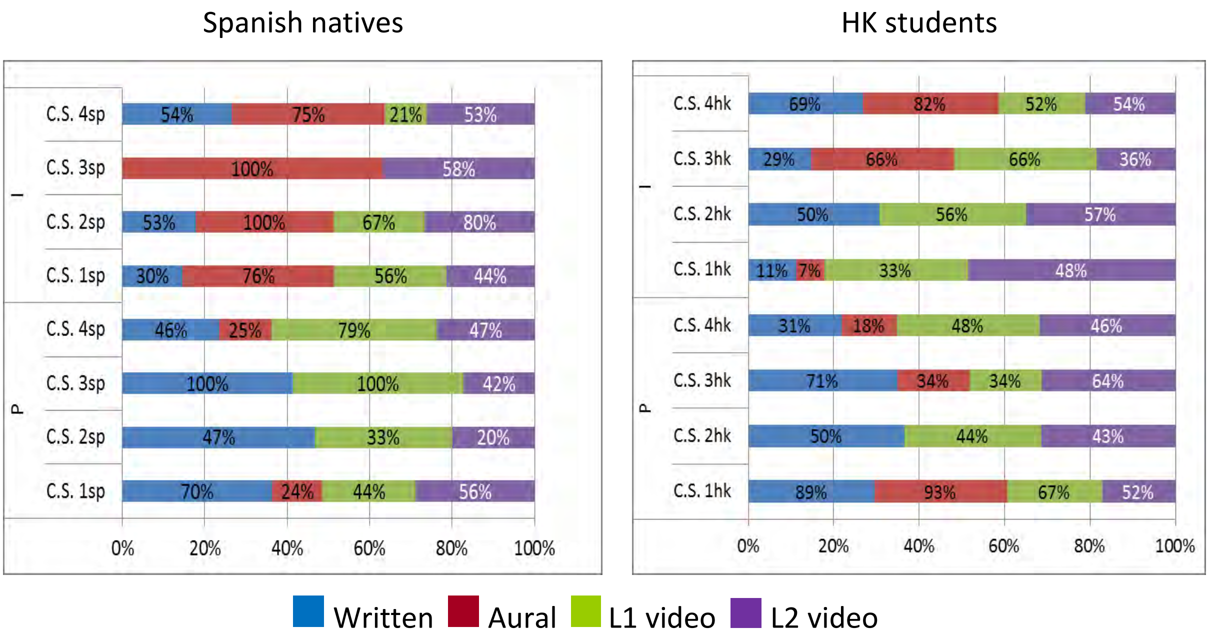


Table 24. – Type of gesture occurring with disfluencies by percentage by group.

%	Spanish students		HK students	
	P	I	P	I
Written L1	61%	39%	62%	38%
Aural L1	17%	83%	38%	63%
Video L1	59%	41%	46%	54%
Video L2	45%	55%	53%	47%
Total L1	52%	48%	49%	51%

Note: Distribution of types of gesture within each retelling (total sum P and I for each = 100).

Figure 25. – Percentages of gesture type (pragmatic, P, or Iconic, I) during disfluency.



6.3.5.1 Spanish natives

When analysing the data as percentages (for the total number of disfluencies with gestures each subject used in each retelling), the only consistent pattern among all participants is a preference for iconic (83%), rather than pragmatic gestures, during the aural retellings.

6.3.5.2 Hong Kong students

There is a prevalence of iconic gestures during the aural retellings (63%).

6.3.5.3 Observations

Once more, we observe that there seem to be differences by type of input (in particular with the aural input) and believe that it is necessary to compare the retellings L1 and L2 based on the same type of input. The fact that aural retellings show a higher use of iconic gestures might be significant: the aural input retelling was the hardest for all participants, and might have been the hardest to render into images of any kind (the written text could be associated with the image of the written word).

Comparing the distribution of pragmatic and iconic gestures with and without disfluencies gives similar breakdowns for the Spanish natives and HK students (in L1 retellings) but a reversed picture for the HK students in the L2 is observed (Tables 23 and 24).

6.3.6 Disfluencies with gestures and pauses

This part of the analysis looked at 262 cases of disfluencies with gestures in the Spanish natives' corpus (three fillers other than 'uh' and 'um' were discounted) and 565 from the HK students' corpus (fillers other than 'uh' and 'um' were excluded), excluding pauses.

Figure 26. – Number (bar) and percentage (axis) of disfluencies with gesture followed by a pause in the L1 (all retellings).

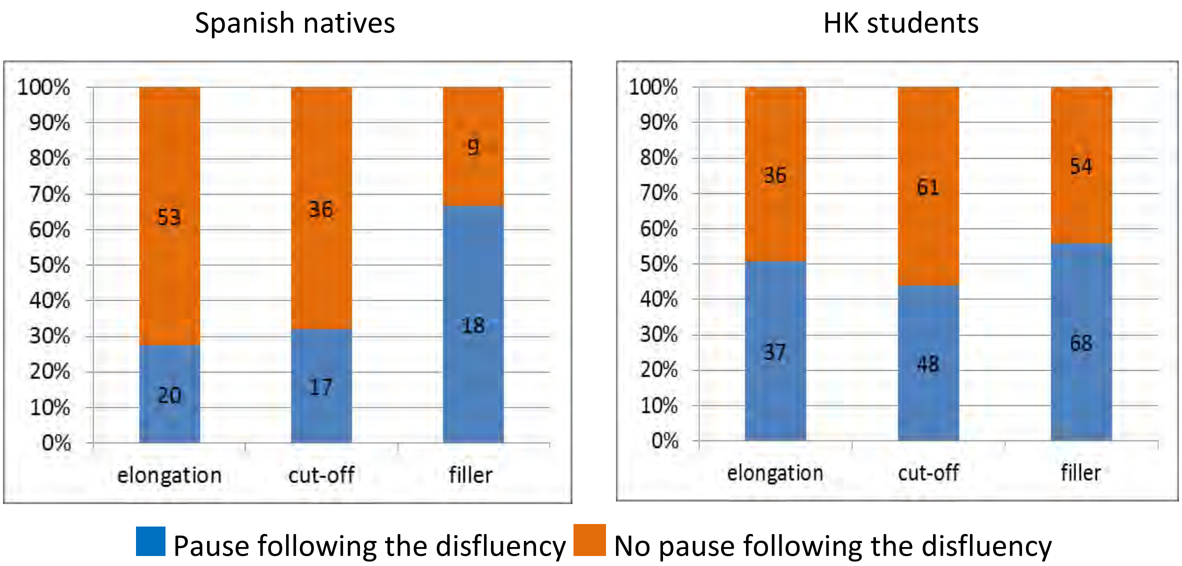


Figure 27. – Number (bar) and percentage (axis) of disfluencies with gesture followed by a pause in the L2.

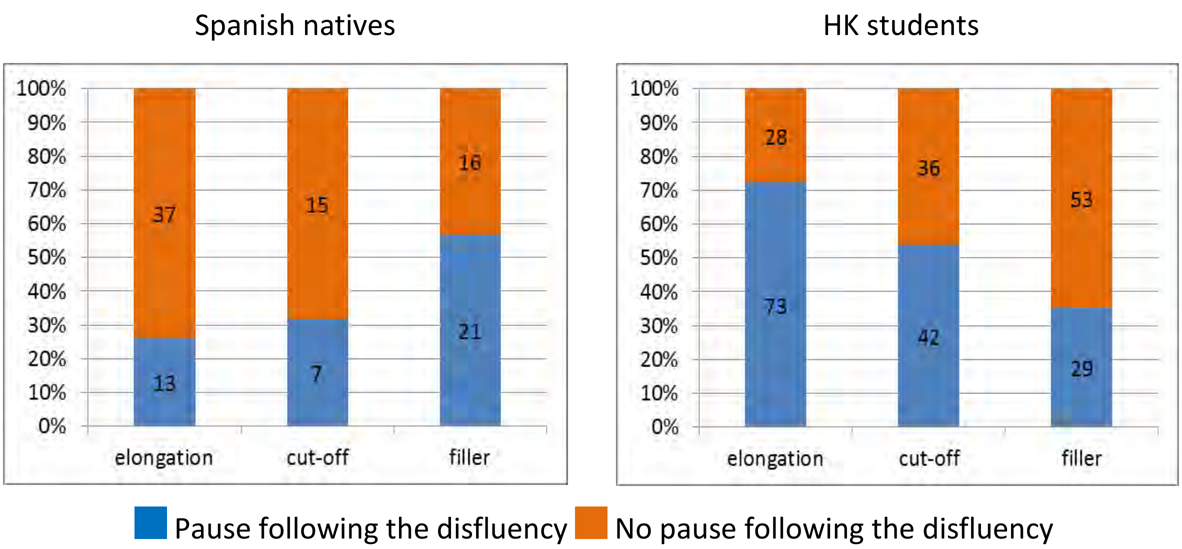


Table 25. – Summary of pauses after disfluencies as a percentage.

	Spanish natives				HK students			
	L1		L2		L1		L2	
	Pause	No pause	Pause	No pause	Pause	No pause	Pause	No pause
Elongation	27%	73%	26%	74%	51%	49%	72%	28%
Cut-off	32%	68%	32%	68%	44%	56%	54%	46%
Filler	67%	33%	57%	43%	56%	44%	35%	65%

Table 26. – Two-tailed t-test, significance test for pauses after disfluencies.

	Spanish natives				HK students			
	L1		L2		L1		L2	
	Pause or not		Pause or not		Pause or not		Pause or not	
	t-statistic	*p*	t-statistic	*p*	t-statistic	*p*	t-statistic	*p*
Elongation	(72) 4.43	0.00*	(49)3.869	0.0003*	(72) 0.17	0.86	(100) 4.9	0.00*
Cut-off	(52) 2.89	0.007*	(21) 1.81	0.085**	(108) 1.262	0.21	(77) 0.71	0.48
filler	(26)1.88	0.07**	(36) 0.86	0.4	(121) 1.335	0.184	(81) 2.85	0.006*

*reliable at the 0.05 critical alpha level **reliable at the 0.1 critical alpha level

Note: in brackets degrees of freedom.

6.3.6.1 Observations

Spanish natives show similar breakdowns of disfluency with or without pause regardless of the language (L1 or L2), and in most cases this breakdown is reliable (with a less than 10% probability of occurring by chance), but in HK students, whose disfluencies are more likely to be followed by pauses than those of Spanish natives', we observe a more complex relationship: elongations in the L2 are more likely to be followed by a pause and fillers are less likely. However in HK students only the data on elongations and fillers in the L2 seems to be significant (with a 95% probability of not occurring by chance).

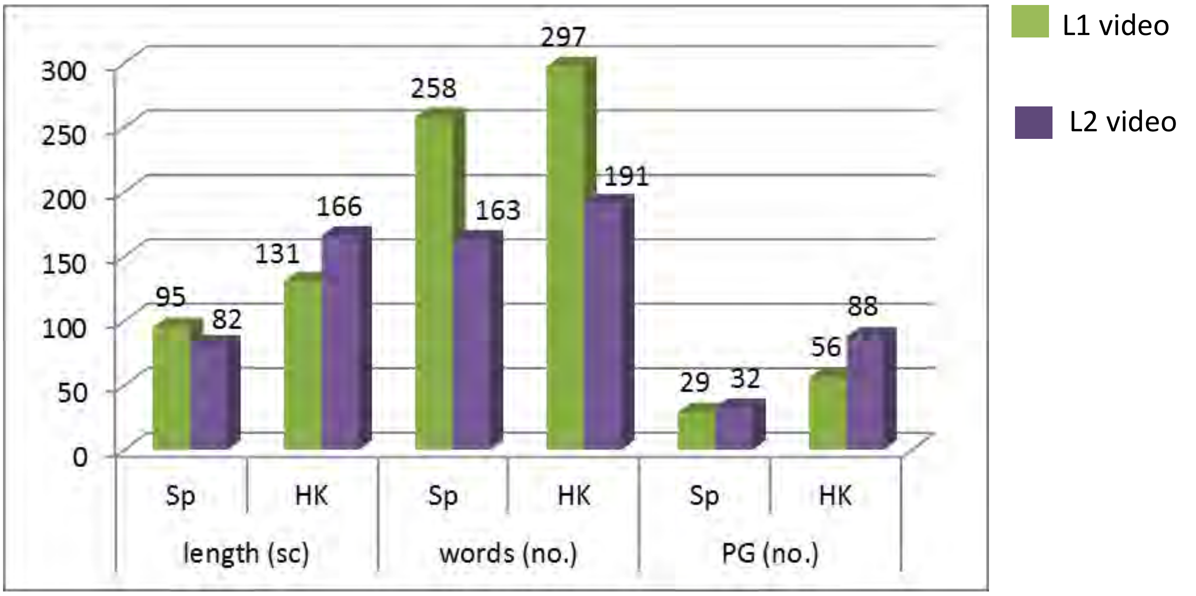
6.4 Video Corpus - Disfluencies

This section analyses the disfluencies and gestures recorded in the retellings based on the video input, both in the L1 and the L2.

6.4.1 Basic data for video L1 and L2

On reviewing the analysis above it is clear that there are inconsistencies in the pause length and type of gesture results that might be attributable to the type of input (written, aural or video). It is known that the nature of content can lead to disfluency differences in the same individual (Bortfeld et al., 2001), and perhaps the type of input on which the production of the content is based is also affecting the participants' disfluency production. In order to be able to cancel the possible effects of this variable the video data alone, in both the L1 and L2 was analysed next. Retellings of the video in both the L1 and L2 provided data that fell between the results obtained for the other retellings in terms of the length, number of words and phonic groups - having the additional benefit of excluding extreme results. As it was the third retelling in all cases, participants ought to have been quite comfortable with the process by this point and any negative affective issues should have been minimized. Figure 28 gives a summary of the basic data for the video L1 and L2 retellings.

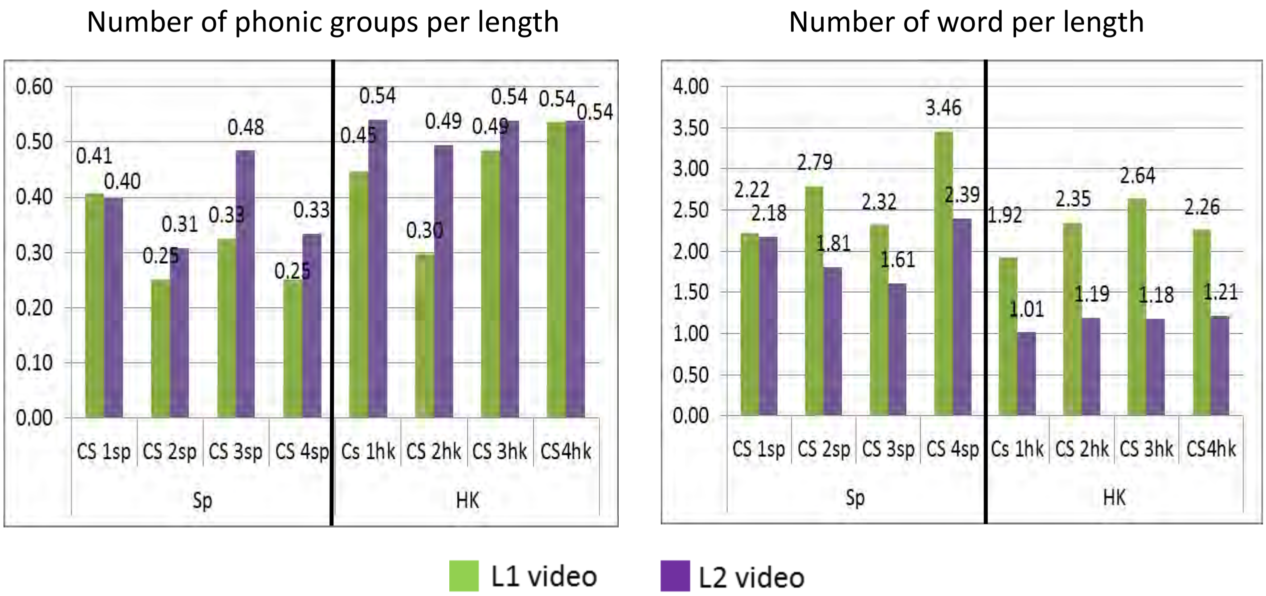
Figure 28. – Comparison of video retelling data.



6.4.1.1 Observations

HK students produced longer retellings in both languages than Spanish natives, and as expected they also used more words and phonic groups as a total. However, the actual lengths, number of words and phonic groups are not really relevant to calculating their correlation, as the retellings of the HK students in the L2 were cut short, and it was necessary to normalise the data. Figure 29 presents the ratios word by length (seconds) and phonic group by length (seconds) for each participant. As can be seen, similar rates are observed among most participants within each group by L1 and L2 data. We believe this is not directly related to the content of the story. If it was related to the content we ought to observe mirroring rates: an increased rate from L1 to L2 retellings in Spanish natives should correspond to a decrease in the same rate from L1 to L2 retellings in HK students, as Spanish natives retold the first 3 minutes in Spanish and the last 3 minutes in English, while HK students started retelling the first part in English.

Figure 29. – Phonic groups per length and word per length rates in L1 and L2 video retellings.



It was observed that the differences between the ratios of phonic group per second are smaller and more consistent in both the L1 and the L2, suggesting less variability based on the language. The differences in the word per second ratios (L1 vs. L2) are larger among HK students than Spanish natives and the difference between the two groups is also large, see Table 27.

Table 27. – Comparison of PGs/second and words/second average ratios.

	PGs/length		**words/length**	
	Spanish natives	**HK students**	**Spanish natives**	**HK students**
Video L1	0.31	0.44	2.70	2.29
Video L2	0.38	0.53	2.0	1.15

Table 28 presents the Pearson correlation coefficients between the L1 and the L2 ratios calculated by both number of phonic groups per second and number of words per second.

Table 28. – Pearson Correlation for rates for video L1 and L2.

	Spanish natives PG/second	**HK students PG/second**	**Spanish natives words/second**	**HK students word/second**
Correlation coefficient	0.633	0.934	0.57	0.764
Standard error	0.3	0.063	0.337	0.208
***t*(2)**	1.156	3.712	0.981	1.677
p	0.3672	0.0655*	0.4299	0.2355

*reliable at the 0.05 critical alpha level **reliable at the 0.1 critical alpha level

The strongest correlation between L1 and L2 video retellings found was in HK students' phonic group per second rates (0.934, suggesting a very strong correlation). With a probability of p = 0.0655, this is considered reliable at the 0.1 critical alpha level (90% probability of not being a chance result), therefore we conclude that there is a correlation between the video L1 and L2 retellings, as expected, as we were comparing retellings by the same individuals, suggesting that different rates do give different results. It also suggests that phonic group per second might give more reliable results.

The difference in the ratios is smaller and more consistent in the phonic group per second rate, suggesting less variability based on the fluency of the language spoken. Table 29 provides the results from the t-test analysis showing the significance of the difference between the L1 and L2 ratios by phonic group per second and words per second by group (comparing the differences between the four Spanish native participants and the four HK students). The results of the test indicate that the difference between L1 and L2 among HK students is quite high (3.313) in the word per second rates, and we can reject the null hypothesis of H_o = the difference between the two ratios is 0, as the probability of seeing a difference of less than 3.313 is quite low (0.0453), making the results significant to a critical alpha level of 0.05. That is we are 95% confident that the large difference in the rates calculated between the L1 and the L2 is significant. For Spanish students and for differences calculated by phonic rate per second the data is not significant but the differences between the L1 and the L2 are smaller.

Table 29. – Paired t-test results for L1 and L2 retellings per participant as PGs/length and words/length ratios.

	PGs/length		**words/length**	
	Spanish natives	**HK students**	**Spanish natives**	**HK students**
***t*(3)**	2.077	1.06	1.239	3.313
p	0.129	0.3668	0.3034	0.0453**

** Significant at 95%; critical alpha level = 0.05

6.4.2 Disfluencies

In Figures 16 and 17 the disfluency rates were presented by phonic group per second. The data was not very conclusive and it is not obvious that there are more disfluencies in the L2, as might be expected. To test whether there is a correlation between disfluencies in video L1 and L2 retellings from participants of the same group a Pearson one-tailed probability correlation analysis was carried out (only disfluencies without pauses were included in this test). There doesn't seem to be a strong correlation: in the case of the Spanish natives' data the correlation is negative, indicating a mirroring rates of disfluency in L1 and L2, but it is not significant (correlation coefficient = -0.242; standard error=0.471; $t(2)$ = -0.352; p = 0.379), the HK students' data also indicates a negative correlation, but not significant (correlation coefficient =-0.666; standard error = 0.278; $t(2\ )$= -1.262; p = 0.165). In both cases the probability of the Pearson's t-tests suggests low reliability of the correlation.

One of the objectives of this study was to determine whether either group produced any particular type of disfluency when speaking the L1, therefore we compared type of disfluencies by language, L1 versus L2. As the calculations on rate of disfluency did not seem to be significant, the data was calculated as an average number of each type of disfluency per group and as a percentage of the total number of disfluencies observed within each group. The Spanish natives' corpus included 141 cases analysed in the L1 and 176 in the L2, pure pauses were not included and fillers other than 'uh' and 'um' and their variations were eliminated (there were 29 in the L1 and 1 in the L2) to limit the analysis. The HK students' corpus included 215 cases in the L1 and 393 in the L2, pure pauses were again not included and fillers other than 'uh' and 'um' and their variations were eliminated, (46 in the L1 and 1 in the L2). The average number of disfluencies by speaker by language is presented in Figures 30 (total numbers) and 31 (percentages). There is an increase in both average number and percentage of disfluencies in both groups going from the L1 to the L2 (although both groups used less words in the L2, but those retellings were longer among HK students).[70]

Figure 30. – Average number of disfluencies per speaker in L1 and L2 video.

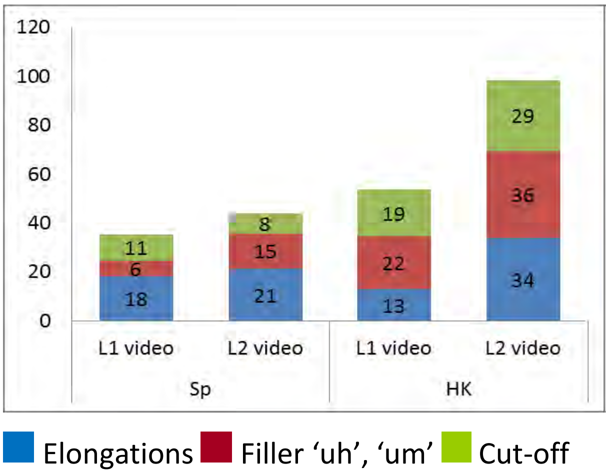
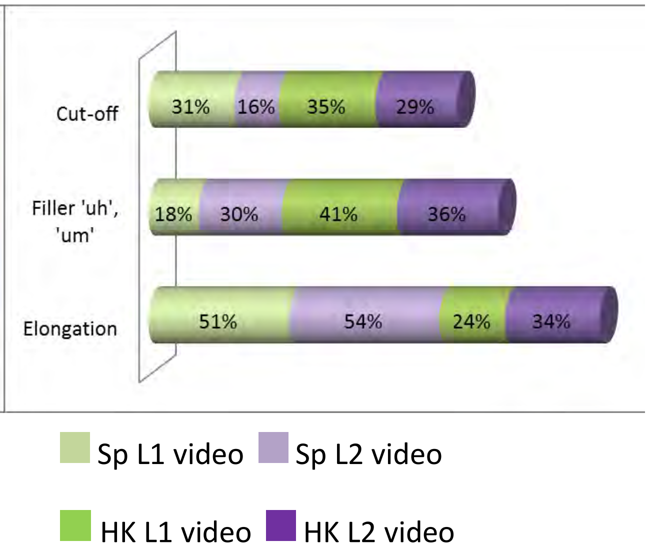


Figure 31. – Average percentage of disfluencies by speaker in L1 and L2 video.



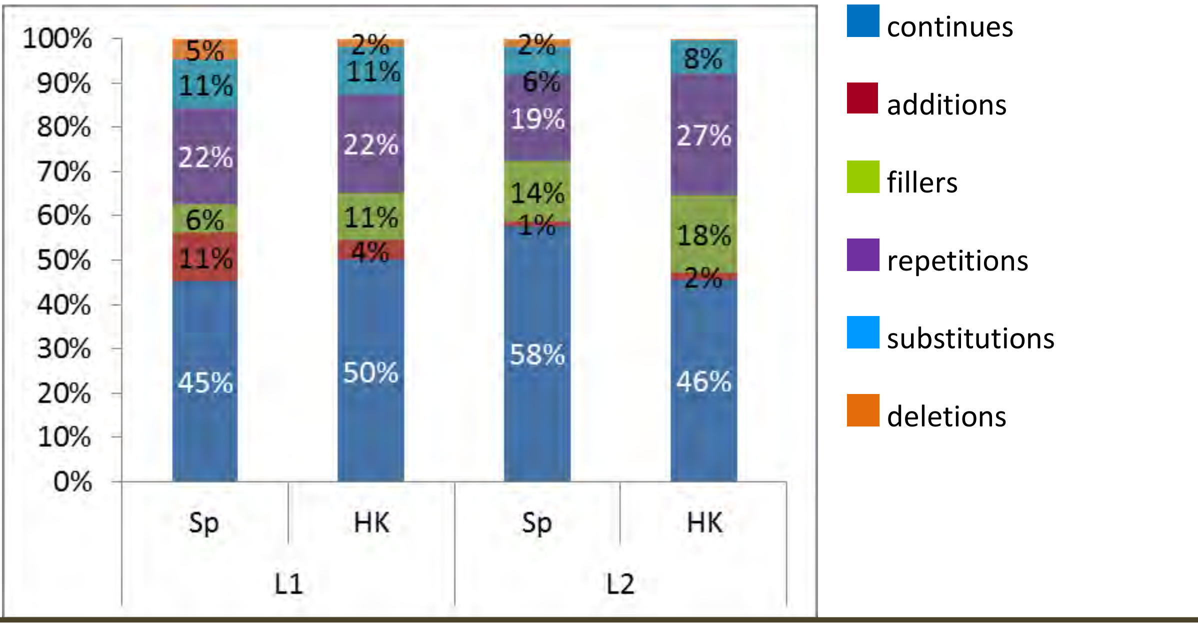
[70] These results do not contradict previous results on disfluency rates. They use actual number of disfluencies and do not account for number of words, phonic groups or lengths of the retellings.

The results shown in Figure 30 indicate that there are more disfluencies of all types produced in the L2 retellings (in particular by HK students), with cut-offs the least preferred. Figure 31 indicates that Spanish natives favour elongations over other disfluencies, while in the HK students' data it is possible to observe a tendency to use fillers rather than other types of disfluencies. There seems to be a difference in the use of disfluencies between the two groups, and also by language. These are explored in more detail below.

6.4.3 Repairs

The repairs in the video data were analysed, excluding repairs after pure pauses, to verify whether the classification of disfluencies had been correct (it was possible that some disfluencies might be speech traits, in particular elongations), see Figure 32.

Figure 32. – Comparison of video retelling data.



It can be observed that in almost half of all cases, in both groups, the utterances just continue after a disfluency. It is therefore possible that some disfluencies might have been wrongly classified as such.

6.4.4 Elongations

The calculations on elongations (excluding fillers that have been elongated) are summarised in below. Initially just the number of elongations was counted. The transcription of the elongations included information on whether they were followed by a pause (e) or not (e x) and the length of the pause. In the Spanish natives' data 158 elongations were recorded, slightly more in the L2 (54% of the total for all data from Spanish natives). A total of 187 elongations were observed in the HK students data, the majority in the L2 (72% of the total for all data from), Figure 33. However, when it became necessary to normalise the data, using the number of phonic groups by second (normalisation was calculated with various denominators and the results were similar), the original differences observed by group (Spanish natives vs. HK students) disappeared and became specific to HK students using the L2. It was also found that the data was very similar in terms of average (and total number of elongations) for Spanish natives' data L1 and L2, while a marked increase in the use of elongations was observable in Hong Kong students using the L1, Figure 34. The average length of the pauses (see Table 30) is similar for the L1 data for both groups but in the L2 it decreases for Spanish natives (sp) and increases in HK students.

Figure 33. – Number of elongations.

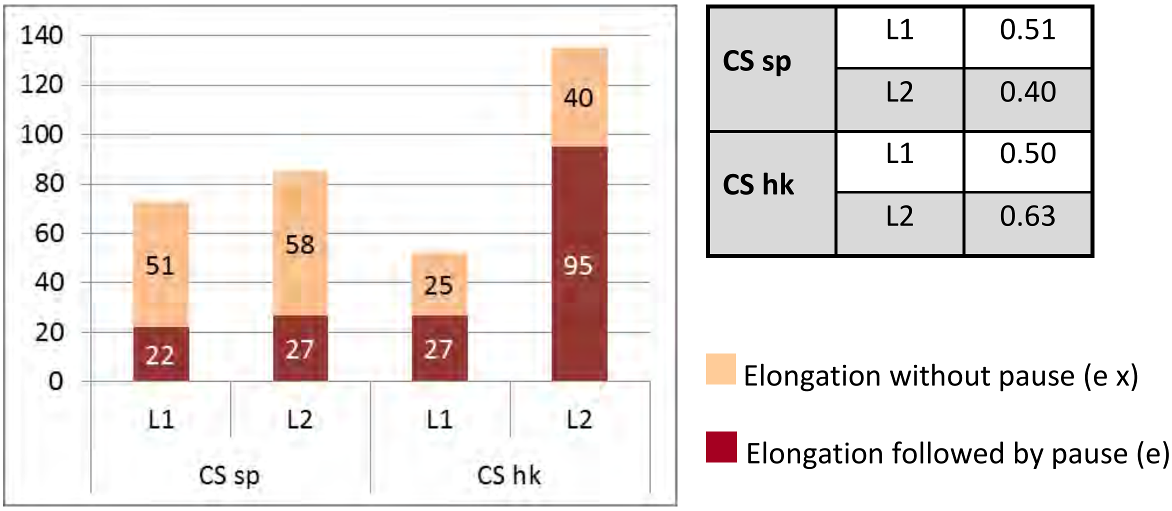


Table 30. – Average length of pauses after elongations (seconds).

CS sp	L1	0.51
	L2	0.40
CS hk	L1	0.50
	L2	0.63

Figure 34. – Average number of elongations by phonic group per second.

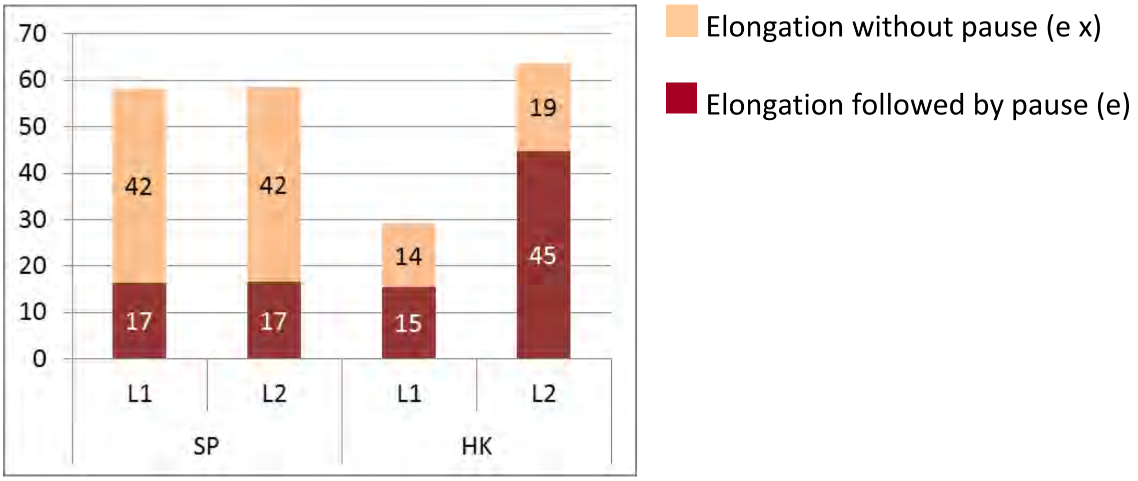
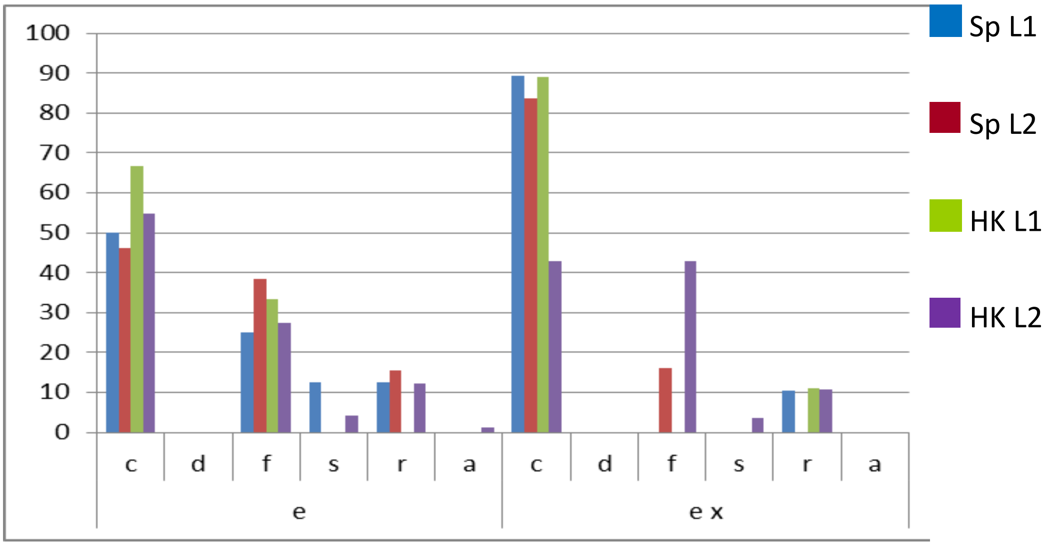


Table 31. – Percentage of types of elongations (followed or not by pauses) - as a percentage of the total rate of elongations observed (by PG/second).

As a percentage of the total rate of elongation observed (by PG/second)	Spanish natives L1	Spanish natives L2	HK L1	HK L2
Elongation followed by a pause (e)	28%	28%	53%	71%
Elongations without a pause (e x)	72%	72%	47%	29%

As can be seen from Table 31, HK students prefer to use elongations with a pause. In Spanish natives there is a preference to use the elongation without a pause — this could suggest that the elongation is not so much a disfluency but a speech trait. In order to test this hypothesis we looked at the repair that followed these elongations.

Figure 35. – Percentage of elongations by repair.



wdc = continuation d = deletion f = filler s = substitution r = repetition a = addition

As shown in Figure 35, the majority of elongations were followed by the continuation of the utterance without repair. Repair was not often used with elongations except if the elongation was followed by a pause, then it was also quite likely to be followed by a filler. In the L1 if there was no pause there was never a filler (only in the L2).

6.4.5 Cut-offs

The use of cut-offs as a proportion of the total number of disfluencies (in relation to other disfluencies), decreases in both groups when using the L2, from 31% in the L1 to 16% in the L2 in Spanish natives and from 35% to 29% in HK students, Figure 31. The data on cut-offs was analysed first as the actual number of occurrences (raw data) and then as a rate (cut-off per phonic group per second). The cut-off data was subdivided into cut-offs within word (wi) and cut-offs after word completion (i) and the pauses following both analysed. See Tables 32 and 33 and Figures 36 and 37 below.

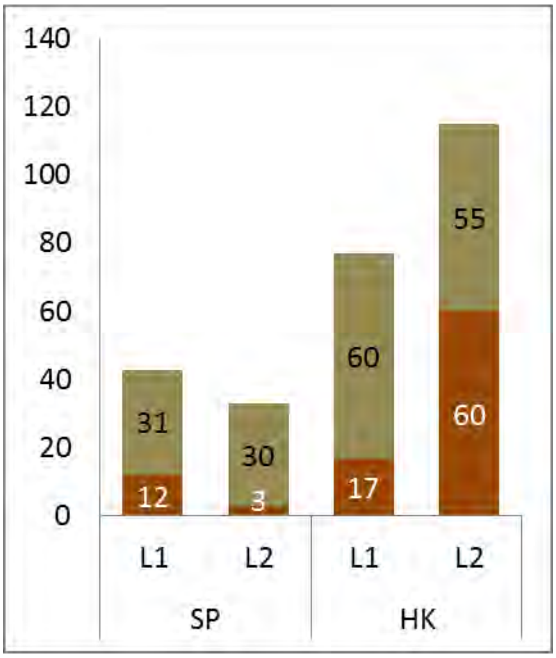
Table 32. – Summary of cut-off data for Spanish natives.

	Cut-off within word		Cut-off after word completion	
	Within word + pause (wi)	Within word no pause (wi x)	After word + pause (i)	After word no pause (I x)
L1	2	10	13	18
L2	0	3	10	20

Table 33. – Summary of cut-off data for HK students.

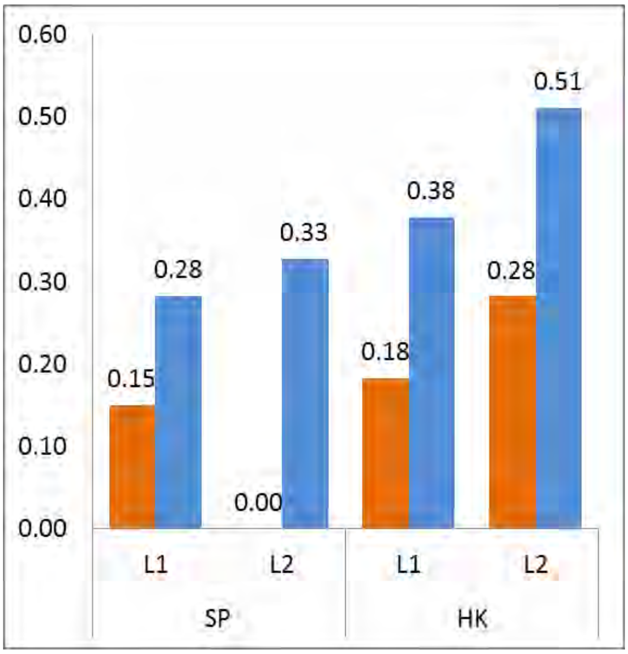
	Cut-off within word		Cut-off after word completion	
	Within word + pause (wi)	Within word no pause (wi x)	After word + pause (i)	After word no pause (I x)
L1	3	14	27	33
L2	27	33	36	19

Figure 36. – Summary of cut-off data in L1 and L2 video.



Cut- off within word (wi)
Cut-off after word completion (i)

Figure 37. – Pause length after cut-off in L1 and L2 video.



Pause length following cut-off after word completion
Pause length following cut-off within word

More cut-offs were observed in the HK students' data in the L2 than in the L1 (but less in Spanish natives' L2 than in the L1). In particular there is an increase of cut-off within words in HK students in the L2 and a decrease of these in Spanish natives. These differences are significant, as tested with a t-test between percentages of one sample, with a critical alpha level of 0.05 in all cases except for HK students in the L2.[71]

Figure 37 provided the average pauses after cut-offs within word (wi) or after word completion (i), excluding cases where there was also an elongation. Pauses are longer in the L2, as expected, but the difference is greater among HK participants. Moreover there is also a noticeable difference in length between pauses after cut-offs within word which were shorter (an average of 0.15 seconds for Spanish natives and 0.25 for HK students) than pauses after word completion (0.3 seconds for Spanish natives and 0.44 for HK students).

The differences in pause length are significant in the HK students data, both when comparing L1 and L2 lengths by type of cut-off and when comparing the pause length of after word completion cut-offs between the L1 and the L2 (to an alpha level of 0.05). The Spanish natives' data is not significant at this level. Table 34 gives the results of the independent group t-test between means (two-tailed probabilities).

Table 34. – Significance testing of differences in cut-off lengths.

	Pause length L1 vs. L2	**Pause length L1 vs. L2**	**Pause length L1 vs. L2**	**Pause length L1 vs. L2**	**Pause length - i vs. wi**		**Pause length - i vs. wi**	
	wi	**I**	**Wi**	**i**	**L1**	**L2**	**L1**	**L2**
t-test	__	(21) 0.717	(29) 1.355	(61) 2.57	(13) 0.984	__	(29) 6.573	(61) 4.144
Probability	__	0.481	0.189	0.0126*	0.343	__	0.0000*	0.0008*

*reliable at the 0.05 critical alpha level **reliable at the 0.1 critical alpha level

Note: degrees of freedom in brackets.

It is observed that on normalising the data there is a change in the relationships between L1 and L2 in Spanish natives' data. Overall the total number of cut-offs per average

[71] The results of the test were:

Spanish natives L1: $t(4) = 3.213$ $p = 0.0025$; Spanish natives L2 $t(32) = 8.23$, $p = 0.0000$; HK students L1 $t(76) = 5.931$, $p = 0.0000$; Hong Kong students L2: $t(114) = 0.429$, $p = 0.429$.

phonic group per second in Spanish natives is similar to HK students' data in the L1. The data was further subdivided into type of cut-offs followed by a pause or not see Figures 38 and 39 (where the same data is shown but grouped differently).

Figure 38. – Number of cut-offs per average phonic group per second in L1 and L2 video.

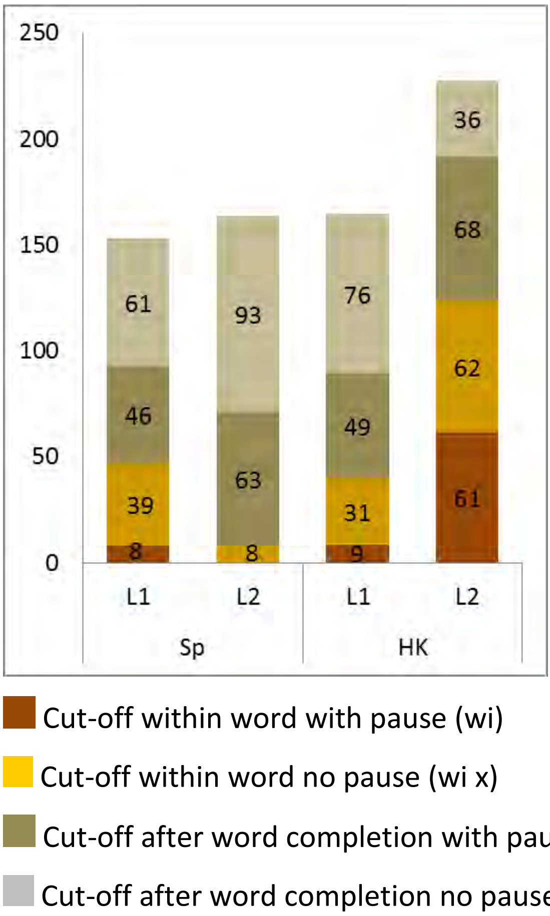
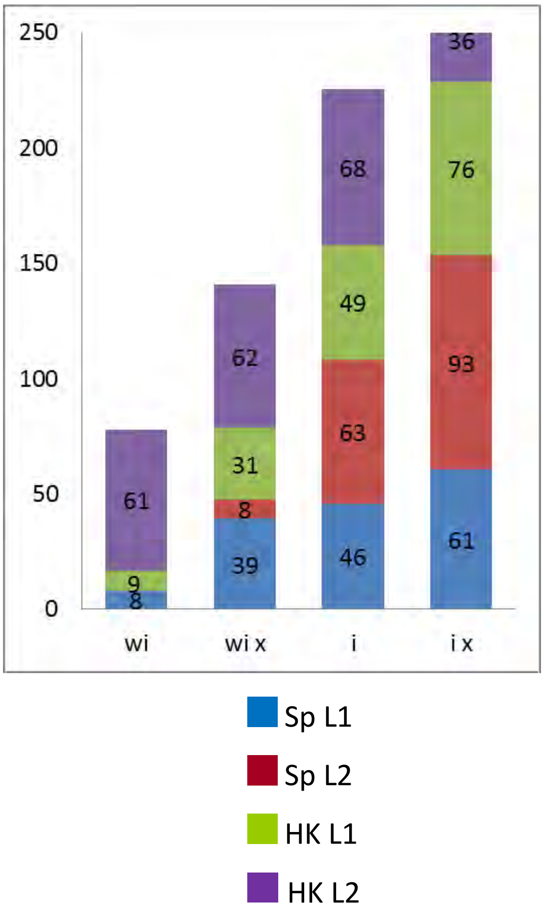


Figure 39. – Number of cut-offs per average phonic group per second by type of cut-off.



The normalised data shows a decrease in the number of within word cut-offs in the L2 amongst Spanish natives, from 47 in the L1 to 8 in the L2, while the number of after word interruptions increases from 107 in the L1 to 156 in the L2. With HK students on the other hand we observe a considerable increase in the number of within word cut-offs (40 to 123) and a decrease in cut-offs after word completion (125 to 104). The correlation of the results between the L1 and the L2 was tested (Pearson's correlation two-tailed probability test), the levels of correlation are high (cc=0.861 for Spanish students' data and cc=0.705 for HK students', indicating a somewhat strong correlation but the significance levels are not high enough to confirm the correlation (Spanish natives *p* = 0.14, HK students *p* = 0.295).[72] A correlation test was also carried out between the results of the two L1 groups (Spanish natives versus HK students) and the results of the L2. It was found that there was a strong correlation in the L1 data (0.96), which was significant (*p* = 0.042),[73] in the L2 data an even stronger correlation was found, with a coefficient of 0.977 which was significant (*p* = 0.02), in both cases the significance was calculated at an alpha level of 0.05 (95% probability of not being a chance occurrence). A third check for correlations was carried out, this time by language, data from English retellings of Spanish natives and HK students and data from Spanish retellings. For the data in Spanish we did not find a correlation (*cc* = -0.53) and the results were no significance either (*p* = 0.462), however with the data in English we found a strong correlation (*cc* = 0.96) which was significant (*p* = 0.0405), at an alpha level of 0.05 (95% probability of not being a chance occurrence).[74] To summarise: it seems that the distribution of cut-offs is used similarly in both English and Spanish by native speakers; speakers using an L2 also show similar cut-off distributions (regardless of the language); but HK students using the L2 are not comparable to native speakers of that language (although Spanish natives using the L2 are comparable to HK students using English).

To ensure that the transcription of the Spanish natives' cut-offs after word completion had been correct, and were indeed cut-offs, the repairs that followed them were analysed. Whether they were cut-offs or not could have been thought a subjective decision, as the indications of a cut-off, especially if there is no pause following it, are very subtle (such as a glottal stop). Table 35 gives the percentage of repairs by type that followed cut-offs after word completion. In Spanish natives only in 5% of the cases does the utterance continue (without any repair), suggesting that the majority of cut-offs transcribed were indeed disfluencies.

[72] Spanish natives: standard error = 0.129; *t*(2) = 2.398. HK students: standard error= 0.252; *t*(2)= 1.405.

[73] L1 Spanish natives versus L1 HK students: standard error=0.041, *t*(2) = 4.733; L2 Spanish natives versus L2 HK students: standard error = 0.023, *t*(2) = 6.477.

[74] Spanish data: standard error = 0.356, *t*(2) = -0.9; English data: standard error=0.04, *t*(2) = 0.4818.

Table 35. – Repair following after-word completion cut-offs in the L1 in Spanish natives.

Utterance repair	i	i:	i: x	i x	Total	%
Continues	1	1	0	0	2	5
Addition	0	1	2	2	5	13
Filler	4	0	0	1	5	13
Repetition	3	2	2	8	15	39
Substitution	4	1	1	3	9	24
Deletion	0	0	1	1	2	5

Note: The data used for this was that of disfluencies occurring with gestures in L1 retellings in Spanish natives. Cut-offs with elongations are marked ‘i:’ (followed by a pause) or ‘i: x’ (if not followed by a pause).

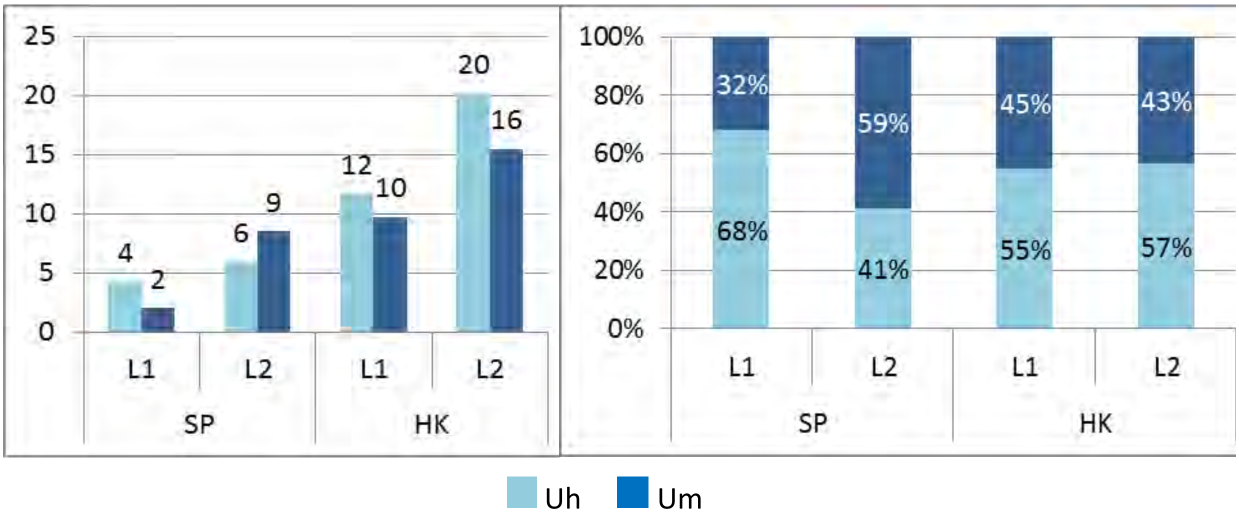
6.4.6 Fillers

The use of fillers is analysed below, only ‘uh’ and ‘um’ fillers (and their variations) were considered. The raw data indicates that the use of fillers increases in both cases from L1 to L2 and HK speakers use more fillers than Spanish natives. In all groups there were more ‘uh’ than ‘ums’ recorded, see Figure 40). When the total number of fillers is calculated per phonic group per second these results are confirmed. Comparing ‘uh’ and ‘um’ as the total number of fillers produced in each group, it is observed that Spanish natives have a preference for using ‘uh’ in the L1 (68%) but to a tendency to use ‘um’ in the L2 (59%), see Figure 41. Hong Kong students on the other hand use them in similar proportions in the L1 and L2. The differences in the percentage of ‘uh’ and ‘um’ by group and language were found to be significant to a critical alpha level of 0.1 in the L1 results of Spanish natives ($t(24) = 1.929$, $p = 0.06$) and the L2 of HK students ($t(142) = 1.69$, $p = 0.093$).[75] Furthermore, when calculating the significance of the differences in filler percentages by language (using a two sample percentage t-test), it was found that the differences in Spanish natives were significant to a critical alpha level of 0.05 ($t(81) = 2.257$ and $p = 0.027$). HK students differences by language were not significant ($t(227) = 0.295$, $p = 0.77$). Further analysis of fillers was carried out to investigate these results, by taking gestures into account.

[75] These were calculated using a one sample percentage t-test. Other results for Spanish natives in the L2 were: $t(57) = 1.39$, $p = 0.169$; and for HK students: $t(85) = 0.93$, $p = 0.354$.

Figure 40. – Average number of fillers by type in the L1 and L2.

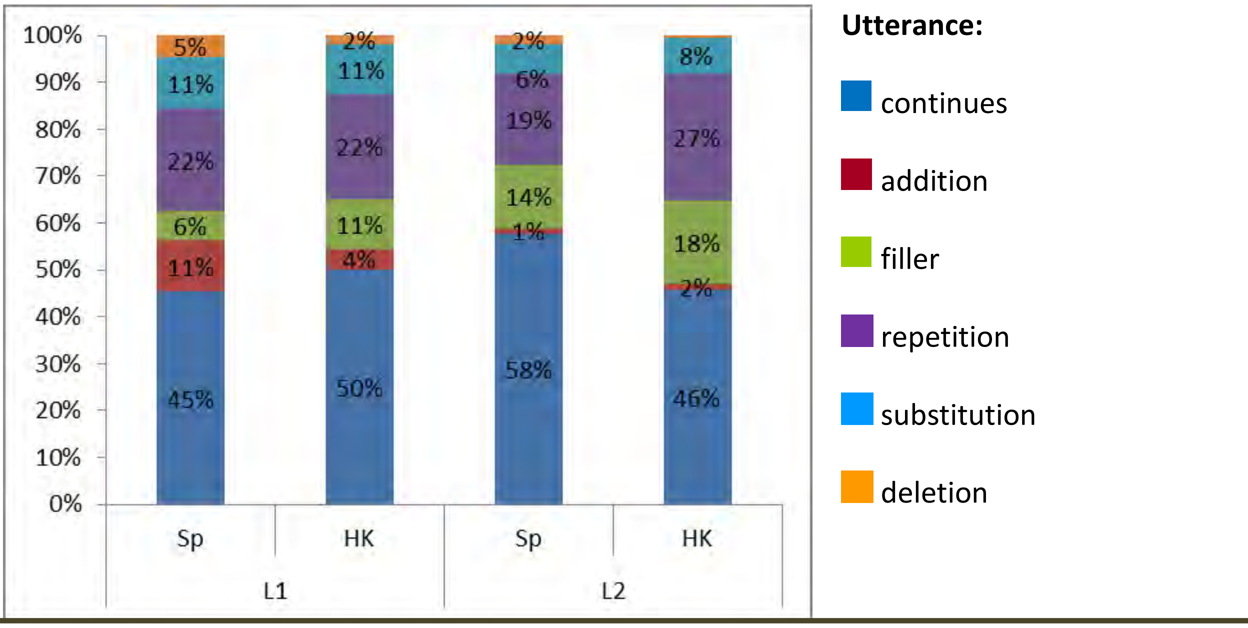
Figure 41. – Type of fillers in L1 and L2 (percentages).



6.4.7 Repairs

The nature of the repairs that followed fillers ‘uh’ and ‘um’ is shown in Figure 42. As can be observed, in all groups close to half of the fillers are followed by continuing utterance, not a repair.

Figure 42. – Type of repair in fillers with gestures (percentages) in L1 and L2 video.



6.5 Video Corpus - Gestures

Following the decision to look only at the data from the video exercises, gestures produced by both groups of participants in the L1 and the L2 video retellings were compared in more detail. There were more gestures analysed among the HK students, in both the L1 and the L2, as their retellings were longer, see Figure 43.

Figure 43. – Total number of gestures in the L1 and L2 videos.

Figure 44. – Total number of gestures per phonic group per second in the L1 and L2 videos.

L1 video L2 video

Gesture rates were recalculated per phonic group, per length and per number of words. In the first two cases we observed that gesture rates in Spanish natives were higher than in HK students, but the calculations of gesture per word shows higher rates of gestures in HK students in the L2 than in Spanish natives in the L2, see Table 36.

Table 36. – Average no. of gestures.

	Per average number of PGs		Per average length		Per average number of words	
	Spanish natives	**HK students**	**Spanish natives**	**HK students**	**Spanish natives**	**HK students**
L1 video	1.49	0.92	0.46	0.40	0.17	0.17
L2 video	1.35	0.91	0.53	0.48	0.27	0.41

When the average number of gestures was divided by the number of phonic groups per second (Figure 44), it was observed that Spanish natives in the L1 gestured more than HK students. It has to be noted that HK students produce more phonic groups per second (0.44 in the L1 video and 0.5 in the L2 video) than Spanish natives (0.31 in the L1 video and 0.4 in the L2 video).

Table 37. – Average no. of gestures per PG or word per second.

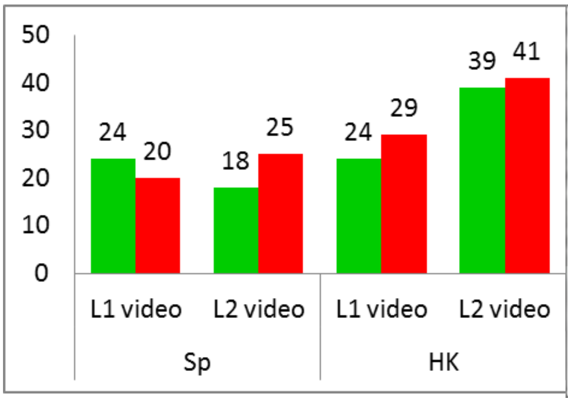
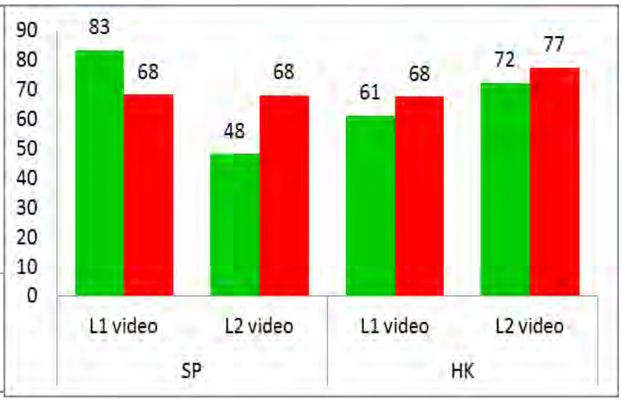
	Per average number of PG/second		Per average number of words/second	
	Spanish natives	HK students	Spanish natives	HK students
L1 video	151	128	15	22
L2 video	116	149	19	70

There are more pauses in the L2 of HK students, so phonic groups are shorter in the L2, therefore there will be less time in which to gesture. However HK student rates of gesture increases, indicating also a gesture rate increase (gestures are faster). In Spanish natives the gesture rate decreases per phonic group per second in the L2 but increases per word per second. Obviously there is a difference in the use of gestures in the L1 vs. the L2 in the two groups.

6.5.1 Types of gestures in the video corpus

One of the possible reasons for there being more gestures in the L2 in HK students could be an increase of pragmatic gestures to help in lexical retrieval. To test this the gestures were divided according to their relationship with the speech content, if this was of an iconic nature gestures were labelled iconic (i), otherwise they were labelled pragmatic (p). The results indicate that out of the eight retellings (four participants each doing two retellings, one in the L1 and one in the L2) among the Spanish native speakers there is a prevalence of iconic gestures (in six retellings) and when comparing L1 retellings against L2 retellings a prevalence of more iconic gestures in the L2 is observable in three cases (out of four, except in C.S. 2sp); the percentage of pragmatic gestures goes down in two cases and up in another two.

Among the eight retellings done by HK students the prevalence of one or another type of gestures is equally divided (in four retellings there are more pragmatic gestures, in the other four more iconic). When comparing L1 retellings against L2 retellings in most cases both the number of pragmatic and iconic gestures increases in the L2 (except for pragmatic gestures in C.S. 2hk). Figure 46 presents the data as the average number of gestures for each group, and Figure 47 as a percentage of the total of gestures per group.

Figure 45. – Average number of pragmatic and iconic gestures in the L1 and L2 videos.	Figure 46. – Average number of gestures per phonic group per second by type in the L1 and L2 videos.
	

Iconic (i) Pragmatic (p)

Calculations were also carried out to convert the average number of gestures to the average number of gestures per phonic group per gesture, the same patterns were observed in the rates, see Figures 46 and 47. Thus it was considered that for pattern comparison purposes the average number gave an accurate picture.

Table 38. – Average number of gestures by type per phonic group per second in fillers and within word cut-offs.

Average number of gestures per phonic group per second	All gestures				Gestures with disfluencies			
	Spanish natives		HK students		Spanish natives		HK students	
	L1 video	L2 video	L1 video	L2 video	L1 video	L2 video	L1 video	L2 video
Pragmatic (P)	83	48	83	48	34	34	28	65
Iconic (I)	68	68	68	68	21	42	33	57

6.5.2 Disfluencies with gestures in the video corpus

In the video corpus most participants' disfluencies fell within a gestural phrase (in 10 out of the sixteen retellings). Figures 48 and 49 breakdown the data by participant.

Figure 47. – Number (bar) and percentage (axis) of disfluency (D) with gesture (D wi G) and without (D wo G) in Spanish natives.

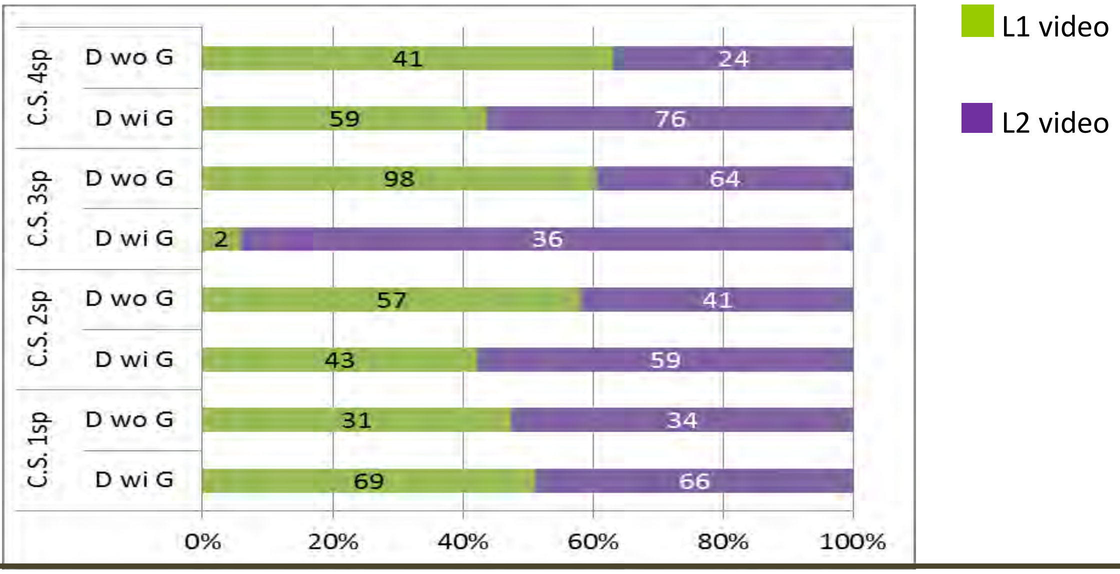
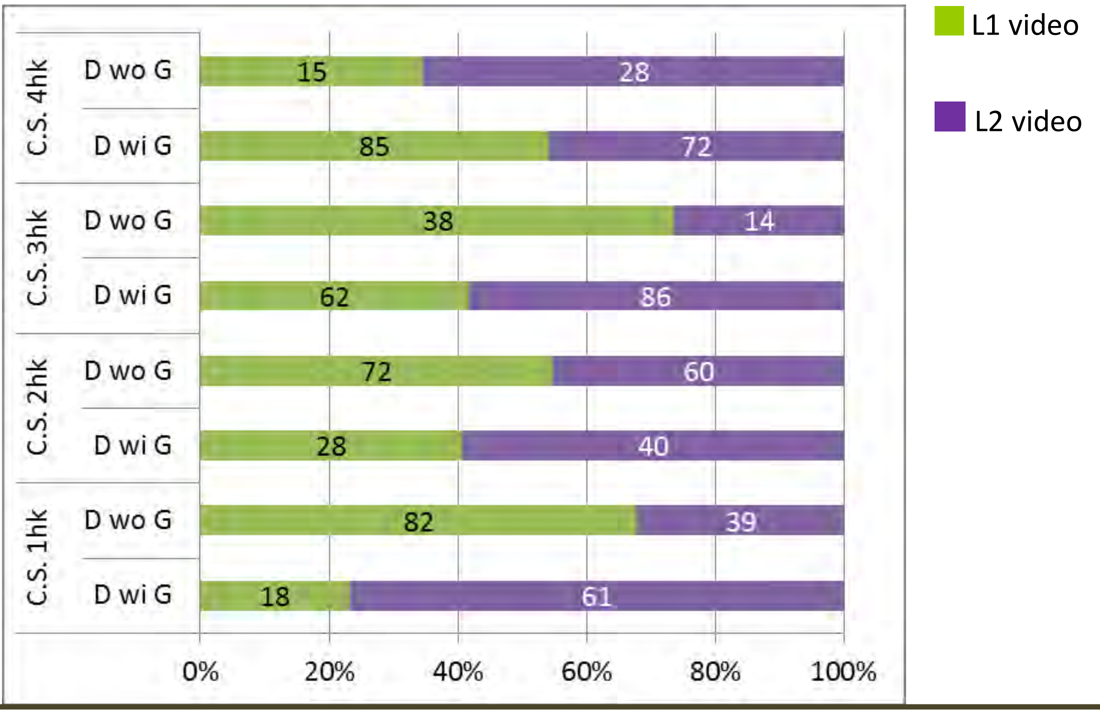


Figure 48. – Number (bar) and percentage (axis) of disfluency (D) with gesture (D wi G) and without (D wo G) in HK students.

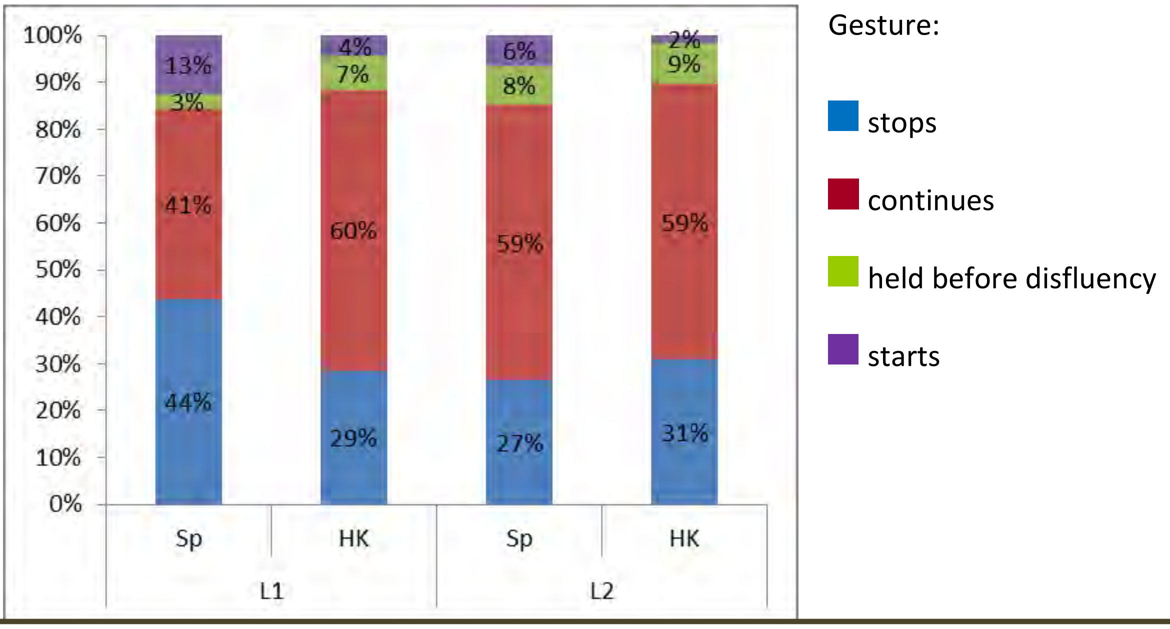


There is a tendency to use disfluency with gesture in the L2 and disfluencies without gestures in the L1 in most participants.

6.5.3 Gesture movement at point of disfluency

The analysis detailed whether the gesture continued at the point of the disfluency or whether it was put on hold or started (it also recorded if the gesture had been put on hold previous to the disfluency, so the gesture was in mid-phase and remained on hold during the disfluency). Figure 45 gives the results. As it can be observed, in a very high percentage of disfluencies, the gesture just continues (in close to 60% in all groups except Spanish natives using the L1), corresponding to similar percentages of the utterance just continuing after the disfluency.

Figure 49. – Gesture action at point of disfluency in L1 and L2 videos.



6.5.4 Type of gesture during the disfluency

Figure 50. – Average number of type of gestures in disfluencies in videos.

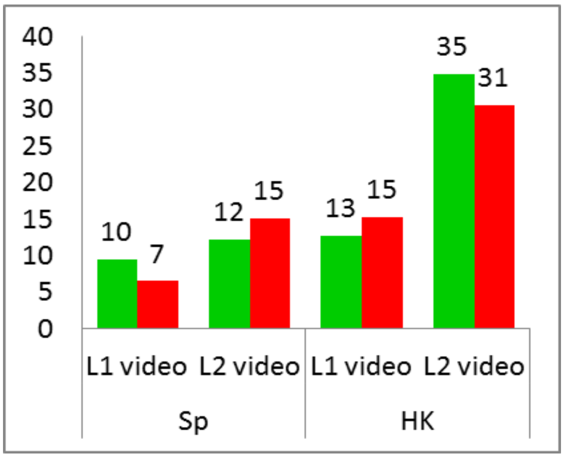
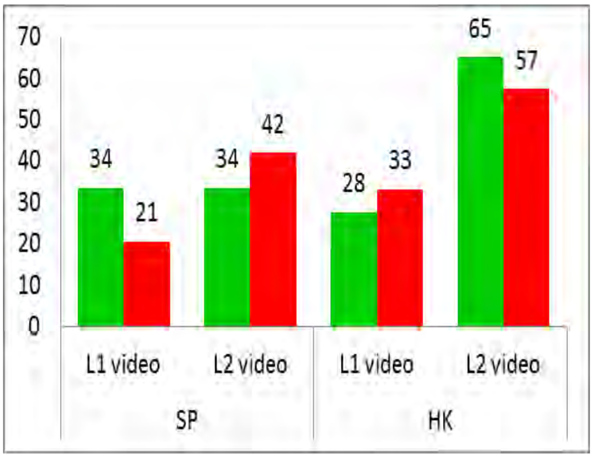


Figure 51. – Average number of gestures per phonic group per second in disfluencies.



Comparing the relationship of iconic versus pragmatic gestures in gestures with a disfluency (Figures 50 and 51) to all the gestures (Figures 46 and 47) a very similar pattern can be observed, except in the gestures in the L2 of HK students. Overall in both the L1 and the L2 gesture with disfluency data there seems to be a prevalence of pragmatic gestures (all averages add up to 70 pragmatic gestures and 66 iconic), while in the all gesture data there was a higher number of iconic gestures (all averages add up to 115 iconic gestures and 105 pragmatic). Although on an individual basis this is observed only in three retellings (with more iconic gestures used in the other five retellings), in both groups, suggesting one individual might be skewing the data. In all cases but one (C.S. 4sp, pragmatic gestures) there are more gestures in the L2 than in the L1, of both types.

6.5.5 Gestures with overt disfluencies

One possibility that was considered was that certain disfluencies might have been speech traits and so distorts the pragmatic vs. iconic relationship. In order to address this, disfluencies with gesture were reanalysed using only those which were overt, either fillers ‘uh’ or ‘um’ or within word cut-offs.

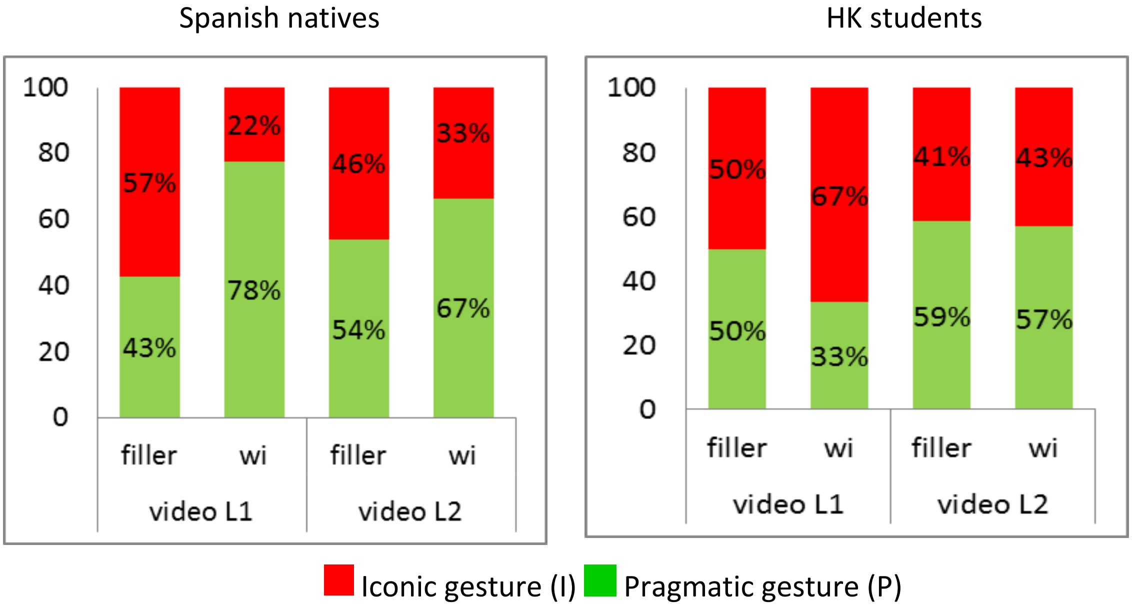
Both fillers and cut-offs in the L2 in most cases showed a greater percentage of pragmatic gestures. A one sample t-test between percentages was carried out to see whether the differences were reliable (if they could have happened by chance). Only for Spanish natives L1 within word cut-off (wi) is the difference between the two percentages significant at the 0.1 critical alpha level, but as the sample is small (9 data points) and the results are not significant in any other group, we would advise caution when using these results.

Table 39. – Number of gestures by type (p, i) in fillers and within word cut-offs.

	Spanish natives				**HK students**			
	Video L1		**Video L2**		**Video L1**		**Video L2**	
	filler	**wi**	**filler**	**wi**	**filler**	**wi**	**filler**	**wi**
P	3 (43%)	7 (78%)	20 (54%)	2 (67%)	24 (51%)	4 (33%)	48 (59%)	24 (57%)
I	4 (57%)	2 (22%)	17 (46%)	1 (33%)	23 (49%)	8 (67%)	34 (41%)	18 (43%)
t	(6)0.374	(8)2.028	(36)0.488	(2)0.626	(46)0.137	(11)1.252	(81)1.657	(4)0.9
p	0.721	0.0771**	0.628	0.595	0.891	0.236	0.101	0.36

**reliable at the 0.1 critical alpha level

Figure 52. – Average number of type of gestures in disfluencies in videos.



Although the results are not significant, there seems to be a prevalence of pragmatic gestures in cut-offs within word in Spanish natives, but a preference for iconic gestures in HK students L1 data(in the L2 the difference between the usage of different type of gestures is very small), for fillers the difference between the two percentages is relatively small, suggesting that there isn't a clear difference in the use of pragmatic vs. iconic gestures.

One of the questions raised earlier was whether a difference between the filler 'uh' and 'um' could be seen in gesture when the disfluency is uttered. As previously, only data from overt disfluencies was included, Figure 53 and Table 40 give the results for within word cut offs and fillers. The gesture was observed when the disfluency started and it was recorded whether it continued (c) throughout the disfluency or stopped with it (s). In some cases the gesture started with the disfluency in which case it was transcribed as such but for the purposes of the analysis it was included within the gesture-continuing label.

Figure 53. – Percentage (axis) and numbers (on bars) of gesture at within word cut-offs.

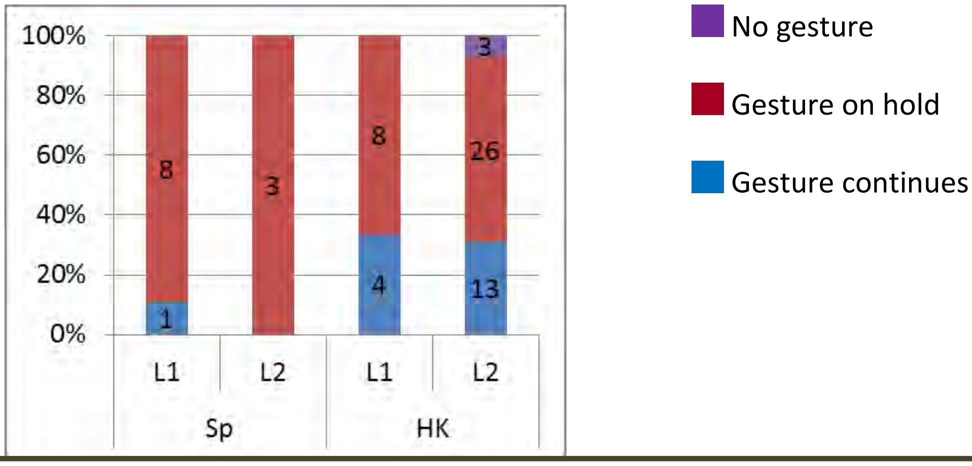


Table 40. – Gesture move by type of filler.

Gesture continues, stops, or no gesture						
%		Spanish natives		HK students		
		L1	L2	L1	L2	Average
uh	Continues	14	32	45	48	35
	Stops	7	16	13	10	11
	No gesture	0	14	4	15	8
um	Continues	14	32	23	18	22
	Stops	14	5	13	5	9
	No gesture	50	0	2	5	14

Note: columns add up to 100% (differences are due to rounding up of figures).

There seems to be a prevalence of stopping the gesture with the within-word cut off, while with the fillers there seems to be a prevalence of continuing the gesture. We tested whether there was a significance to the percentages obtained in continuing and stopping gestures in the 'uh' and the 'um'. A one sample t-test between percentages with a two-tailed probability was used.

The results indicate that within the 'uh's the percentage difference obtained was reliable at the 0.051 critical alpha level (95% chances of not occurring by chance) in HK students. And the difference in 'um' in eth L2 of Spanish natives was reliable at the 0.1 critical alpha level (90% chances of not being an accident). None of the other results were shown to be significant, see Table 41.

Table 41. – One-tailed Pearson's correlations for continuous vs. stopping gestures.

	Spanish natives				**Hong Kong students**			
	uh		**um**		**uh**		**um**	
	L1	**L2**	**L1**	**L2**	**L1**	**L2**	**L1**	**L2**
t	(2) 0.27	(17) 1.01	(3) 0	(13) 1.85	(26) 2.4	(46) 3.9	(16) 0.69	(18) 1.22
p	0.814	0.33	1	0.09**	0.02*	0.003*	0.49	0.23

*reliable at the 0.05 critical alpha level **reliable at the 0.1 critical alpha level

Note: degrees of freedom in brackets.

It seems that the differences observed in gestures during 'uh' in HK students' data are significant but we would like to caution that the sample is small (as can be seen by the degrees of freedom). To create a larger sample the data for all fillers with gestures was totalled, this time to look at the type of gesture with the disfluency and whether it stopped or continued. Table 42 shows the compiled results for all fillers (regardless of group or language).

Table 42. – Type of gesture with disfluency and its action (percentages).

Uh (100%)				**Um (100%)**			
g continues		**g stops**		**g continues**		**g stops**	
i	**p**	**I**	**p**	**i**	**p**	**i**	**p**
35%	43%	11%	10%	28%	47%	15%	9%

Note: g = gesture

There seems to be a prevalence of continuing pragmatic gestures in both 'uh' and 'um', while there is a preference for stopping 'um' iconic gestures. This preference is seen in both groups and languages. A one sample, one-tailed, t-test between percentages gives the pragmatic and iconic difference in the 'um' continuing significance. ($t(65) = 1.827$ and $p = 0.072$, which is significant at 0.1 critical alpha level, 90% chance of seeing the difference between the two percentages repeated in further samples). The iconic and pragmatic percentage of the gesture stopping do not prove to be significant ($t(65) = 1.003$ and $p = 0.32$).

6.6 Summary

Figures 54 and 55 provide a summary of some of the key data in terms of numbers of disfluencies produced with gestures and the repairs that followed them. In addition the gesture is categorized by type (pragmatic, p, or iconic, i). The data included all retellings in the L1, to ensure a sample as large as possible; when only the video data is included we observe a similar pattern).

Figure 54. – Number of repairs by type in disfluencies with gestures (by type) in L1 and L2 video, Spanish natives.

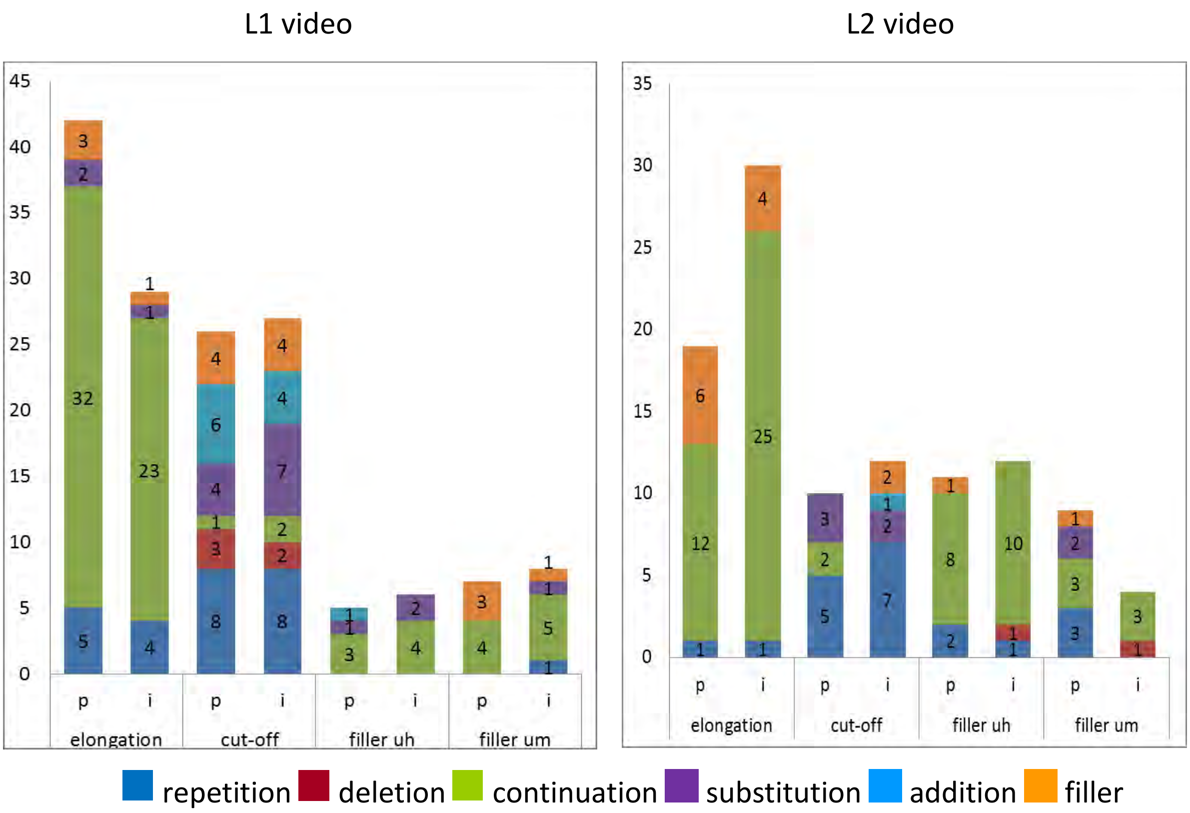
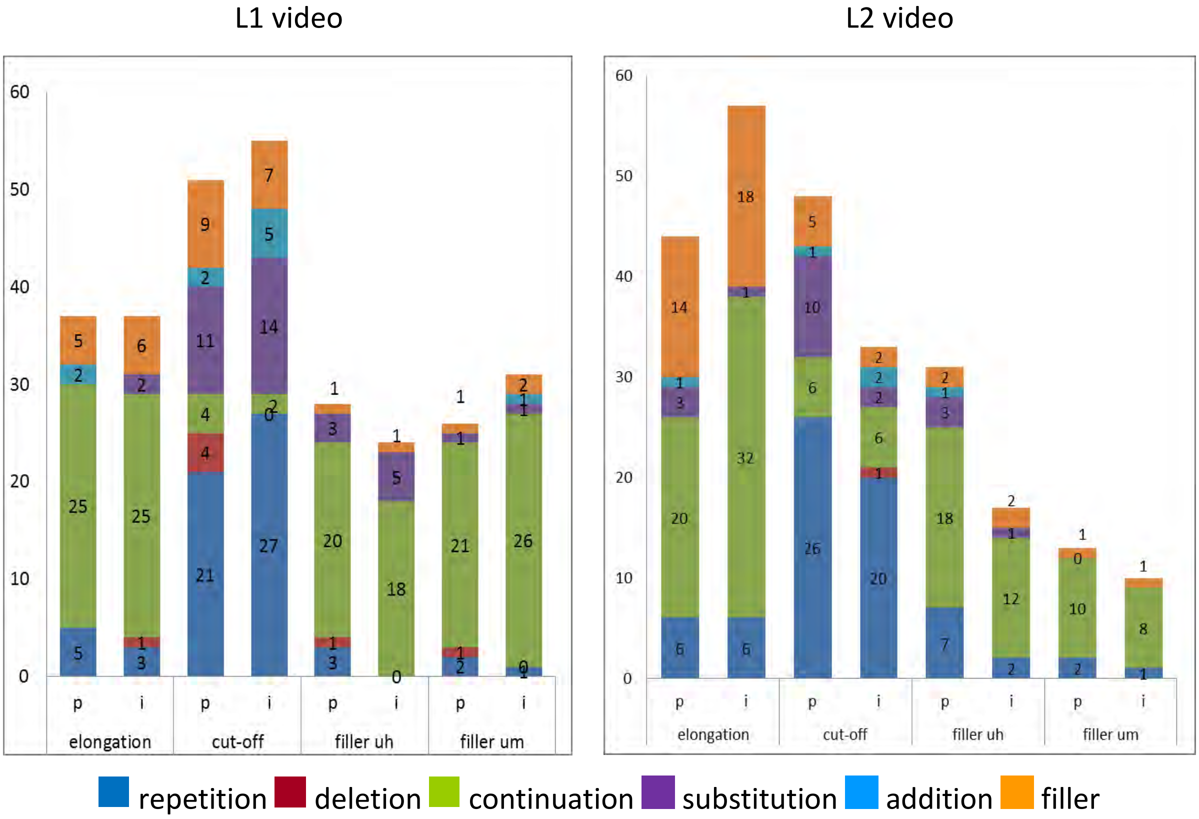


Figure 55. – Number of repairs by type in disfluencies with gestures (by type) in L1 and L2 video, HK students.



6.6.1 Observations

In terms of total numbers there are more instances of disfluencies with gestures in the data of HK students, but their retellings were often longer. What is interesting to note is that in HK students the number of disfluencies in the L2 is not dissimilar to the total from all of the L1 retellings, suggesting a significant increase of disfluencies in the L2. Spanish natives' data shows a prevalence of elongations, in both the L1 and the L2. While HK students' data shows a preference for cut-offs in the L1 and elongations in the L2.

In terms of the type of gesture we observe that in general in Spanish natives and in the L1 of HK students the distribution of pragmatic versus iconic gestures in each disfluency is quite similar, except in the elongations of Spanish students. In HK students in the L2 we observe a tendency to use pragmatic gestures in all disfluencies except elongations.

Analysing the data by repair it seems that elongation is usually followed by continuing utterances, questioning its disfluency categorization. In both groups in both the L1 and the L2, cut-offs are most often followed by a repair: repetition, substitution or addition. Fillers on the other hand seem to be differently used by each group in the L1 and the L2, although the tendency is to continue the utterance.

7 DISCUSSION

Based on the studies mentioned in the *Literature Review* (above) it is clear that a number of questions concerning disfluency and the gesture-speech link are still to be answered. Just as speech processing mechanisms were based on the study of slips of the tongue and other errors, we believe that more could be learned about the gesture-speech link from disfluency gesture-speech studies. One of the biggest hurdles is that disfluencies are different from individual to individual and so are gestures, and both are also dependent on content and context. Consequently studying data points from mixed speakers, as was carried out in the preliminary study, might not reveal much. The content and context of the utterance in which the disfluency was recorded might have varied and thus would have an impact on the performance of the individual. In order to identify significant patterns applicable to more than just one individual, it is necessary to look at comparable data from more than one source. Thus we have considered it essential to look first at the data from an individual and content specific (by type of input) point of view before generalising our results to the group.

Some studies control variables to obtain specific data from participants. In most cases this creates a somewhat constrained and forced event (such as attaching electrodes to speakers' hands, to prevent gestures). Nonetheless there are some variables which might not hinder spontaneity or reduce naturalness, such as comparing a dialogue versus a monologue (where the production of turn managing gestures would vary), or asking participants to use two languages (creating a different cognitive situation). In the preliminary study we failed to control a number of variables which impacted the results, such as having a dyad set up. A dyad means a dialogue, where turn management gestures (and possibly some disfluencies) are part of the communicative process. Participants were also allowed to hold pieces of paper which greatly hindered the use of the hands. Thus, although we were able to confirm that HK students of Spanish as a L2 performed disfluencies and gestures, we were not able to see any significant patterns in the data collected.

The main study controlled as many variables as possible and analysed the data accordingly. In the main study, participants (Spanish natives and Hong Kong (HK) students) were asked to recount three stories in their first language (L1).[76] The input for the first story was a written text, the Fable of The *Lion and the mouse*; the second an audio (with a made up *Tweety and Sylvester* story); and the third, the first three minutes of a video cartoon with *Tweety and Sylvester*, *Canary Row*, with no sound. Then participants watched the last three minutes of the video and retold it in the second language (L2). Finally they were asked, in their L1, a number of cognitive questions relating to the exercise, a debriefing in the form of an interview. Therefore we were able to analyse the data by language and by type of input.

This chapter begins with observations on various variables associated with the research, gender and rates used to normalise the data. It continues with a detailed discussion of the various findings, based on our results on the use of disfluencies and gestures and on the information given in the *Literature Review*. The final recommendations for further research and the conclusions are given in the next chapters.

7.1 Variables

This section provides a brief explanation on variables, such as gender and rates used to normalise the data, which ought to be considered when interpreting the data.

7.1.1 Gender

Even though the focus of this study was not to find gender differences we are aware that there might be differences by gender and so the participants were specifically chosen to ensure gender equality, in terms of the data. In certain disfluency studies (Tottie, 2011; Bortfeld et al., 2001) gender differences have been found in relation to fillers. These are associated with the male tendency to lead the conversation. After reviewing our results we could find no recurring differences between the sexes, confirming the observations of scholars, such as Kendall (2013, p. 32), who point out that studies on gender-based speech rates present conflicting results. Speech rate perception is highly subjective depending on the listener, and the social and communicative expectations (p. 17). It is not at all clear that differences in speech can be explained by gender differences; Weatherall (2002) presents a comprehensive review of gender studies from a discourse and linguistic point of view and

[76] Spanish natives spoke English as the L2. HK students spoke English as the L1 and Spanish as the L2.

concludes that speech differences have “social meaning other than gender” (p. 95). Within language acquisition, differences in acquisition and processing between the sexes have been noted but there is no conclusive evidence to confirm any patterns. Saville-Troike states that “There is a wide spread belief in many Western cultures that females tend to be better L2 learners than males, but this belief is probably primarily a social construct, based on outcomes which reflect cultural and socio-psychological constraints and influences.” (2006, p. 90).

A review of the literature suggests that during the mid-twentieth century much research was done on gender differences (not just in linguistics) but the results were often contradictory or inconclusive at best. As awareness of feminist, lesbian and gay perspectives grew, these differences became almost a taboo subject, to the extent that many authors shy away from them. Even the authors mentioned above advise caution on stating gender differences. Weatherall (2002) points out that “asking whether sex differences in verbal abilities exist, or whether there are sex differences in the organization of the brain for language, might not be very good questions because they are too general and too simplistic” (p. 48). Kendall (2013), in his study of speech rates finds gender related differences and states that “this possibly indicates that males and females distribute their speaking time differently […] but beyond making this suggestion I resist further speculation” (p. 119). Bortfeld et al (2001, p. 140) offer their results on gender differences “with caution and because of their descriptive interest”. Taking their cue we will shy away from analysing the data from a gender point of view and speculating on the results.

7.1.2 Rates

In order to compare the data from different participants it was necessary to neutralise as many variables as possible. Variables associated with age, content, conversation management and socio-economic factors were controlled (by ensuring they were similar in all participants). But as lengths of retellings were different these had to be normalised. Speech in the L1 and L2 was being compared, with the expectation that speed of speech, number of words, disfluencies and gestures would vary between the two. Therefore these also had to be normalised. As can be seen from Tables 6 and 7, lengths of retellings differed greatly, even by participant. In HK students’ L2 data the lengths had been manipulated (to ensure manageable transcripts) and as number of words and of phonic groups depended on the length of the retelling it was necessary to use it as the denominator to calculate rates per second.

As the results show, the denominator used to calculate disfluency and gesture rates, whether it is number of words, length or phonic groups, can significantly alter the findings. This study has not included a rate based on clauses but we predict that similar results would be observed. These findings support Kendall (2013) who, on calculating speech rate based on various denominators (number of words, syllables), also found differing patterns in comparing Finnish and English speech rates (p. 17). Tables 12 to 15 provide information on speech rates, phonic group per second and word per second, calculated for the L1 and L2 data. Data from L1 retellings by input was compared to investigate whether the differences observed were reliable and if there was a correlation between the different types of retellings in the L1 by participant (see Tables 16 a & b). Depending on the rate used (number of phonic groups per second or number of words per second) some strong correlations were found, in particular between written and aural retellings in Spanish natives – this was significant using both rates. The same type of retellings did not show a strong correlation (although the results were not very reliable) with HK students' data. These results lead us to suspect that there are significant differences in the retellings based on the type of input given to participants and that comparisons between data from different studies ought to bear in mind the type of input used to generate the corpus. This would be in line with the results of Bortfeld et al. (2001) who found different utterance lengths and numbers of disfluencies depending on the topic speakers were discussing (photographs of their offspring or tangrams).

As the objective of this study was to compare L1 and L2 production, to eliminate possible differences related to the type of input, only data from the video retellings, in the L1 and L2, was used for the detailed analysis. When comparing the phonic group per second and word per second speech rates for the L1 and L2, it was found that the HK students' data showed a greater difference than that of Spanish natives between L1 and L2 speech, measured both as word[77] per second and phonic group per second (see Table 27). However, only phonic group per second rates between the L1 and the L2 showed a significant correlation (tested using Pearson's coefficient) at a 0.1 critical alpha level (see Table 28). As Spanish natives were quite proficient in the L2, their speech pattern was similar in both the L1 and the L2. While in HK students' speech in the L1 was more fluent, faster and with longer utterances, than in the L2.

Rühlemann, Bagoutdinov, and O'Donnell (2011) considered that pauses in narration might be indicators of thought units; because phonic groups are determined by pauses we understand them to relate to thought units. They also believe there might be a difference in the use of disfluencies between narrations and dialogues and have demonstrated that pauses differ whether the speaker is narrating a story or conversing. The average pause length is shorter in narrations. Altogether narration shows more short and filled pauses than conversation. The authors suggest that pauses are used differently in narration for a number of reasons: first it is necessary to switch registers from a conversation to a narration, and this entails changes in turn-taking; 'and' is used more often to facilitate coordination of clauses, which is seen as the "essence of storytelling" (p. 221). So pauses are considered as "indexing the cognitive efforts" that generate the thought units (p. 226). If by changing the type of input the cognitive process is likely to be affected, our belief is that the results obtained from different types of retellings cannot all be analysed as equal examples of L1. It is likely that type of input introduces an extra variable that results in varying frequencies of types of disfluencies. Therefore we consider this as further support to the decision to use only L1 and L2 data from the video retellings to carry out the in-depth analysis.

Comparing pauses after disfluencies in the L1 and L2 video data, it was found that Spanish natives (L1 and L2) and HK students in the L2 prefer to use fillers followed by a pause (while elongations and cut-offs are more often used without pauses). Spanish natives decrease the use of pauses after fillers in the L2, while there are no changes for elongations or cut-offs. These results are significant at a 0.1 critical alpha level in most cases. In HK

[77] Guidelines for word count also vary from study to study. This study included as words cut-offs as well as fillers, following the practice of Bortfeld et al. (2001).

students' there are more pauses following elongations and cut-offs than fillers and their proportion increases in the L2 data, however there is a decrease in fillers followed by pauses in the L2. The percentage differences in the L2 are significant to a 0.05 critical alpha level for elongations and fillers (see Tables 25 and 26). As expected, we observed more pauses following disfluencies in the L2 in both groups. However, what was not expected was a decrease in the proportion of fillers followed by a pause in the L2, compared to those not followed by a pause. These results suggest that fillers are perhaps being used differently than elongations and cut-offs and this is more obvious in the L2. This point will be further expanded below (under *Fillers,* Section 7.3.2.3).

If phonic groups indicate thought units, we would have expected to observe a stronger correlation between the two languages, especially in Spanish natives. The rate of phonic group per second differs for the two groups, Spanish natives' rate being lower. One explanation could be that the differences in Thinking for Speaking (Slobin, 1996) between English and Spanish affect the length of thought units. But we believe it more likely that the differences are due in part to the use of elongations in Spanish – Spanish links syllables, even from different words, often reducing adjacent vowels and pronouncing them as one syllable (Dauer, 1983, p. 57). HK students were less proficient in the L2 and so it is possible to observe the traits of the L1 in their L2 (they are still making use of the interlanguage). We believe that the differences in phonic group might be related to the interference of Cantonese on HK students. Cantonese is a syllable-based language (Setter, 2012), like Spanish, but with marked pauses between syllables (or at least that is the perception), so there are likely to be more phonic groups in Cantonese speech. Although our HK students were fluent in English, theirs is a Hong Kong English variety. A key difference with other English varieties, affecting this study, is the lack of reduced vowels that result in the perception of a syllable-based speech (Deterding, Wong & Kirkpatrick, 2008; Setter, 2012), rather than stress-based as is the British English variety. Thus we did not include non-reduced vowels in the list of disfluencies analysed.

A further significance paired t-test was carried out comparing L1 and L2 speech rates for each group of participants (see Table 29). A smaller difference – although not significant – was observed again between the L1 and L2 in Spanish natives when looking at the phonic group per second, confirming the results above (similar use of pauses and breath patterns in both languages). A significant difference is observed between the L1 and the L2 among HK students when measuring the number of words per second. It is our belief that this higher difference in HK students is not so much related to the type of language but to the proficiency levels (HK students being less proficient in the L2 than Spanish natives in their L2). If it had been language based the reverse relationship ought to have been observable in the data for Spanish natives. The gesture results obtained when comparing the L2 and L1 retellings also varied by number of phonic groups, number of words or length. Figure 22 shows a decrease in gesture rate from L1 to L2, an increase, or no change depending on the rate used.

The number of words used seems to be very dependent on the proficiency of participants, thus for the Spanish natives, more fluent in the L2, the difference between the L1 and the L2 words per second rates is not as large as for the HK students. These results suggest that phonic group might be a better denominator to use as it is less dependent on the proficiency of the language spoken, Spanish or English (at least for these participants). However, the length of the utterance, related to the phonic group, has been linked to the content (increasing with easier content or a lesser cognitive load) (Bortfeld et al., 2001). Therefore, even looking at the data using phonic groups as the denominator might not be adequate, if the retellings are based on different input. Obviously further research is necessary to identify a suitable conversion rate to compare results from speakers of different languages.

7.2 Disfluencies from all Retellings

The expectation was to see more disfluencies in the L2 than in any other type of retelling in the L1.[78] This is not always the case, suggesting that disfluencies might be related to much more than proficiency issues. Another possibility is that not all disfluencies are so related. We suggest that most often, especially in cases of continuing utterance after cut-offs and elongations, this indicates a cognitive conflict related to the conceptualization or retrieval from memory of the idea, rather than an error. Most of these disfluencies are used

[78] Initially data from all L1 retellings were analysed for disfluencies.

similarly in English and Spanish. Other disfluencies, 'uh' fillers in particular, might also function as words with a specific meaning, especially in English, that aid the management of the conversation.

Each participant produces a different number of disfluencies (by phonic group per second), per retelling. But we can observe a similar pattern among participants within each group (except for one individual in each group): a decrease in disfluency from the written to the aural retellings; an increase in video L1 (although not as many disfluencies as in the written retelling); and a decrease in disfluency in the L2 (Spanish natives) or an increase (HK students). The higher rate of disfluencies in the written retelling could be explained by anxiety, this was the first retelling and participants might have been nervous – which might affect the rate of some disfluencies.[79] The video depicted a more complex story than the previous two and this might affect the cognitive load. In addition no words were provided in the video so participants had to translate the concept into words, thus adding to the cognitive load. Finally, for HK students, less proficient than Spanish natives in the L2, there was also an element of needing to search for the right word, explaining the increase in disfluency in the L2. Participants from both groups skipped the description of at least one of the scenes in the retelling in the L2, another attempt to catch the bird (Sylvester tried another three times to get Tweety in the second part of the video). This could be because the cognitive load involved in using the L2 had become too great, affecting working memory. However, it did not seem to affect the translation of the concepts into words, as there is no increase in disfluency rates among Spanish students. This suggests that Spanish natives were more concerned with fluency than accuracy in both languages.

Brown et al. (2014) suggest that there might be differences in disfluency (and gestures) between monolingual and bilingual[80] speakers. Their study found that the average number of disfluencies was the same in both groups at 6.5 but monolinguals averaged 167 words[81] per narration and bilinguals 178. This gives an average rate of 0.039 disfluencies per word for monolinguals and 0.036 disfluencies per word for bilinguals, which are

[79] Except fillers (Mahl, 1987).

[80] Equally fluent in two languages from infancy.

[81] Their definition of 'word' is not known.

considerably lower rates than those we observed in the L1 for both groups of participants. If anything, our results contradict those of Brown et al. as rates for HK students (all bilinguals and some trilinguals, Cantonese, English and Mandarin) are higher than for Spanish native monolinguals. However, Brown et al.'s study focused specifically on the impact of cognitive load which could also affect the results.

Table 43. – Average number of disfluencies per number of words in the L1

	Spanish natives	HK students
Written	0.16	0.19
Aural	0.16	0.23
Video	0.19	0.21

The main difference between the results of the present study and those from Brown et al. is no doubt related to the different definition of disfluency used in the two studies. Their count of disfluencies included: "[sic] filled pauses (only 'hums', 'er's', 'eh's'), unfilled pauses (1 sec or more), repetitions/stutter, speech repairs" (Church, personal communication, August 2014). Brown et al. counted both disfluency and repair under one category and most importantly they did not count elongations as a disfluency. As the categorization is different from ours, we feel a comparison will not shed much light on the present study. Another study, Bortfeld et al. (2001), found disfluency rates of 0.06 per word, in English as the L1. Oviatt (1995) noted 0.055 disfluencies per word in face to face dialogues (also in English as the L1). Rodríguez and Torres (2006) reported similar disfluency rates per word in Spanish, 0.053. However, all the studies mentioned above included in their count disfluencies and repairs.[82] A comparison with their data is not relevant either as our count of disfluencies did not include repairs. Nevertheless our rates are quite different, suggesting a different measuring methodology that might be addressed through collaborative studies.

[82] Bortfeld et al. included as disfluencies: "repeated words or phrases, restarts, and fillers" (2001, p. 135).

This categorization issue is one of the biggest obstacles in getting a clear picture of disfluency (and gesture) performance. Data on gesture is even harder to come by, even if in most studies with speech-gesture transcriptions the disfluency is also recorded, it is seldom analysed for its part in the communication act. Even when recorded, categorizations tend to be general, such as bringing together all filled pauses under one heading. The relevance of this observation is that researchers cannot compare results from different studies if the categorisation of disfluency varies and if rates (of disfluency or gesture) are not calculated based on the same denominators.

7.2.1 Gesture – disfluency relationship in all retellings

The main hypothesis was to confirm whether there is a relationship between speech and gesture during disfluency. There are two types of relationships we have discussed. One is casual, the gesture is being produced because there is a disfluency, in which case the Lexical Retrieval hypothesis (Krauss, Chen & Gottesman, 2000) would be supported. The other is inter-dependency between gesture and the speech, both forming a single unit, in which case the gesture would be likely to show the same patterns as speech during disfluency (a suspension and a resumption point at the least, supporting Kita and Özyürek, 2007, and McNeill and Duncan's, 2000, theories). If there is no relationship and gesture is independent of speech, there should not be any significant match between gesture and speech performance during disfluencies.

If the gesture was caused by the disfluency, as a strategy to compensate for a speech related problem, then the data should show more disfluencies occurring with gestures than without them. In the L1 data (including all types of retellings) this is not the case, when comparing the number of gestures each participant produced we find that more disfluencies are produced without gestures in both the Spanish natives' data (on average 62% of all disfluencies do not have a gesture) and in the HK students' – 53% of all disfluencies do not have gestures (see Figure 22). We found these differences to be significant in all retellings for Spanish natives and in the written input retellings for HK students (see Table 20). Furthermore, if the gesture was caused by the disfluency most gestures should be starting with it, and instead we find that in the L1 only 13% of gestures start at the disfluency in the Spanish natives' data and just 4% in the HK students' (see Figure 45). Looking at all the gestures produced in the L1, we observe that in all participants there are gestural phrases not containing any disfluencies. From this data we would suggest that there does not appear to be a causal relationship between the disfluency and the gesture in the L1 (the gesture appearing because there is a disfluency). Rather it would seem that the gesture is paralleling the speech and gets caught up in the same conflict that caused the disfluency, so it is likely that the function of the gesture is not just to aid speech during conflict.

In the L2 the results suggest a more complex explanation, there is an increase of disfluencies accompanied by gestures, 60% of all gestures in Spanish natives and 65% in HK students (see Figure 22), significant in both cases (see Table 20). Although the gestures do not usually start at the point of the disfluency, just in 6% of cases in the data of Spanish natives and only in 2% of cases in the HK students' data (see Figure 45). For most participants (3 out of 4 among Spanish natives and all HK students) there are more gestural phrases that include at least one disfluency, than gestures without disfluency – as was observed in the L1. From these results it appears that the L2 disfluency in speech production might be different from the L1 disfluency in terms of the gesture produced, at least for lower proficiency levels (the HK students). Thus at this point in the analysis the data does not allow us to reject the Lexical Retrieval hypothesis for the L2 data. Maybe for lower proficiency levels the gesture is used as a strategy to compensate for speech difficulties.

In both the L1 and the L2 there seems to be a relationship between the disfluency and the gesture. In order to conclude what type of relationship exists, a detailed analysis of the gesture and the disfluencies is necessary. It has been identified that the performances in the L1 and the L2 are not necessarily the same and the data ought to be analysed separately and compared at each point of the analysis. In addition there might be differences in disfluency production according to types of input. Also, different disfluencies might signal different conflicts, or they might not be disfluencies at all. Thus the data needed to be analysed bearing in mind all these possible variables.

7.2.2 Types of disfluency in the video L1 and L2 retellings

The effects of different types of disfluencies (fillers and also pauses and some repairs) on the addressee have been studied to identify their possible effect in understanding and remembering the utterance. Clark and Wasow (1998) suggested differences between the fillers (a disfluency) and the repetition (a repair) which establish continuity and are used as a device to ensure smooth and timely speech (pp. 236-239). Bortfeld et al. (2001) indicated differences in disfluency by content. Therefore as well as analysing the data by language it was also relevant to analyse it by input and type of disfluency (although Corley and Hartsuiker (2011) indicated that it might not be the type of disfluency that helps process speech but just the fact that there is a delay).

The analysis of disfluencies in the L1 versus the L2 began with an overall assessment of disfluency preference, it was found that there were considerable differences by language, speaker and type of input (see Figure 18), but we could not find any particular patterns by participant or gender. There were too many variables to get a clear idea as to disfluency use so each type of disfluency was analysed separately. As our interest was specifically in disfluency with gesture by language, in most cases only these events were used to develop the discussion. If further questions arose or if it was deemed necessary to have a bigger sample to confirm any observations, the full corpus was analysed (all disfluencies with and without gestures and, if necessary, all the data from all the retellings using the L1).

One of the objectives of this study was to find whether Spanish natives produced any particular type of disfluency when speaking Spanish that might be different from disfluencies produced by HK students, thus we compared disfluencies by language, L1 versus L2. As before, what we observed was that differences exist but these might not be related to the language itself but to the proficiency level of the participants.

7.2.2.1 Pauses

A comparison of the average pause length (all pauses in all retellings) showed significant differences between pause length in L1 and L2 retellings in HK students but similar lengths in Spanish natives (see Table 19). This result was expected: in participants with lower proficiency levels speech is less fluent, with more and longer pauses. Pure pauses, those neither followed nor preceded by another type of disfluency, had similar average lengths in the data from the written and video L1 retellings in both groups. It has been stated that pauses are dependent on the cognitive load (Heldner & Edlund, 2010), with longer gaps signalling heavier cognitive loads. We believe that longer pauses in the video L2 are related to the increased cognitive load related to the level of proficiency. However this is just speculation and would require more accurate measurement of the pause (as Kendall, 2013, details in his study) and a larger sample to confirm any results. In an attempt to limit the scope of this study only those pauses followed by or preceding a disfluency were analysed, the pause being considered part of the disfluency. The average pause length per type of disfluency varies considerably, in particular by type of filler (see Figures 10 and 20). This will be discussed below under each type of disfluency.

7.2.2.2 Cut-offs

Overall we observed a higher proportion of cut-offs in the HK students than in the Spanish natives, this could be due to a combination of factors, including the lower proficiency level in the L2. Data on cut-offs was analysed as percentages which were compared between groups and types of cut-off (within word and after word completion). The distribution by type of cut-off is similar in both groups in the L1 and in Spanish natives L2. Although Nooteboom (1980) believed suspension of speech appears as soon as the trouble is detected (Main Interruption Rule, MIR), evidence suggests that it is more likely that speech is interrupted once repair processing has been completed (Seyfeddinipur, 2006, 142), known as the Delayed Interruption for Planning hypothesis (DIP). This seems to indicate that most speakers are more interested in fluency than correctness (except HK students in the L2).

Levelt (1989, pp. 478-480) analysing his own 1983 study results on self-repair, found that the majority of cut-offs (or interruptions, as Levelt refers to them) occur just after the completion of the whole trouble word (making up 51% of overt[83] self-repair cut-off cases). His other categories of cut-offs included:

- Immediate within-word interruptions (totalling 18% of overt cut-offs in the corpus)
- Delayed interruptions, by one or more words (31% of overt cut-offs in the corpus).

[83] Levelt did not include covert repairs in this analysis as it was not possible to identify the source of the trouble, such as in his example: *Here is a –er a vertical line*. And therefore not possible to say at what point the trouble had been identified (1989, pp. 478-480).

- Delayed within-word interruption (a subcategory of the above, 4% of the 31%).

Adding[84] all instances of cut-offs after word completion and within the word his corpus showed 78% of cut-offs after word completion and 22% of within word cut-offs. Our results (bearing in mind we had included covert and overt errors or conflicts) are very similar for the L1 results, from all the retellings (see Tables 32 and 33).

Table 44. - Types of cut-off by percentage in the L1.

	After word completion cut-off	**Within word cut-off**
Spanish native speakers	72%	28%
HK students	78%	22%

However the L2 results are very different, showing a significant increase of within word cut-offs in HK students and a decrease among Spanish native speakers.

Table 45. - Types of cut-off by percentage in the L2.

	After word completion cut-off	**Within word cut-off**
Spanish native speakers	91%	9%
HK students	35%	65%

We believe that part of the explanation might lie in the proficiency level of each group. In the case of HK speakers, with lower L2 proficiency levels, there seems to be an overall preference for correction rather than fluency (the opposite is true in proficient speakers). In addition we speculate whether the syllabic nature of Cantonese might be playing a part as well, leading to more within word cut-offs. Although Spanish is also a syllabic language, native Spanish speakers tend to concatenate syllables, which HK students do not do[85]. It could be that under higher cognitive constraints, as is the case in lower

[84] To compare his data with ours this was necessary as our results on cut-offs did not refer back to an error to identify whether the cut-off was a delayed repair or not.

[85] Indeed, one of the author's main challenges as a Spanish L2 teacher to this group is to ensure they learn to join syllables when speaking in Spanish.

proficiency levels, HK students' speech articulation mechanism was trying to emulate what it already knew. In addition, it is possible that, trying to pronounce clearly and carefully – leading to repetition of syllables, stuttering almost – might have resulted in more within word cut-offs, some corresponding to disfluencies (and repairs) while others might be related to an increase awareness of correction. Spanish natives were more proficient in English than HK students in Spanish, so there were less cases of stuttering, and Spanish natives seemed more concerned about fluency, albeit their speech rate was slower in the L2 retellings (calculated as words per second) than in the L1. It could also be that they are concerned about the addressee's perception, and were not confident enough to interrupt words.

According to Seyfeddinipur, Kita and Indefrey (2008), there ought to be an obvious difference in the number and length of the pause that follows the two types of cut-off. When speakers notice a conflict and have a repair ready they might interrupt themselves mid-word (within word cut-off), and if there is a pause this will be relatively short, as the repair is ready. If speakers do not have a repair ready they are more likely to finish the word (another reason why the Spanish natives might have a preference for completing words in the L2) and will probably pause as well, a longer pause than with a within-word cut-off, as they need time to prepare the repair. Thus the expectation was to see more within word cut-offs in the L1 than the L2. Among Spanish natives this prediction is met, there are indeed more within word cut-offs in the L1 than in the L2. In HK students it is not met, as there are more within word cut-offs in the L2 (see Figure 38). As previously mentioned, this could be due to the low proficiency of this group of speakers in the L2 which results in frequent stuttering (each counted as a case of within word cut-off).

In terms of pauses the expectation was to see more pauses following cut-offs after word completion than following within word cut-offs. In all participants' data this prediction is met in both the L1 and the L2 (although in almost all categories there are more cut-offs of both types without pauses). The length of the pauses was expected to be longer with after word completion cut-offs, this is indeed confirmed in all cases, with results proving to be significant at the 0.05 critical alpha level (see Table 34). When speaking in an L2, it is expected that speakers will have more conflicts linguistically (more or less the same conceptually) than when speaking the L1, therefore we would expect to observe longer pauses. However, average pause length in the L2 cut-offs was longer than in the L1 but not significantly so (see Figure 37 and Table 44).

Thus for both Spanish natives and HK students' data we can state that the results – excluding HK students L2 data – support the Delayed Interruption for Planning hypothesis (DIP): speakers stop mid-word if the repair is ready and the pauses following these cut-offs is shorter than in the cut-off that occurred after the word had been completed. We believe that the differences observed in the HK students' data in the L2 are related to their lower proficiency level resulting in articulation issues that provoke stuttering. This was considered a within word cut-off and perhaps it is distorting the results.

To rule out the possibility of having miss transcribed the cut-offs as disfluencies (instead of just longer pauses), the after word completion cut-offs in the Spanish natives L1[86] were analysed by type of repair that followed them (see Table 35), only in 5% of cases did the utterance just continue (mostly there was a repetition, 39% of all cases). When taking all the data points for disfluencies with gesture for all participants (for video L1 and L2 data) we found that 4% of all after word completion cut-offs were followed by continuing the utterance. Thus we confirm the unlikelihood of a large error in the transcriptions of the L1 data for Spanish natives. It is also unlikely to be an error just in the transcriptions of one group as all transcriptions were carried out by the same researcher for both the L1 and L2, and the same parameters were followed in both cases to identify the interruption.

7.2.2.3 Fillers

Clark recorded 'uh' and 'um' type fillers as accounting for 30% of all repairs – his term (1989, p. 483). The equivalent Spanish filler to 'uh' and 'um' is what Rodríguez and Torres (2006, p. 338) described as: "sounds like the vowel /e /", that in our corpus has been transcribed as 'eh' and 'em' (although we also found cases of other vowels such as 'uh' and 'um', 'ah' and 'am'). If their assumption is correct, then fillers are one of the most frequent

[86] The data used for this was that of disfluencies occurring with gestures in L1 retellings in Spanish natives.

disfluencies in their corpus.[87] In our video corpus we found that on average 31% of disfluencies were fillers (although by language and group of speakers the range went from 18% in Spanish natives L1 to 41% in HK students L1), as shown in Figure 31. In order to understand the data better, a more in-depth analysis of fillers was carried out, using only the L1 and L2 video data (to avoid possible differences due to the content and type of input of the retelling). An initial look at the results would suggest that HK students use more fillers than Spanish students. When the data was normalised to account for different number of phonic groups per second these results were confirmed.

Based on the work done by Clark (1996) the expectation was to see a significant difference between the use of 'uh' and 'um', and the pauses preceding and following them (Clark & Fox Tree, 2002, p. 82). These authors suggest that the two fillers fulfil different functions, with the 'uh' indicating a lesser disfluency than 'um'. If that is the case we would expect to see a lower incidence of 'um' during the L1 retellings (lesser cognitive issues), compared to 'uh'. We would also expect an increase of both 'um' and 'uh' in the L2 retellings, as the syntax, lexical and phonetic conflicts in the L2 would be more serious than in the L1. Indeed, overall we observed more fillers in the L2 (see Figure 40). We were also expecting a higher incidence of 'um' in the L2, in comparison to 'uh', as linguistic difficulty increased. The expectation was met in the data of Spanish natives, where the percentage of 'uh' decreased to 41% (from 68% in the L1) and there was an increase in 'um' to 59% (from 32% in the L1).[88] However in HK students' data the proportion of 'uh' versus 'um' was similar in both languages, with a slight decrease in 'um' in the L2 (from 45% to 43%).[89]

The data was explored further with the aid of pauses following and preceding the fillers. If 'um' indicates a more serious conflict then the pause following and preceding it ought to be longer than those of 'uh'. There should also be more pauses with 'um' than 'uh', as suggested by Clark and Fox Tree (2002, p. 95). Figures 10 and 20 show the average length of pauses before and after each type of filler. Contrary to what Clark and Fox Tree found,

[87] It is hard to give a percentage that compares to those of Levelt (1989) as the Rodriquez and Torres (2006) classification is not identical.

[88] These results were significant to a critical alpha level of 95%.

[89] These results were not significant.

our results indicate longer lengths before ‘uh’ than after it,[90] however this could be related to the fact that ‘uh’ often follows another type of disfluency. Apart from this discrepancy, our results support those reported by Clark and Fox Tree. Pauses after ‘um’ are longer than those following ‘uh’, in particular in the L2 retellings where the average pause length in Spanish natives is 0.25 seconds (after ‘uh’) and 0.73 seconds (after ‘um’). It was also observed that Spanish natives prefer to keep pauses following ‘uh’ much shorter than HK students (in both the L1 and L2), while HK students make use of longer pauses before and after ‘um’.

As speech difficulty increases, the use of fillers increased in the L2 and the pauses associated with them were longer. But, considering their proficiency level in the L2 was quite high, we were surprised to observe such an increase in filler use from L1 to L2 in the Spanish natives. In particular we were surprised at the large differences in the length of the pauses, and the drop of number of fillers followed by pauses (compared to an increase of fillers used without a pause) in the L2. In HK speakers the use of fillers followed by pauses also decreased significantly in the L2.[91] In most past disfluency studies, fillers have been included in the same disfluency category as cut-offs and elongations. However, these results indicate that fillers behave differently from other disfluencies.

Our results support those of Clark and Fox Tree (2002), suggesting a different use of ‘uh’ and ‘um’. They associate the filler ‘uh’ with lesser conflicts, perhaps the difference is not a ‘lesser’ conflict but a different type of conflict. We speculate whether ‘uh’ could be related to the linguistic difficulties and ‘um’ to cognitive content issues. This would mean that HK students, who use slightly more ‘uh’ in the L2 are less worried about cognitive content than linguistic difficulties. On the other hand, Bortfeld, et al. (2001) suggest that fillers might be used as signals of speech difficulty to addressees, helping them “coordinate with their addressees”, and might not be related to cognitive difficulties alone (p. 142). But as our data is based on monologues this hypothesis does not explain the differences observed in our results.

[90] Except in HK students L1 video retellings.

[91] As mentioned in the above section on pauses. See Figures 26, 27 and Tables 25 and 26.

There might be other variables at play that we are not considering, such as linguistic differences that lead speakers to use different disfluency strategies. One possibility could be that there is a preference to use fillers in English, while in Spanish other disfluencies (such as elongations) are preferred. This would explain the overall lower number of fillers in Spanish natives' L1 data and their different breakdown in HK students' data in the L1. In the L2, Spanish natives show a different pattern from the L1. This could be because they experienced the L2 in a naturalistic context (or had spent considerable time among English natives) and would have had the opportunity to pick up on the use of fillers from native speakers, therefore their use in the L2 is closer to that of HK students. This explains differences in the rates of 'uh' versus 'um' but does not clarify the difference in pauses. Another explanation, proposed by Kjellmer (2003), is that fillers might have the linguistic function of indicating a new thought unit (the assumption being that a new thought will convey new information). Nevertheless this does not help explain why there are overall more fillers in HK students than in Spanish natives. Also we believe that it is more likely that the newsworthiness element is linked to the pauses (and phonic groups) rather than with fillers.

Another theory suggests that disfluencies are an aid to the addressee, rather than the speaker. The team Corley, MacGregor and Donaldson carried out a series of experiments testing how disfluencies affect the memorability of words by analysing neural responses to repetition, pauses and fillers. Their results showed that listeners respond differently to repetitions than to fillers and pauses. Listeners are more likely to recognise words after pauses (MacGregor, Corley & Donaldson, 2010, p. 3982), and fillers help increase recall of the preceding words, perhaps aiding with the difficulty of integrating unpredictable words (more likely to be newsworthy) into the utterance (Corley, MacGregor & Donaldson, 2007, p. 666). If that is the case we would, if anything, expect more fillers in the L1 where the use of a wider vocabulary is more likely.

Returning to the differences between 'uh' and 'um', if 'um' indicates a more serious conflict it might be possible to observe this reflected in the use of gestures. It had been noticed that not all gestures stopped with the filler, and we speculated whether a relationship based on the use of 'uh' or 'um' could be found. The fillers with gestures from the video data were analysed looking for fillers where the gesture had been put on hold during the utterance of the filler or had continued (if the gesture had been on hold when the filler commenced this was also recorded). The data shows that on average in 35% of all 'uh' and in 22% of 'um' the gesture continued during the filler (see Table 40). This is an interesting result, which feeds into our speculation as to the categorisation of fillers as other than a disfluency. If gesture and speech are part of one system (following McNeill's GP theory that speech and gesture are one unit externalised in two modalities), then there ought to be a hold in the gesture during the filler (as there is during cut-offs, see Figure 53). The fact that gesture is more likely to continue through the 'uh' could indicate that the filler 'uh' is not a disfluency but a word, this supports Clark and Fox Tree's concluding remarks (2002) that 'uh' and 'um' ought to be considered words, their meaning being "I am initiating what I expect to be a minor delay in speaking" (p. 107). We would add that perhaps 'um' signals a conceptual issue, reflected on the gesture being less likely to continue during the 'um'. The differences in percentages between continuing and stopping gestures are only significant in some cases (see Table 41) so these conclusions are just speculative.

7.2.2.4 Elongations

Elongations were the most frequent type of disfluency encountered in the data of Spanish native participants, in both the L1 and the L2 (this supports the results of Rodríguez and Torres, 2006, p. 346). There were considerably more elongations in the utterances of Spanish natives in both the L1 and L2 than in HK students in the L1 (see Figure 34). Furthermore, it was found that the average number of elongations by phonic group per second among Spanish natives was the same in the L1 and the L2. In all retellings of HK students in the L1, elongations were used less than other types of disfluencies (in most participants), but their use increased in the L2 (see Figure 28). Further analysis of elongations,[92] including the pauses following them, suggested that Spanish natives and HK students use elongations differently. In particular Spanish natives prefer to use elongations without a pause (see Figure 34). We speculate whether the elongation itself might be taking the function of the pause. HK students in the L1, where elongations are used less than other

[92] For this part of the analysis only disfluencies occurring with gesture in the L1 and L2 video data were used.

types of disfluencies, do not show a tendency to produce elongations with pauses rather than without them. It seems that they are not as familiar as Spanish natives with the 'elongation instead of pause strategy'. However, in the L2 the use of the elongation in HK students' data increases to a 34% of all disfluencies observed, from 24% in the L1 (see Figure 31). There seems to be a prevalence of elongation and pause together (71% of all elongations with gesture in the L2), with longer pauses than in the L1, comparable to the length of pauses Spanish natives use in the L1 (see Figure 34 and Tables 30 and 31).[93]

When considering the type of repair that followed elongations, it was found that in over 80% of elongations without a pause, the utterance just continued (this was true in all the groups except HK students in the L2). For elongations with a pause the results were less clear, although there was also a high prevalence of the utterance continuing. We observed that in over 20% of all cases the utterance continued with a filler (see Figure 35). If the elongation is in itself a disfluency, analysing the gesture that accompanies the elongation should provide support evidence for treating it as a disfluency: as the gesture and the speech ought to be synchronous – with a higher likelihood of the gesture preceding speech (Seyfeddinipur, 2006). If the elongation indicates a conflict, the gesture ought to also indicate the conflict by being put on hold or stopping altogether. When looking at the gesture, whether it continued or was put on hold with the elongation, it was found that the prevalence in all speakers was to use elongations with continuing gestures, although in a smaller proportion in HK students (see Figures 54 and 55). The fact that the gesture continues through the elongation (it is not put on hold) might indicate that elongations might be speech traits and not necessarily a disfluency: either elongation is a disfluency more likely to occur in Spanish or it could be that the elongation is not so much a sign of disfluency, but a regional speech characteristic of our participants (all from Southern Spain). If this is so, then the data related to elongation (in particular elongation without a pause) ought not to be considered as it is not really a disfluency, but a speech trait. The preference of Spanish natives to elongate a sound rather than to create a pause is likely to be related to the tendency to link syllables, perhaps as a strategy to keep the floor by minimising the number of pauses. Elongation is also used more by HK students in the L2, in particular by

[93] A possible explanation for this is mentioned below.

those who have spent time studying in Spain (a month-long Spanish language course in Málaga in C.S. 2HK's[94] case and six months in Jaén as an exchange students in the case of C.S. HK4). It is likely that access to native speakers has increased their subconscious awareness of the use of elongations in Spanish adopting it without realising.

7.3 Gestures

De Ruiter (2007) and Melinger and Levelt (2004) proposed a trade-off relationship between gesture and speech in terms of communicative content. This hypothesis states that the harder speech becomes the higher the likelihood of gestures taking over the communicative function. That is: the more the speaker has to think or the more thought units, the more the hands come into use. If we also believe that the disfluency (pauses, fillers) and repairs might indicate thought units, the more complex the cognitive demand the more disfluencies will be expected. Nicoladis, Pika, Yin and Marentette. (2007) also report observing more gestures with higher cognitive loads. This is indeed the case, when comparing the number of disfluencies per phonic group and that of gestures per phonic group, by speaker, type of input and language, we find similar patterns in the data (see Figures 17 and 21). Based on this hypothesis we expected to observe an increase in the number of gestures in the L2 video, as, we assume, participants had more cognitive and lexical difficulties when using the L2. This would confirm results like those of Slezak-Swiat who, in a study of paralinguistic strategies, found that L2 speakers of English involved in successful communications used gestures 20% more than speakers who were unsuccessful in their communicative intent (2010, p. 72).

The retelling of the video in the L2 tended to be longer in HK students, than in Spanish natives (see Tables 6 and 7). When the data was normalised as gestures per phonic group per second it was found that in only four cases were the gesture rates higher for the L2 than the L1 video (see Figure 21).[95] When calculating the number of gestures per second

[94] Participants were numbered one to four: 'case study 1 Hong Kong' (C.S. 1HK) or 'case study 1 Spain' (C.S. 1Sp).

[95] Gesture rates per number of words indicated that both groups performed more gestures in the L2 than the L1, supporting the results of Slezak-Swiat (2010).

or per phonic group, Spanish natives seemed to gesture more than HK students, and produced more gestures in the L1 than the L2 (in HK students the rate increases from the L1 to the L2). The number of phonic groups per second increases in both groups (Spanish natives and Hong Kong students) from the L1 to the L2, although total retelling lengths do not vary significantly. Therefore the phonic groups are shorter in the L2. Similarly the number of words per second decreases in the L2 and speech speed becomes slower. As in both groups there is a decrease in number of words and an increase in number of phonic groups per second from the L1 to the L2, we have to conclude that the variations are probably attributable to the differences in proficiency level (there could be other variable we have not considered, such as type of language and further research with different proficiency groups and languages would be required to confirm this). HK students speaking in the L2 increase the number of gestures per phonic groups, but phonic groups are getting shorter, this means their speed of gesturing is increasing (more gestures in shorter phonic groups). Meanwhile Spanish natives in the L2 are gesturing less per phonic group (than in the L1), but as the phonic groups are shorter it is likely that the speed of their gestures might not have changed, or might have decreased (further calculations outside the scope of this study would be required to confirm this). Our hypothesis is that the less proficient L2 students gesture faster than the more proficient ones.[96] Faster gestures could result in less well defined gestures. There could also be another explanation related to the type of gesture, as an iconic gesture might take longer to be realized than a pragmatic one, but it seems unlikely. A different gesture speed would help to address some of the inconsistencies that have appeared in the literature as to whether speakers gesture more in the L2 than in the L1 (related also to inconsistent denominators being used to compare different studies).

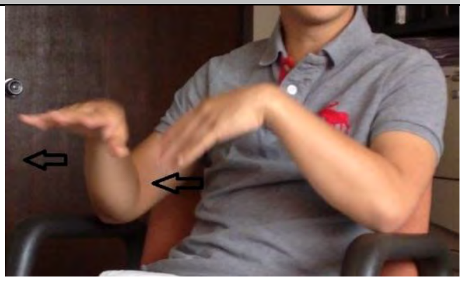
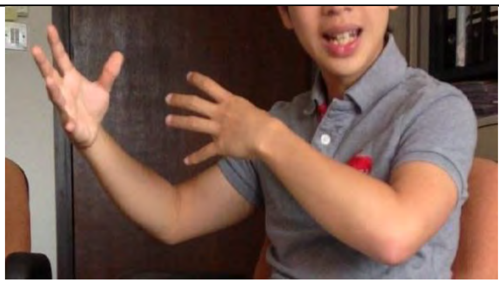
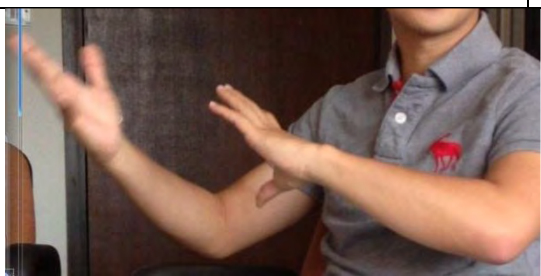
Graziano and Gullberg (2013) found that adult L2 learners (Dutch speakers learning French in a formal context) performed more gestures during disfluencies than other groups tested. We found our results supported their findings (see Figure 22). Both Spanish natives and HK students performed more gestures co-occurring with disfluencies in the L2 than in the L1. Gullberg (2014) suggests that this could be due to an increased use of gestures to elicit help from the addressee, she indicated that in their corpus a high proportion of these

[96] Nicoladis (2007) found that bilinguals gesture at higher rates than monolinguals. It is possible that HK students, all bilinguals in Cantonese and English, perform gestures faster than monolinguals.

were iconic in nature (speakers trying to explain the missing word with the gesture). Gullberg (1998, 2003) also indicated that lower proficiency L2 speakers are likely to use more noun phrases and with them more referent gestures. These would have been considered iconic gestures in our data. Based on these studies we expected some differences in the type of gestures by fluent and disfluent speech and by language, L1 versus L2.

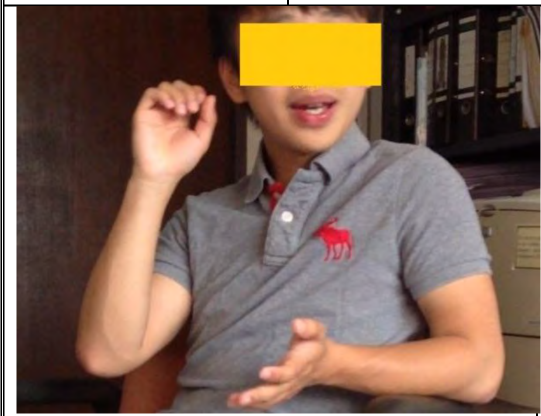
What we observed was that the less proficient HK L2 speakers gestured constantly but their gestures did not always represent metaphorically or iconically the content of their speech. These speakers were not using gestures to construct a visual representation of the content they were unable to express verbally. An attuned observer could recognize a number of gestures as being of a deictic nature, where the speaker was using them as referents to the actors or objects of the narration, or their position (however another interpretation could be that most of these are similar to the other pragmatic gestures observed). Our belief is that often they are being used as referents. These gestures coincide with nouns in the speech and had been used in the English version of the video narration (a similar shape had been observed), therefore they had been classified as iconic.

For example, in the L2 video retelling of C.S. HK3 we observed four instances of iconic gestures representing action verbs:

Line No.	Speech	Gesture	Description
116	voy a: I am going to:		*Hands move to the right.*
203	coge l: Grabs th:		*Both hands grab Tweety's cage.*
213	Salil Go out		*Hands move to the right and right hand flips palm up.*
215	saliendo del getting out of the		*Hands move to the right and right hand flips palm up (as in previous gesture).*

However the gestures that accompany these utterances were badly defined, imprecise in their shape, in contrast with this participant's gestures during the English narration which were sharp, well defined and much easier to recognize. The only gesture which was clear and well defined in this L2 retelling was the iconic gesture that accompanies the utterance that described the receptionist answering the phone.

Line No.	Speech
25	eh dígame uh hello



The rest of the gestures in this data from this participant were classified as pragmatic. Only in one case was there an obvious lexical retrieval gesture,

Line No.	Speech
42	en [la: in [the:

where the speaker circled the hand by his head as if trying to extract the next word. Other gestures correspond to the rhythm of the speech and seem to follow the syntax construction of the clauses, seemingly directing the beginning and end of clauses and the position of the various elements. Further analysis of this phenomenon, which supports Kita and Özyürek's IP hypothesis, falls outside the scope of this study but it has been observed in other low proficiency speakers (López-Ozieblo, 2013b)

In the data from the video L1 and L2 retellings[97] the results show that there seems to be a prevalence of iconic gestures, rather than pragmatic, in the L2 (see Figures 23, and 45 to 48 and Tables 21 to 23 and 38). Gullberg (2006) found that in the L2 there are usually more deictic gestures than in the L1, as we had included deictic gestures in the iconic category our results support Gullberg's. Iconic gestures amounted to 58% of all gestures

[97] Including gestures during fluent and disfluent speech.

performed during the L2 retellings in Spanish natives and 51% in HK students (see Table 23). In seven out of the eight participants, there is an increase of iconic gesture use from the L1 to the L2 (see Tables 21 and 22). These results are confirmed when the data is normalised as gesture per phonic group per second (see Table 38 and Figure 46). A reason for the use of iconic gestures might be the need to ensure the addressee understands and so resorting to 'drawing' the action, space or character with gestures (even if there is no disfluency in the speech). This being the case, we would expect a higher proportion of iconic gestures in the HK students (whose proficiency level in the L2 was lower than that of Spanish natives), as is the case. We would suggest that HK students might also use more iconic gestures in English because their addressee was a Spanish native and they might have exaggerated their iconic gestures to ensure their addressee understood them.[98]

Although there is a prevalence of iconic gestures there is also an increase in the number of pragmatic gestures (as a percentage of the total number of gestures in the retelling) among HK students from video L1 to video L2 retelling; we believe this is due to the inclusion of what could be called lexical retrieval gestures, such as waving the fingers or flicking the hand to aid tip of the tongue lexical item, retrieving words the student knows but does not remember (while iconic gestures might be used to explain words they just do not know, as suggested by Graziano and Gullberg, 2014). This hypothesis, however, does not hold true in the case of Spanish natives who produce on average less pragmatic gestures in the video L2 retellings than in the video L1 (see Table 38). We would like to advance two theories to explain these results. The first is that their gestures might be faster in the L1 and these might not be as well formed as those of the HK students and thus not recognised as iconic. Perhaps if the gesture was slowed down their iconicity might become clearer. The other possibility is that they are indeed a type of pragmatic gestures, beats, and that there is a prevalence of their use for keeping the rhythm or providing stress in Spanish. As the data did not differentiate beat gestures and other types, such as lexical retrieval gestures, we cannot confirm this observation.

[98] If the addressee is affecting gesture production we suggest this could be easily tested by repeating the exercises with an English native addressee.

7.3.1 Types of gestures during disfluencies

When analysing only gestures occurring with disfluencies, data from all retellings by Spanish natives in the L1 and HK students in the L1 shows a prevalence of iconic gestures with disfluencies. Spanish natives tend to use iconic gestures in the L2 as well but show a prevalence of pragmatic gestures (see Table 24). However the breakdown of pragmatic versus iconic gestures is quite close and not significant. Comparing the data just from the L1 and L2 video retellings, the main difference observed was a higher proportion of pragmatic gestures in HK students in the L2 than in the L1. For other groups we observed the same pattern as that obtained when analysing all gestures occurring both with and without disfluencies (see Figures 50 and 51). Taking the average number of gestures co-occurring with disfluencies for all groups, a higher incidence of pragmatic gestures, rather than iconic, is observed (see Table 38), although the difference could be due to the data from HK students in the L2. A prevalence of pragmatic gesture during the disfluency would tie in with the Lexical Retrieval hypothesis, the pragmatic gesture being used as a strategy to prime the lexicon. When analysing the description of these pragmatic gestures we observe finger waving and hand flipping, but most gestures are just marking the syllable, the gesture stroke coinciding with the syllable. It is interesting to note (although it falls outside of the scope of this study) that during the aural retells there was a significant preference of iconic gestures in both groups.[99] One possible explanation could be that the aural input contained more low frequency words than the written text. Nicoladis et al. (2007) suggest that more iconic gestures are observable with low frequency words as these are harder to access, so an iconic gesture is used to provide their meaning. The second part of the video, the part retold in the L2, also contained a scene (involving a fulcrum and a weight falling) that might require low frequency lexicon. This might have increased the use of iconic gestures in the L2. However, as that scene was often missed in the retellings, we cannot confirm that low frequency lexicon is a decisive factor.

As discussed under *Disfluencies* (Section 7.3), there seemed to be differences in the use of the various types of disfluency analysed: elongations in Spanish natives might be speech traits, not disfluencies; and fillers might not be disfluencies but words. Therefore it

[99] Spanish natives: $t(29) = 4.81$, $p = 0.0000$. HK students: $t(87) = 0.24$, $p = 0.018$.

was decided to analyse the gesture type again, but just in overt disfluencies: fillers and within word cut-offs (see Table 39 and Figure 52). Most pragmatic gestures during disfluencies in Spanish natives in the L1 seem to be related to within word cut-offs (we observed a significant 78% of pragmatic gestures versus 22% iconic). In the L2 the percentage of pragmatic gestures in cut-offs descends to 67% (not significant). Among HK students there is a preference to use iconic gestures in the L1 in within word cut-offs (67%) which is reduced to 43% in the L2, where more pragmatic gestures are observed. It is puzzling to observe this opposing trend between Spanish natives and HK students in gestures particularly as the use of within word cut-offs is similar in both groups in the L1. In the L2 the differences might be related to the classification of the within word cut-off (as mentioned previously, stuttering might have been counted erroneously as disfluency), but not in the L1. Therefore it seems that the difference is in the gesture. Apart from the suggestion mentioned above, that HK students were keen to ensure their Spanish native addressee understood, and so used more iconic gestures, we are at a loss as to why this could be, and suggest this be a topic of further research.

The breakdown in type of gesture by filler ('uh' versus 'um') is relatively similar in all cases (and the differences are not significant). In Spanish natives we observe more iconic gestures in the L1 (57%) than in the L2 (46%), and also in HK students (50% in the L1 and 41% in the L2). These results help explain the increase in pragmatic gesture in the L2. A further look at the type of gesture by type of filler did not provide any other conclusive results (see Table 42), the breakdown by type of gesture being very similar for both 'uh' and 'um'. However, in many cases the sample size was small and as the differences in the use of pragmatic versus iconic gestures was not large, we cannot categorically conclude that pragmatic gestures are preferred during disfluencies, although we cautiously indicate that pragmatic gestures are observed more often with overt disfluencies in the L2. These results indicate that some of the gestures might be used to aid lexical retrieval in the L2. A larger sample size would help clarify this issue.[100]

[100] The apparent prevalence of pragmatic gestures might be related to C.S.3HK who performed 32% of the total pragmatic gestures (seen in the L1 and L2 retellings together).

7.4 Speech and Gesture Processing Mechanisms

All speech processing models suggest a monitoring function that allows the speaker to identify and correct errors before they have been uttered. A signal of these and other conflicts might be a disfluency. Levelt suggested that the monitoring model asks itself the following questions (1983, pp. 51-55):

Do I want to say this now? – This is an error at a conceptual level.

Do I want to say it this way? – Here the speaker would worry about the pragmatics, and appropriateness.

Am I making an error? – These types of error could be at lexical, syntactical or phonetical levels.

The first two questions assume that the concept has been fully formulized. We agree with Clark (1996, p. 267) in that speakers don’t usually formulate the whole presentation in advance of uttering it. We would go a step further and argue that not only does the speaker not have the whole presentation ready but often not even the fully formed concept; this would tie in with a parallel speech processing framework. Some speakers’ ideas might be more dynamic, forming and dissipating without reaching a conclusion or concluding point and a new thought already appearing. Therefore the ideal utterance might be only in hindsight, and disfluencies and repairs might reflect the new shape of the concept. The work of Van Wijk and Kempen (1987) points in this direction: in their study, speakers were asked to describe a changing picture and the instances of self-repair were analysed. For example, the picture initially showed a man without glasses and as the speaker started describing the scene, the appearance of the man changed and glasses were added. Speakers’ resumptions following a disfluency based on the change in input, were directly related to changes in the visual stimuli. These resumptions were truly add-ons or deletions, reflecting a different concept from the original one the speaker had started to utter.

In the L2, especially at lower levels of proficiency, it is likely that speakers are going through a similar process: their original concept might not have been fully formed, or might change as speakers realise that they either cannot remember the event they had started to describe or lack the vocabulary to do so and need to rephrase. Therefore more disfluencies are expected when using the L2, as is the case when comparing most retellings of L1 versus retellings in the L2 (see Table 6). As, during the debriefing sessions, most participants denied completing the thought in their L1 and then translating it, this provides proof of the unlikelihood of speakers forming complete thoughts before uttering commences. If that was the case they would not start the utterance without knowing the words needed for it. The results based on cut-offs also indicate that speakers often start utterances without planning them in advance, otherwise there would be fewer after word completion cut-offs. We suggest that speakers are more likely to make use of a parallel speech process that works as a network of nodules that activate each other, those with the strongest activation becoming externalised as utterances.

Clark (1996, p. 241) argues that speakers make use of two simultaneous tracks in their communications: the primary track to cover the topics of the discourse and a collateral track referring to the performance which would include meta-communicative acts. He states that within the second track it is possible to find gestures, facial gestures and gaze as well as speech related pragmatics (pitch, intonation, etc.) suggesting two processors: one for gestures and one for speech. Other scholars (Kita & Özyürek, 2003; de Ruiter, 1998) agree with him in that speech and gesture are processed by different mechanisms. On the other hand McNeill and Duncan (2000) proposed a unified gesture-speech mechanism, where gesture and speech are the thought externalised.

In our results we have seen that there are differences in gestures produced with and without disfluencies and also seemingly by proficiency level (L1 and L2). We did not find any specific differences in gesture co-occurring with disfluency by language, although they are known to exist when co-occurring with action verbs. Negueruela et al. (2004) found that lower proficiency speakers of an L2 used gestures with motion verbs following Thinking for Speaking[101] of the L1, instead of the L2, resulting in gestures that would seem out of synch or redundant for a native speaker of the L2. There seems to be some evidence that the onset of the gesture occurs at slightly different points in fluent and disfluent speech, and it might also vary by language and even by type of gesture. Schegloff (1984), Hadar and

[101] For more details on the Thinking for Speaking theory see Slobin, 1996.

Butterworth (1997) and Morrel-Samuels and Krauss (1992) found that the onset of iconic gestures with fluent speech occurred mostly before the word (in Indo-European languages). However, Chui's (2004) results based on Chinese speakers' data contradict theirs. He found gesture onset preceding speech in only 35.6% of his corpus. Butterworth and Hadar (1989) observed that in non-fluent speech the onset of iconic gestures was more likely to occur during the disfluency pause, while beats (pragmatic) gestures occurred after the disfluency. Chiu, on the other hand, observed that in Chinese speakers the iconic gesture onset occurred mostly before the disfluency.

Our transcription recorded what happened to the gesture at the time of the disfluency (whether it continued, stopped, started or if it was held before the disfluency). Few instances of gesture being held before the disfluency were found, 7% of all L1 and L2 cases. Mostly the gesture continued during the disfluency (this study excluded pauses) in 56% of all L1 and L2 cases (see Figure 49). Regarding the onset of the gesture, although it is transcribed, no calculations were made and we suggest this ought to be the topic of further research. Few gestures started with the disfluency (6% of all L1 and L2 cases), but a difference was observed between the L1 and the L2. In the L1, three times as many gestures as in the L2 started with the disfluency. As the sample is very small, the significance of these results is questionable. We speculate whether this difference could be explained by the type of gesture used, whether iconic gestures (more used in the L1) are more likely to be stopped with the disfluency than pragmatic gestures. Our analysis on fillers does confirm that there is a slight tendency to stop (rather than continue) an iconic gesture during the filler (see Table 42). Spanish natives in the L1 are more inclined to stop the gesture at the point of disfluency (44% of their gestures) than other groups, who showed similar patterns averaging 28% of gestures within each group (see Figure 49). Further manipulation of the data might help shed light onto these results as there might be differences observable by type of gesture. In particular, data relating to elongations might be distorting the data if they are not disfluencies, so we propose to exclude these in further analysis. Thus we cannot state that our results provided conclusive evidence of a strong gesture-disfluency link or for a unified or dual (one for gesture one for speech) processing mechanism.

7.5 Second Language Instruction

In second language[102] instruction very little attention is paid to either disfluencies or gestures, probably because neither element has had linguistic status until recently and so was ignored by traditional instruction methods (López-Ozieblo, 2013a). With the emergence of the communicative method in L2 classrooms there ought to have been an inclusion of gestures in syllabi; this has not been the case. Only a handful of Spanish as a Foreign Language textbooks mention gestures, and these are always emblematic (García García, 2009). Monterubbianesi (2013) reported that in a study of 107 textbooks of Spanish as a Foreign Language (various levels) only 19% explicitly mentioned aspects of nonverbal communication (including intonation, eye contact, proximity distances, as well as gestures). Gesturing like a native projects an impression of fluency (Neu, 1990) and using a higher proportion of formulaic expression is perceived to be a trait of native speakers (Foster, 2001). Therefore we believe that both disfluencies and gestures ought to be part of the instructional curriculum, as they are part of the target language.

In a natural context, students of an L2 have access to native speakers and much is learnt from them, including the speech and gesture traits we are discussing. HK students usually do not have the possibility of contact with native speakers (except their teachers) and so have few opportunities to acquire communicative traits of the culture being studied other than in the classroom. Under these circumstances it might be necessary to draw students' attention to possible differences. Norris and Ortega in a metastudy of 49 studies concluded that "explicit modes of instructions are more effective than implicit types" (2001, p. 249). These terms usually refer to grammar teaching. Implicit instruction avoids grammar instruction, believing learners will pick it up naturally, while explicit instruction, including deductive and inductive methods, does specifically address grammar. Our belief is that if explicit grammar methods work better it is likely that this is the case in other areas of instruction. If in disfluency there exist a number of set expressions specific to native speakers, teachers should be able to transfer them to learners through explicit instruction.

[102] The term foreign language, rather than second language, is preferred when teaching a language in a formal context, however to provide consistency the term second language will continue to be used, whenever possible.

There is little conclusive literature on the success of teaching gestures to L2 learners (Gullberg, 2006, p. 110). But casual personal observations in the classroom indicate that after awareness raising and explicit instruction as to the use of both gestures and emblems, students do make a conscious effort to use them more.

Ellis identified ten principles to serve as the "basis for language teacher education" (2005, p. 292). The first one was to ensure that instruction results in learners' rule-based competence and knowledge of a range of formulaic expressions. Learners begin by internalizing set expressions and generally by focusing on the acquisition of vocabulary. Eventually they are able to analyse the structure of these set expressions and focus on the form as well. It seems that learners structure the acquisition process by dividing the new input into chunks, absorbing the easier ones first. It has been suggested that to ease the cognitive load on the learners, teachers ought to use both gestures and speech, and encourage students to use both as well (Nicoladis, 2007, p. 452). Most teachers do use both iconic and pragmatic gestures – scholars refer to this usage as "Teachers' talk" (Gullberg, 1999, p. 112) – just as many speakers would when talking to a foreigner or when using an L2. However, in the Spanish L2 classroom, HK students do not usually gesture, especially in lower proficiency levels. We believe gesturing needs to be encouraged, not only for students' benefit but also for the benefit of the teachers, who might be able to recognise conflicts by observing the gesture (Gullberg, 1999, p. 115). In addition, if it is believed that gesture and speech are one unit they ought to be taught together (Gullberg, 2006, p. 108).

Regarding disfluencies, Rodríguez and Torres (2006, p. 336) propose that understanding disfluency will lead to a better design of spoken dialogue interface. The practical applications of this in the language classroom might be to improve students' communication, based on a raised awareness of cultural idiosyncrasies that might exist in disfluency. Strategies in disfluencies do include formulaic expressions, such as 'I mean', 'or rather' (Clark, 2002, p. 12). Jefferson pointed out their importance when she said that these "non-lexical fillers", often seen as failures of the foreign language student, are "instances of the vast repertoire of conventional, standardized, abstract mechanisms which people use to interact not only coherently and competently, but with some degree of finesse" (1974, pp. 192-193). As discussed above, it is likely that elongations and fillers might be conventionalised mechanisms preferred by some Spanish and English speakers respectively, and ought to be taught as such in the classroom.

In the case of elongations we are aware that linguistics differences exist throughout Spanish speakers in the Iberian Peninsula (García Mouton, 1994) and it is quite likely that the use of the elongation might be particular to southerners from Málaga. This is just a hypothesis that ought to be considered. However, we feel it is still relevant in our case as, if teachers of second languages are to have an accent, this is more likely to be that of their place of origin. In our case, it would be an accent from Málaga, which is likely to be picked up by students. In addition the institution Hong Kong students attend has exchange programmes with universities in Spain, both in the South. Therefore we feel it is relevant to highlight specific characteristics of the Spanish Peninsula southern accent. Attention is drawn upon elongation and syllable separation differences between English and Spanish when instructing HK students in Spanish, from a beginners level (A1 in the Common European Framework of Reference for Languages). Mixed results are observed throughout the levels. What can be confirmed is that there is an improvement in students who have been made aware of the need to link syllables in Spanish (personal classroom observations). We believe that if attention is brought to the use of elongations as a speech trait (maybe also as a disfluency strategy) their use will increase.

Fillers on the other hand seem to be less frequently used by Spanish speakers than HK students. However as there seem to be other variables affecting it, possibly cognitive load and intercommunication aspects, we would recommend further studies before they are included in the curriculum.

As the results show (see Table 38), during disfluencies in the L2, HK students tend to use more pragmatic than iconic gestures, which might not aid the communicative act as much as if they were iconic, 'showing' the addressee what they mean if they cannot explain it with speech. We suggest that increasing HK students' awareness of the power of gestures, might increase their communicative success in these instances.

8 RECOMMENDATIONS AND FUTURE RESEARCH

In a study of this nature the results often raise more questions than they answer. Throughout the discussion a number of ideas for future research have been noted, both in disfluency and in the field of gestures, as well as suggestions to improve the methodology. The key ones are summarised below.

8.1 Recommendations

This study has been conducted in a thorough, accurate and factual manner. Every step that could be duplicated was repeated in a slightly different format to ensure reliability: recordings were done with two pieces of equipment, transcriptions were done and then checked using two different software programmes, and calculations were constantly crosschecked. However, as it is the work of one individual it is possible that errors have crept in. It is recommended that studies of this nature are always carried out by two or more researchers. We believe it is important that all data is checked by more than one individual as a number of decisions are of a subjective matter and would benefit from a consensus. In this study, whenever there was ambiguity about a point in the data this was not included in the analysis. A team of researchers would have allowed ambiguous data to be kept, if a consensus had been reached. It is also suggested that when using a confederate (the addressee in this case) this person has the same type of relationship with all participants. In this study the confederate was also the researcher, who was well known to HK students but a stranger to Spanish natives. Both groups knew the confederate to be Spanish native. A more objective approach would be to use L1 natives as confederates in both cases, if possible a stranger to the speaker. Bavelas and Gerwing (2007, p. 286) recommend using a different confederate for each speaker to ensure the confederate is not giving any signals of having heard the content before.

It has been mentioned in the discussion on types of gestures, that there was a prevalence of iconic gestures during the L1 aural retellings in both groups and also in the L2 that could have been related to the use of low frequency lexicon. Further studies should take this observation into account and ensure that the vocabulary needed in the retellings is consistent in terms of its frequency.

There was an expectation that we would find a transitional state between the L1 and the L2, interlanguage, where a systematic use of speech and gesture was being used, reflecting the Thinking for Speaking of not quite the L1 but not yet the L2. This interlanguage would be idiosyncratic to each learner as it changes with the learner's position in the acquisition continuum. The results indicate that there are differences in both gesture and disfluency that could be related to the level of proficiency. Therefore we suggest that future studies involving an L2 carefully state the proficiency level of participants, and strive for consistent proficiency levels. This ought to be an important variable to be considered in data comparison. We believe that it was a significant variable in our data that might be distorting many of the results (as mentioned during the discussion). We were not able to find participants with a higher proficiency level in Hong Kong and our access to Spanish natives was limited. Discussions are already in progress to carry out similar studies with colleagues in the US and the Netherlands, which will also give us access to speakers with similar proficiency levels as well as speakers of other languages. One of the issues to bear in mind is that HK students were bilingual (English-Cantonese); it is recommended that studies of this type are carried out with monolinguals, to avoid possible interferences from other languages (as we believe might have been the case in this study).

A key recommendation, which has been mentioned a number of times, is the need to ensure consistent rates as well as description of events that will make the data comparable to other studies. Whenever possible it is suggested that rates are calculated using words and phonic group (and clauses if considered relevant) to make results easier to share. [103]

For future research into conversational analysis we strongly suggest that the speech is analysed together with the gesture. Within our study, in the analysis of disfluency it was found that often the speech transcript was not enough to judge whether a disfluency, in particular pauses, could be an indication of a conflict. In cases like that it was necessary to look at the gesture for additional information that left no doubt as to the fact that there was a conflict, often also providing an insight into the type of conflict it might be, conceptual, semantical or articulatory. The following examples illustrate this point:

[103] All our data can be easily recalculated per word or length if necessary.

Example 1 from L1 video CS 2HK.

Line no.	Time start (seconds)	Time end (seconds)	Duration (seconds)	Transcript
33	36.964	39.612	2.648	um the cat was-
34	39.612	39.948	0.336	/
35	39.948	41.804	1.856	was um thrown out

In the above example we encountered an interruption in the speech, after the word 'was' had been completed. The word is repeated and is followed by a filler, and then speech continues. The speaker is describing an action which is the consequence of a previously described situation (the building does not allow cats or dogs), therefore we can assume that the concept has already been formulated (being thrown out would have been generated as an associated concept to the concept of the 'ban on cats and dogs'), therefore this is not a conceptual issue. That is as much as we can guess from the transcript. However if we add the gesture transcript we gain invaluable information:

Line no.	Time start (seconds)	Time end (seconds)	Duration (seconds)	Transcript	Gesture
33	36.964	39.612	2.648	um the cat was-	
34	39.612	39.948	0.336	[/	A whole gestural phrase is carried out during the pause, both hands are lifted from lap to chest height, palms facing, they move quickly from left to right, hands in a loose open palm.
35	39.948	41.804	1.856	was] um thrown out	

Most of the gesture, which mimics the action of someone throwing something out, is carried out during the pause, with the speech resuming with the post-stroke hold. It is clear that the conceptualization is not the issue in this instance, but the speaker might be having problems formulating it into words, but not into gestures.

Example 2 from L1 video CS 2HK.

Line no.	Time start (seconds)	Time end (seconds)	Duration (seconds)	Transcript
50	55.156	56.1	0.944	relaxing-
51	56.1	56.628	0.528	/
52	56.628	58.172	1.544	relaxing and swinging in

In this example the speaker is describing how Tweety was relaxing in her cage, but repeats 'relaxing' without any obvious reason. Again, it is the gesture that gives us the clue:

Line no.	Time start (seconds)	Time end (seconds)	Duration (seconds)	Transcript	Gesture
50	55.156	56.1	0.944	relaxing-	
51	56.1	56.628	0.528	/	Before the gestural phrase starts the right hand is holding the left by the wrist, both resting on the lap, the left hand palm is facing up with thumb and forefinger touching,
52	56.628	58.172	1.544	[relaxing and swinging] in	with 'relaxing' the palm opens up and slides forward as if to emphasize the word (giving an idea of affirmation, as if to indicate that 'relaxing' was indeed the appropriate word for this occasion) with 'and swinging' the stroke continues or merges into a new one, the fingers curl and the forefinger and annular swing slightly with the word 'swinging', after that the stroke ends very quickly going back to the resting posture at the starts of the phrase.

Interpreting speech and gesture together we believe that the speaker had doubts as to the appropriateness of the word 'relaxing' in this context and after considering it, decided that it was adequate.

As can be seen from the examples above the gesture analysis provides invaluable data that should be recorded alongside speech transcription. Most multimodal studies are aware of the need to incorporate all the signals and symbols used in the communication but within discourse analysis multimodal approaches are not very common. This is even truer of Spanish L2 studies where the emphasis is still on the speech production rather than the speech-gesture production. In particular the official Spanish proficiency examinations (DELE)[104] warn examiners against accounting for students' nonverbal communication (which includes gestures) when assessing their proficiency level. We believe that as gesture is an inherent part of the speech act, its use should also be taking into account in language proficiency assessments.

8.2 Future Research

Many of our conclusions are of a speculative nature and would need to be supported by additional studies, in particular by larger samples. In doing this the key will be to ensure that the rates used to calculate and compare the data are clear and kept consistent, and that the classification of types of disfluencies and gestures can be compared with those used in this study. Throughout the discussion a number of future research topics have been highlighted, these are summarised below:

Although the data on all pauses had been transcribed, with information of each pause by disfluency and resumption and length, these were not included in the analysis. This study focused on the analysis of pauses occurring with a disfluency. The main reason was an attempt at limiting the scope of the research and also in the knowledge that recent thorough studies have been carried out on pauses (Kendal, 2013). It is suggested that further analysis of the data, including pure pauses, be carried out. In particular, differences

[104] From instructor's advice received during a course for DELE examiners taken during May 2012. Instituto Cervantes, Beijing.

in pause length by language, group and type of input should be addressed. We believe that if the longer pauses observed in the L2 are related to the increase in cognitive load (as proposed by Heldner and Edlund, 2000), longer pauses will also be observed in retellings considered more challenging by participants. In addition, it is suggested that data from other participants (recorded but not included in our analysis) should also be analysed for pauses associated to disfluencies to verify whether the pause lengths in our limited sample are representative averages of each group.

This study did not measure the length of the elongation itself. It is suggested that future studies compare elongation lengths in English and Spanish to see whether there are any significant differences: we expect that elongations in Spanish will prove to be longer. This would confirm their role in Spanish natives as pause replacements.

Filler results suggest that Spanish speakers might have less of a tendency to use 'uh' and 'um' fillers than English speakers, preferring other disfluencies or even other fillers such as 'entonces', 'pues', 'este'. This is not to say that English speakers do not use equivalent fillers (Clark, 2002, p. 12); we identified 'and then' as a common alternative to fillers in some speakers (the equivalent of the Spanish 'entonces'). Due to the high degree of subjectivity in identifying these words as fillers – they could be just part of the narrative functioning as connectors – it was decided not to include them in the analysis. This is not to minimize their importance, and we suggest fillers is an area requiring further research, the literature on 'uh', 'um' fillers used in Spanish is scant and even more so on the use of other fillers and whether their use mirrors their suggested use in English.

It has been suggested that the filler is used to elicit help from the addressee (Bortfeld et al., 2001); if iconic gestures are also used to elicit help from the addressee (Gullberg, 2014) then there should be more gestures with fillers than with other disfluencies. Further research could be carried out to confirm this hypothesis.

There are a number of studies covering cut-offs, but they tend to look at specific sub-sets of the data, a meta-study is probably required to take stock of all the findings. We suggest further work in this field could also be linked to cognitive load and lexical difficulties. A specific topic could be analysis of whether the type of conflict might be reflected in preferences for a particular cut-off and gesture pattern. Our believe is that conceptual conflicts might be associated with extreme reactions in terms of cut-offs: either the word is finished and there is a long pause (indicating a complex cognitive conflict) or the interruption is mid-word and there is no pause (indicating a sudden thought change), in both cases the gesture might be put on hold. Cut-offs within word followed by a pause might be related to phonetic or articulatory issues and then the gesture might continue or start just before the disfluency.

There is evidence to suggest that when using the L1, speakers prefer to finish the word before interrupting speech and starting the repair, supporting the Delayed Interruption for Planning hypothesis (DIP). This seems to be the case as well in the L2 with higher levels of proficiency (Spanish natives speaking the L2). Speakers do not interrupt themselves immediately on detecting trouble, but seem to weigh fluidity versus correction before stopping, preferring fluidity. On the other hand Levelt (1989) reports that erroneous words are three times more likely to be interrupted than just inappropriate words. As our data did not analyse the type of error we cannot confirm this observation, but it would suggest that speakers consider appropriateness errors less important to maintain accuracy of content than fluidity - otherwise the balance of erroneous words to inappropriate ones ought to be the same (Levelt, 1989; Seyfeddinipur, 2006). The data used in this study could be further analysed to identify the sources of errors and verify Levelt's results.

The observation of gestures leads to various questions regarding their speed and possible relation to the type of gesture. One possibility is that pragmatic gestures might be faster, iconic ones taking longer to be performed, and this speed is increased in lower proficiency speakers. We observed that in HK students in the L2 the gestures were less well defined than in other groups. We propose to analyse the data further to establish whether there is a relationship between the sharpness of the gesture and the type of input. It would also be interesting to determine whether there are differences by type of gesture and speed. It has also been suggested that Spanish natives in the L1 perform more pragmatic gestures, probably to keep the rhythm (as they do not seem to be used as lexical retrieval gestures). An in-depth analysis of the description of the gesture, including the various phases and their timings, might shed some light on these issues.

The bulk of the analysis did not include the data analysed from the L1 retellings from the written and aural inputs. Throughout the analysis it has been noted that data from the written and aural retellings might be different from the video L1 retellings. Most participants found the aural retellings the hardest, followed by the video in the case of Spanish natives and the written text, in the case of HK students. We believe the processing of these different types of input, each resulting in a different cognitive load, plays a key role in the subsequent speech-processing activated in the retellings. Our hypothesis, supporting the work of Nicoladis et al. (2007, p. 447), is that differences in disfluency and gesture rates and types might be dependent on the type of input and cognitive load. The type of gesture might vary depending on the type of input given (audio, video, written).

Data from the debriefing interviews was also recorded and casual observations suggest that there are even bigger differences in gesture use during the debriefings (the data for these has not been included in this study, as it would have added too many variables). These observations will be verified in a future study. The data from the debriefings probably has higher gesture rates than data from the retellings and gestures seem to have more iconicity (a superficial analysis indicated there were fewer pragmatic gestures). The gestures seem to be sharper, better formed than the ones used in the retellings, this might indicate a stronger link gesture-speech-thought. In the retellings the participants had to access their short term memories but in the interviews they were also being requested to think, to analyse and give a plausible explanations as to what they did and how they did it. Our belief is that the retellings necessitated accessing short term memory and little else, the stories had already been structured in the input and the speaker was not required to use higher skills than comprehension and memorization (in some cases there was a clear attempt to summarize and rephrase). However, in the interviews, the speakers required the use a much more sophisticated cognitive process in order to answer the questions (an introspective view of the exercise they had just been through), thus needing to active short term and long term memory (they were asked why they thought they had found one retelling harder than others and their answers often led them to their school practices), as well as higher level processing to make logical connections and inferences. The increase in the cognitive load, in thinking, should carry with it an increase in gesture, if the Growth Point theory (McNeill & Duncan, 2000) is correct, and a higher number of disfluencies if these are marking though units.

As well as the data from the transcripts of the debriefings, the content of these should also be analysed to try and understand the thinking processes of the participants in each task. From the original 28 participants, Spanish native speakers all except one (out of 14) agreed that the easier text to retell was the written. The written story was simpler, with frequently used words and a clear beginning, middle and end, which might have been a factor in participants' evaluation of difficulty (the storyline of the other inputs was more complex). In contrast all except two Hong Kong English speakers (12 in total) thought the video the easiest. This difference is in itself interesting; it was speculated whether it might be related to the educational system but both Spanish and Hong Kong educational systems rely mostly on written texts. Participants were all in their early 20s and all seemed to be comfortable with digital technology; therefore it is unlikely that the preference would be related to an increased use of images on screens as a source of entertainment. Further work on this issue might clarify the differences. It is suggested that the content of the debriefings ought to be analysed and correlated to the results from the *VARK*[105] questionnaire some participants took, on their learning preferences.

Neurological studies of how speech is processed with gestures show that the two events may be integrated, triggering activity in the same parts of the brain (Kelly, Kravitz & Hopkins, 2004, p. 1). On the other hand Bergen (2012, Ch. 7) defends a link between gestures and visual simulation in speech processing models. It is known that cognitive abilities and styles differ; some people report solving cognitive tasks through vision while others use language (motor and auditory modes). The neural circuits corresponding to visual simulation and the words used to verbalize the scene are capable of being active simultaneously. If one text is easier than the others the question is why and whether this is related to how the story is remembered (via images, concepts or seeing/hearing words), and whether this is reflected in the gestures. Many participants claim to remember the exact words they heard in the aural input (and it is true that some use them): further

[105] An online questionnaire used to assess learning preferences (visual, aural, reading/writing and kinesthetic).

research could calculate how many words are exactly the same and what gestures were used with those words. In our study we could not control what cognitive abilities our participants were using, but through providing them with different types of inputs, we tried to ensure at least partial use of some of them. As the results show differences in the use of gestures by type of input, we believe that there might be a link between cognitive styles and gesturing. This is a hypothesis that requires further research with specialized neurological resources, such as MRI scanners.

Despite recognising that nonverbal elements are essential in the interaction, very little is explicitly taught about the use of hand gestures in the language classroom. Speakers searching for a word display a thinking face and withdraw their gaze, these could be signals to the listener, or just cognitive unloading (Goodwin & Goodwin, 1986, pp. 57-58). However this might not be so obvious in all cultures. In a study of nonverbal communication (gaze, head turns, intonation and gestures) elements used by HK students of Spanish, we (López-Ozieblo, 2012) found that in turn management (where these strategies are frequently used by western speakers) they were often missing. If in a normal dialogue there is no eye contact then withdrawing one's gaze would hardly be read as a sign of word searching. Nonverbal communication is different between HK students and Spanish natives. Despite a colonial history, many HK students' culture is based on Confucian principles and the use of elements like gaze, proximity or gesture is not the same as in western cultures. In particular gesture is likely to be more restrained, in modesty and partly an attempt at self-effacing (see López-Ozieblo, 2013a). Further research on the success of gesture instruction, and best methodologies to implement it would be very useful in the foreign language classroom in Sino-speaking contexts.

9 CONCLUSIONS

We consider that this study is a small progressive step in the field of gesture and disfluency studies. It provides a summary of key findings from both fields and it brings these together under the umbrella of second language acquisition. Our findings open up new directions for future research which we hope to follow ourselves.

9.1 General Objective

The main objective of this work was to confirm or reject the hypothesis that there is a relationship between speech and gesture in disfluency. The results seem to indicate that there is a relationship, but as there are a number of gestures produced without disfluency, we conclude that the relationship between gesture and speech is not unique to the disfluency. The function of the gesture is more than just an aid to lexical retrieval during disfluency (as gestures are produced without disfluencies). The gesture speech relationship observable during disfluent speech, might be dependent on the cognitive load. The higher cognitive load imposed by the use of the L2 seems to be resulting in an increase of gestures co-occurring with disfluencies. In lower proficiency speakers of the L2 there are more instances of disfluencies co-occurring with gestures than in the L1. Therefore we would like to suggest that in lower proficiency levels gesture might be used more often for lexical retrieval purposes than in the L1.

9.2 Secondary Objectives

One of the secondary objectives of this study was to compare the use of gesture and speech during disfluencies between native Spanish speakers and HK students speaking Spanish. The goal was to adapt and improve the curriculum for the program of Spanish as a foreign language currently taught at the Hong Kong Polytechnic University (HKPU). There was one key finding that will be included in the oral curriculum. It was found that elongations might not be always disfluencies, at least in the native Spanish participants, but might be instead a characteristic of Spanish speech related to the linking of syllables (Dauer, 1983). Spanish natives and some HK students who had spent time in Spain, use elongations often and do not follow them with pauses. Both groups increase the use of elongations in the L2, which is to be expected, as they encountered more speech conflicts in their L2. As elongations by Spanish natives were mostly used with continuing utterances it opens the question as to their status as a disfluency. If they signal a conflict it is obviously one of a small nature as it does not significantly disrupt the speech. The gesture analysis confirms these results. The gesture is seldom put on hold with elongations, although more cases are seen in lower proficiency scenarios. We would like to suggest that perhaps the description of elongations as a disfluency in Spanish natives needs to be revisited. In terms of the HKPU curriculum we believe elongations might be a speech trait that needs to be included in the oral syllabus. HKPU students participate in programmes that are more likely to put them in touch with speakers from Southern Spain, where the Spanish participants were from, therefore it is considered relevant for them to be aware of and know how to use elongations as Spanish natives do. The programme has already included explicit instruction on the use of elongations in Spanish.

Other disfluency results provided interesting insights, but no practical actions that could be included in the curriculum. It was found that cut-offs were used similarly by both groups, by Spanish natives using the L1 and L2 and HK students using the L1. Their use supports the Delayed Interruption for Planning hypothesis (DIP) (Seyfeddinipur, 2006). The DIP hypothesis states that speakers interrupt speech once processing has been completed, thus speakers who are more fluent are more likely to interrupt themselves mid-word and have shorter pauses, if any, between the disfluency and the repair. Fillers, on the other hand, seem to be used differently in Spanish and English. Speakers using English might have a tendency to use more 'um' fillers than speakers using Spanish. Fillers might also be indications of different types of conflicts, 'um' might signal a cognitive conflict; while 'uh' might indicate a lesser lexical conflict, and should perhaps be considered a word meaning 'expect a delay' (as suggested by Clark and Fox Tree, 2002). Thus explaining why the gesture is less likely to be put on hold during the filler. Therefore we reject the hypothesis proposed in the *Objectives* (Chapter 3):

H_2: $r \neq 0$: there is no significant difference in the disfluencies between the two languages as spoken by the same speaker.

And so conclude that there are some differences in the disfluencies between the two languages, as spoken by the same speaker, although this will vary by proficiency level. The results of the gesture analysis lead us to confirm the third hypothesis:

H_3: $r \neq 0$: there is an inverse relationship between the level of proficiency and the number of gestures made with disfluency.

During L2 retellings both Spanish natives and HK students performed more gestures co-occurring with disfluencies than in the L1, supporting the results of Graziano and Gullberg (2013). When the gesture was analysed further to ascertain its synchronicity with the disfluency it was found that the fourth hypothesis:

H_4: $r \neq 0$: the disfluency is synchronous with the gesture, with both speech and gesture put on hold during disfluency,

had to be rejected, for although a hold in the gesture at the point of the disfluency was observed in 33% of all cases, in 55% of all gestures occurring with the disfluencies, the gestured just continued. This does not indicate that there is no synchronicity, as it is likely that many of the disfluencies recorded (elongations and maybe 'uh' fillers) were not such. It is also possible that the gesture was synchronised not in terms of stopping but starting. The start time might precede or follow the speech. Butterworth and Hadar (1989) suggested that in certain types of disfluencies the gesture onset might precede the disfluency, an action we did not analyse (although it was transcribed).

There were few differences in the type of gesture by language or group (except in HK students using the L2), however we observed considerable differences by participant, not only in type but in number of gestures overall. We believe that gesture and speech form a unit (although we were not able to provide conclusive evidence of this in the study) and as such, support Gullberg (2006) in her call to include gestures in the L2 curriculum. Considering the benefits of using iconic gestures in successful communication, we have already included explicit gesture instruction in the Spanish programme at HKPU. This takes the form of specific exercises relating gesture to content, action verbs in particular, such as mimicking and role playing, as well as inductive instruction on the use of emblematic colloquial speech-gesture expressions (see Appendix 10). In addition, to facilitate implicit acquisition of native speech traits, a weekly one-hour oral session with Hispanic natives has been organised for current students of Spanish at the HKPU. Finally, all students at an A2 proficiency level are assigned a 'Skype-mate' from Málaga University, a native Spanish speaker studying English Studies. They are required[106] to *Skype* with each other, ensuring access to verbal and non-verbal input from natives.

We had also hoped to provide further insights into the speech-gesture unity. We believe our data provides some evidence of a relation between gesture and speech, but not enough to conclude that they form a unit, a Growth Point (McNeill & Duncan, 2000). Evidence exists to support different gesture-speech models and one obvious question is whether it could be possible that cognitive processes allow individuals to utilize more than one model depending on the context. An assumption that can be made is that, whatever the speech model, a monitoring system is interacting with the speech processing mechanism, identifying and correcting errors. As mentioned in the *Literature Review* (Chapter 4), we believe that there is an internal monitoring system affecting both gesture and speech that leads to conflict solving. This mechanism is probably sharing resources with speech planning, which allows for completion of reparandum words at the same time as repairs are being planned. The results of the study suggest that there is a significant speech-gesture relationship with some disfluencies, such as cut-offs, indicating that there the mechanism solving the conflict is dealing with both the speech and the gesture at the same time. However, in lower proficiency levels it seems that the gesture is often used as a lexical retrieval aid. It seems to be used alongside the speech (if with it at all), either as a compensatory strategy (explaining an unknown term with gestures), a demand for help from the addressee, or a self-priming gesture used to recall a known word. We speculate that there might be two speech systems: one activated with language fluidity that forms a unit with gesture, and another, used with low proficiency languages, where gesture performs a different function (that of priming oneself, or filling in for an unknown word), so it does not form a unit with speech. This second system would be akin to that of infants younger than 3~4 years old. At this age the child seems to use speech and gesture

[106] Previously students participated in the *Skype* exercise on voluntary bases.

differently than fluent speakers; indeed, speech and gesture seem to be processed by two different mechanisms, each adding a different piece of information (McNeill, 2012, p. 173).

One insight this study has provided is the likelihood of speech being processed in a parallel system, rather than through a series of steps. The fact that speakers do not seem to have the full utterance prepared when articulating it contradicts the serial speech processing mechanism hypothesis (Levelt, 1989). If individuals can choose visual or language modes (or others we might not yet be aware of) to solve cognitive tasks, it implies that ideas might come into being as language, might be language itself (rather than a series of images). During the debriefing with our participants we found that most of them, 90%, reported thinking 'visually' but 10% said to be thinking in 'words'. A serial processing framework would not fit a word based mode of thinking as speech and gesture seem to be generated continuously and not in chunks. When an idea emerges it activates likely elements that might be involved in that idea (lexicon, articulation, syntax), which in turn activate others. The most likely scenario is selected until a new thought sets off another activation reaction. Utterances are constantly being monitored and changed, sometimes causing significant conflicts (marked by fillers like 'um' and maybe pauses) and at other times not even being noticed.

A final objective was to lay the bases for future disfluency-gesture research studies. Our study has generated as many questions as it has answered and it is our intention to use the existing transcriptions for further analysis that might help to find some of the answers. In particular we would like to follow up with in-depth research of disfluencies and gestures in L1 retellings by type of input, as we believe there might be some significant differences that would be explained by cognitive differences. One key recommendation that has resulted from this study is the importance of ensuring comparable measuring rates and event categorisation when analysing either disfluencies or gestures.

10 RESUMEN

Este estudio tiene sus orígenes en la necesidad de entender las estrategias comunicativas de alumnos de español como lengua extranjera[107] en Hong Kong (HK) al encontrarse con conflictos al hablar. Estos conflictos pueden estar relacionados con el léxico, la sintaxis o la fonética, pero también pueden ser de índole cognitiva (al no recordar un acontecimiento o al no poder describirlo adecuadamente). Al enfrentarse a estos conflictos, el hablante suele interrumpirse, pausar, alargar ciertos sonidos,[108] cambiarlos o usar marcadores como 'eh', 'em', 'pues' o equivalentes, rompiendo la continuidad del habla y creando lo que llamaremos *disfluencias.*[109] Tras una disfluencia el hablante puede continuar su discurso o repararlo, repitiendo, añadiendo, sustituyendo o eliminando partes del discurso anterior a la disfluencia. Usaremos el término 'reparar', ya que corregir indica error, lo cual no es siempre el caso; a menudo, las disfluencias son tan solo estrategias usadas por el hablante para llamar la atención del oyente, al utilizar una pausa más larga de lo normal, o para ayudarlo a entender el mensaje.

Dentro del campo del análisis del discurso se está empezando a reconocer que el acto comunicativo va más allá del habla: la comunicación se efectúa gracias a una combinación de signos verbales y no verbales que incluyen la postura, la mirada, movimientos de las manos y la cabeza, etc. (Poyatos, 1993). En particular, el estudio de los gestos, movimientos de los brazos y las manos es un área en crecimiento dentro de los estudios de lingüística, con investigadores como Bavelas, Goldin-Meadow, Gullberg, Kendon, Kita o McNeill, quienes proponen la unidad del gesto y el habla. Tanto si efectivamente estos dos procesos están ligados, o, por el contrario, son tan solo uno (como proponen McNeill y Duncan, 2002), creemos necesario estudiar las disfluencias junto con los gestos que ocurren con ellas.

[107] Para facilitar la lectura se utilizará el sintagma *segunda lengua* (L2) al referirnos a la lengua extranjera.

[108] Se usa el término *sonido* ya que nos referimos al habla y no a la lengua.

[109] Del inglés *disfluencies*.

Durante las décadas de los 60 y 70 del siglo XX, el estudio de los errores en el discurso condujo a la propuesta de un modelo del proceso del habla que establecía una serie de pasos desde el origen de la idea en un módulo conceptualizador hasta su articulación (Fromkin, 1971; Levelt, 1989). Este modelo ha sido la base de otros para explicar el procesamiento del habla y el gesto (de Ruiter, 1998). Sin embargo, estudios neurológicos actuales indican que el proceso del habla seguramente ocurre en paralelo, donde una serie de nódulos neuronales conectados entre sí son activados y, a su vez, activan otros hasta que aquellos que están recibiendo los mayores impulsos activadores son los que resultan en el acto comunicativo (Banich y Compton, 2011; Dell y Oppenheim, 2012), incluyendo el gesto, según la teoría del Punto de Crecimiento de McNeill y Duncan (2002). Nuestra propuesta es que si el estudio de los errores llevó a los primeros modelos del proceso del habla, quizás un mejor conocimiento de las disfluencias y los gestos que ocurren con ellas puedan llevarnos a confirmar uno u otro de los modelos habla-gesto que se han propuesto hasta ahora.

Es muy posible que el gesto esté relacionado con el tipo de disfluencia, ya que parece que no todas las disfluencias son iguales en términos de cómo y por qué se generan. A su vez, la disfluencia puede que esté relacionada con el tipo de contenido del habla (Bortfeld et al., 2001) y la carga cognitiva (Heldner y Edlund, 2010). Dentro del campo de las disfluencias, los marcadores (Clark y Fox Tree, 2002) y las pausas (Kendall, 2013) son dos de las más estudiadas, aunque su función no está ni mucho menos confirmada. En el caso de los marcadores, su uso se ha relacionado con vacilaciones, turnos discursivos o correcciones, mientras que Clark y Fox Tree (2002, p. 107) sugieren que marcadores como ‘eh’ y ‘em’ son palabras cuyo significado es ‘estoy haciendo una pausa’. Dependiendo de cómo de larga vaya a ser la pausa el hablante elige un marcador u otro: ‘eh’ para pausas más cortas y ‘em’ para aquellas más largas, relacionadas con conflictos más serios (p. 95). Las interrupciones, por otro lado, se consideran acciones controladas (subconscientemente) que ocurren cuando el hablante se da cuenta de que existe una opción más apropiada. La hipótesis del Retraso de la Interrupción para facilitar la Planificación[110] (Seyfeddinipur, 2006; Seyfeddinipur y Kita, 2001; Seyfeddinipur, Kita e Indefrey, 2008) explica la producción de las

[110] En inglés: *Delayed Interruption for Planning*.

interrupciones (o truncamiento de sílabas): cuando el hablante reconoce el conflicto, interrumpe el habla; si puede reparar el discurso inmediatamente, puede que haya una pausa, pero esta será corta. Es posible que el hablante reconozca el conflicto pero no tenga una solución a mano, en cuyo caso seguirá hablando hasta que termine la palabra y solo continuará cuando haya encontrado la reparación, en estos casos la pausa será más larga. Otras disfluencias como los alargamientos, los cuales suelen estar acompañados por pausas cortas (Clark, 1996, p. 268), no han sido objeto de muchos estudios, entre otras cosas porque su categorización como disfluencia no es tan obvia. En cambio la reducción de vocales (cambios fonéticos de ciertos sonidos al ser interrumpidos) ha sido ampliamente estudiada, al menos en inglés, hasta el punto de que se reconoce que ciertas variaciones del inglés, como la que se habla en Hong Kong, no presentan tantas reducciones. En el caso de Hong Kong debido a la influencia del cantonés (Setter, 2012).

Existen diferentes tipos de gestos según su relación con el habla; en este estudio se han dividido en dos categorías: aquellos que se refieren al contenido del habla, gestos icónicos, y los que se usan para enfatizar o llevar el ritmo del habla pero no tienen relación con el contenido en sí, gestos pragmáticos. Dentro del grupo de gestos icónicos se han incluido aquellos que representan el contenido visualmente (gestos con iconicidad), o se refieren a una característica del contenido y usan una representación metafórica de esa característica (gestos con metaforicidad); también se han incluido gestos deícticos, usados para indicar. Los gestos pragmáticos incluyen los rítmicos (para enfatizar o llevar el ritmo) y los léxicos, asociados con dificultades en recordar palabras (McNeill, 1992, 2012).

En los pocos estudios existentes sobre disfluencias y gestos en segundas lenguas (Graziano y Gullberg, 2013, 2014) se ha observado que en la L2 los hablantes realizan más gestos durante disfluencias que en la lengua materna (o primera lengua, L1). Es posible que esta diferencia esté relacionada con un mayor uso de gestos icónicos para pedir ayuda al oyente (Gullberg, 2014), 'dibujando' la palabra necesaria con las manos, y específicamente en niveles de competencia más bajos en la L2; destacan los gestos deícticos que se usan anafóricamente (como referencia) junto con una preferencia a usar sustantivos y evitar verbos (Gullberg, 1998). Dentro de los estudios de adquisición de segundas lenguas, ni las disfluencias ni los gestos son temas muy tratados (salvo algunos estudios sobre los gestos del profesor Tellier, 2010). Dado que nuestro interés se centra en la adquisición del español por alumnos de HK, consideramos necesario establecer un corpus con contenidos de hablantes nativos de español, para poder compararlo con un corpus parecido de contenidos de alumnos de español hablando español. Para evaluar si las diferencias en la L2 estaban relacionadas con la lengua en sí o con el nivel de competencia fue también necesario crear un corpus con muestras del habla de los participantes usando inglés, la L1 (en el caso de alumnos de HK)[111] o la L2 (en el caso de los participantes nativos españoles). El objetivo principal del estudio ha sido identificar manifestaciones características de nativos españoles en la L1, en disfluencias y gestos, que se pudieran incluir en el programa de español como lengua extranjera de la institución de los alumnos de HK.

En un estudio preliminar se confirmó que los alumnos de HK usan disfluencias y gestos al hablar en español, pero la pregunta era si estos eran parecidos a los de los hablantes de español nativos. Para contestar esta pregunta se diseñó un segundo estudio con alumnos de español y nativos españoles en los que el input estaba controlado. Los participantes tenían que contar tres historias en su L1 después de haberlas leído, escuchado o visto (sin sonido) respectivamente. Aquellos con un nivel de competencia suficiente en la L2, también contaron la segunda parte del input visual (un vídeo) en la L2. Finalmente todos los participantes reflexionaron, en la L1, sobre el ejercicio y qué parte les había resultado más difícil. Las sesiones se grabaron en vídeo y tanto el habla como el gesto se transcribieron para luego analizarlos. Los participantes habían sido previamente informados, para obtener su consentimiento, de la necesidad de grabarles para estudiar su comunicación verbal y no verbal, pero no se les indicó que el centro de atención eran los gestos. De un total de 28 participantes, se eligieron cuatro nativos españoles y cuatro alumnos de HK. Dado que buscábamos una participación equilibrada en cuanto al género de los participantes, el factor limitante fue la variable sexo, ya que tan solo dos varones alumnos de HK aceptaron contar la historia en la L2. Cada participante relató tres historias en la L1 y una en la L2, por lo tanto el corpus total consistió en 32 narraciones, de las cuales 24 estaban en la L1 y 8 en la L2. La duración de las narraciones fue de una media de un minuto en la L1 y de dos minutos en la L2.

[111] Los participantes de Hong Kong son bilingües en cantonés e inglés, por lo que no es estrictamente correcto usar el término L1, aunque se usará en este estudio para facilitar la lectura.

El corpus se analizó para tabular los casos de disfluencia, de gesto y la reparación. Las disfluencias se categorizaron por tipo: interrupciones, marcadores y alargamientos (Clark, 1996),[112] así mismo se anotaron las pausas relacionadas con estas disfluencias. Los gestos se categorizaron como icónicos o pragmáticos (Bavelas, 1994; McNeill, 2000; Rauschner, Krauss y Chen, 1996) y se anotó la fase del gesto en el momento de la disfluencia (si el gesto comenzaba, se detenía o seguía, según la definición de fases gestuales de Kendon, 2004). La continuación o reparación del discurso, incluyendo repeticiones, eliminaciones, adiciones y sustituciones, se anotó según los estudios de Clark (1996). Los resultados de los análisis se compararon por participante, por tipo de input y por lengua, observándose diferencias significativas en los resultados dependiendo del tipo de input (escrito, oído o visual). Por lo tanto, se decidió que para eliminar la posible variable del input, un estudio comparativo entre la producción en la L1 y la L2 debía incluir tan solo narraciones basadas en el mismo tipo de input, con lo que se estudiaron con más profundidad las disfluencias y los gestos de las narraciones en la L1 y L2 basadas en los vídeos.

Otras variables tenidas en cuenta fueron las duraciones de las narraciones, el número de palabras y el número de grupos fónicos (grupos de sonidos marcados por pausas de más de 200 milisegundos), ya que estas variables pueden estar condicionadas por la lengua (inglés o español) pero también por el nivel de competencia del hablante. Era por tanto necesario neutralizarlas, normalizando los números de disfluencias y gestos dividiéndolas entre el número de palabras o grupos fónicos y la duración. Se decidió usar el número de grupos fónicos en vez de palabras, a raíz del trabajo de Colletta, Pellenq y Cefidekhanie (2014), quienes sugieren que el número de grupos fónicos es un mejor denominador, ya que está menos relacionado con el nivel de competencia en la lengua. Aunque la mayor parte de los cálculos se realizó usando el número de grupos fónicos como denominador, también se calculó el número de eventos (disfluencias y gestos) según las otras variables. Los resultados indican que según el denominador que se use, las conclusiones pueden ser muy diferentes, con lo que la elección de los parámetros de

[112] Las reducciones no se marcaron ya que los hongkoneses son menos dados a usarlas debido a la influencia del cantonés, como se ha mencionado anteriormente.

medición es importante, explicando, en parte, el número de resultados contradictorios que se encuentran en estudios de gestos.

Para lograr el objetivo final –implementar cambios en el programa de español de la institución– se propusieron una serie de objetivos; el principal, el de confirmar una relación entre la disfluencia y el gesto. Los objetivos secundarios eran confirmar (o no) la existencia de diferencias en las disfluencias en la L1 y la L2 de un mismo hablante; confirmar (o no) una relación inversa entre el nivel de competencia en la L2 y el número de gestos con disfluencias; y confirmar (o no) la sincronización de la disfluencia con el gesto. Se comprobó si los resultados de los cálculos, realizados con *Excel*, eran significativos con un valor crítico alfa de confianza de 0,05 (95 %) y 0,1 (90 %) usando el programa *Statistics Calculator*. Otros programas usados fueron *Express-Scribe* y *Video-Pad Editor,* para realizar las transcripciones, y *Praat,* para anotar la duración de las pausas y los alargamientos, ya que presenta la voz en forma de espectrogramas y diagramas de ondas, intensidad y tono. Además de la información básica sobre la duración, número de palabras y grupos fónicos, los cálculos realizados incluyeron:

- Número de pausas con y sin disfluencias y su duración antes y después de la disfluencia (por tipo de disfluencia).
- Número de disfluencias en cada narración de cada tipo con y sin pausas.
- Tipo de disfluencia en cada narración y si ocurrió con o sin gesto.
- Tipo de reparación por tipo de disfluencia con gestos.
- Fase del gesto (si el gesto continúa, empieza o se detiene) al comienzo de la disfluencia.
- Tipo de gesto (pragmático o icónico) con cada disfluencia.

Los resultados indican que existen diferencias en el habla y el gesto dependiendo del tipo de input, la lengua y el nivel de competencia en la misma. El nivel de español de los alumnos de HK era más bajo que el de los nativos españoles en inglés, por lo que una comparación directa de las dos L2 (español como L2 e inglés como L2) no es apropiada, ya que se mezclarían las variables del nivel de competencia con el de las posibles diferencias relacionadas con la lengua. Esto se ha tomado en cuenta en el análisis y se ha señalado en la discusión. Si aceptamos que el gesto y el habla están directamente relacionados con la carga cognitiva (Nicoladis, Pika, Yin y Marentette, 2007, p. 447), y que los diferentes inputs conllevan diferentes cargas cognitivas, claramente tampoco sería apropiado considerar como iguales todas las narraciones.

El número de grupos fónicos por segundo es mayor en los alumnos de HK que en los nativos españoles, seguramente por la tendencia de los españoles a enlazar las sílabas, produciendo grupo fónicos más largos y menos pausas (Dauer, 1983). En la L1 el habla de los alumnos de HK es mucho más fluida, más rápida y con voces más largas que en la L2. En los nativos españoles estas diferencias entre la L1 y la L2 no son tan marcadas, seguramente porque su nivel de competencia en la L2 era más alto (B2), mientras que en los alumnos de HK era un A2 inicial.[113] La previsión era observar un mayor número de disfluencias[114] (en la L2 en relación a la L1, sin embargo este no es siempre el caso según el participante). Parece que en la L2 los estudiantes, en particular los alumnos de HK, están más preocupados con la corrección que con la fluidez de su discurso, lo cual redunda en un discurso más lento, por lo general. Estudios en disfluencias observan ratios de 0,04 ~ 0,06 disfluencias por palabra, tanto en inglés como en español (Bortfeld, Leon, Bloom, Schober y Brennan, 2001; Brown Quiros, Limon, Aspuerez, Church y Mahootian, 2014; Oviatt, 1995; Rodríguez y Torres, 2006). Los porcentajes calculados en este estudio son mucho más altos en las dos lenguas. No obstante, no es posible establecer una comparación directa, ya que la definición de disfluencias no es la misma (algunos estudios cuentan ciertas reparaciones como disfluencias) y desconocemos qué se ha incluido como palabra.[115] Bortfeld et al. (2001) postulan que el número y tipo de disfluencias depende del contenido del habla, mientras que Rühlemann, Bagoutdinov y O'Donnell (2011), en relación a las pausas, han encontrado diferencias en narraciones versus diálogos. Estas diferencias sugieren que la carga cognitiva de distintos tipos de contenido, no de distintas lenguas, afecta al proceso del habla, dando lugar a diferencias perceptibles en el número de disfluencias y gestos. Parece que a mayor carga cognitiva, mayor es el número de disfluencias y gestos, pero si la relación entre el habla y el gesto es causal (el gesto ocurre porque hay una disfluencia), entonces la

[113] Niveles según el Marco Común de Referencia Europeo (MCRE) de lenguas.

[114] Excluyendo pausas sin relación con otro tipo de disfluencia.

[115] En este estudio, siguiendo las pautas de Graziano y Gullberg (2014), se contaron como palabras cada interrupción y marcadores.

predicción es que deberíamos observar más gestos con disfluencias que con habla fluida; sin embargo, este no es el caso, al menos no en la L1 de nuestros participantes. En el corpus de la L1 de los nativos españoles se observaron más disfluencias sin gestos (62 % de todas las disfluencias) que con ellos; lo mismo se observó en el corpus de la L1 de los alumnos de HK (el 53 % de las disfluencias ocurren sin gestos). El gesto tampoco suele comenzar con la disfluencia (tan solo en un 13 % de todos los casos en la L1, incluyendo todas las narraciones), lo que nos lleva a concluir que no existe una relación causal entre la disfluencia y el gesto, al menos no en la L1.

En la L2 parece existir otro tipo de relación entre el gesto y la disfluencia, una relación más fuerte y positiva. En el corpus de la L2 de alumnos de HK, hasta un 65 % de las disfluencias ocurren con un gesto, mientras que en nativos españoles el porcentaje es de un 60 %. Estos resultados concuerdan con los de Graziano y Gullberg (2013), quienes también observaron un aumento de disfluencias con gestos en la L2. Aunque exista una relación es difícil de clasificar su naturaleza, dado que el gesto solo comienza con la disfluencia en el 6 % de los casos en el corpus de los nativos españoles, y tan solo en el 2 % de los casos en el corpus de los alumnos de HK. Una explicación a esta falta de sincronía viene dada por los estudios de Hadar y Butterworth (1997) y Schegloff (1984), quienes observaron que en las disfluencias el gesto suele anticiparse al habla, [116] lo cual coincide con nuestras observaciones, en las cuales el gesto se desarrolla durante la disfluencia, comenzando anteriormente. Nuestras conclusiones son que en la L2 existe una relación diferente entre la disfluencia y el gesto que en la L1; quizás una de sus funciones en la L2 sea compensatoria, cuando falla el habla. Claramente esta no es la única función del gesto en la L1 (donde no se usa el gesto en vez de la voz) lo cual sugiere que su función es más complicada, quizás aligerar la carga cognitiva (Goldin-Meadow, 2001).

Si la función del gesto en la L2 es compensatoria, deberíamos observar un mayor número de gestos icónicos en la L2 (en relación al número de gestos pragmáticos), tanto puramente icónicos como deícticos (como se sugiere en Gullberg, 2014 y 2003, respectivamente). Efectivamente, los resultados muestran que un 51 % de todos los gestos

[116] Aunque este estudio no haya evaluado los eventos donde el gesto se anticipa al habla, estos han sido transcritos y se usarán en un futuro estudio.

de la L2 en alumnos de HK son gestos icónicos, mientras que en los nativos españoles el porcentaje es de 58 %. En los alumnos de HK también se observa un alto número de gestos icónicos en la L1 (55 %), lo cual puede tener su explicación en la oyente, quien era nativa española, pudiendo haber ocasionado un incremento de gestos icónicos por parte de los alumnos de HK para asegurar su comprensión. Para evitar esta posible interferencia, una de nuestras recomendaciones para futuros estudios es asegurar que el oyente sea nativo en la L1 del hablante. Por otro lado, entendemos que la preferencia de nativos españoles a usar gestos pragmáticos en la L1 puede estar relacionada con un diferente uso de los gestos en español, donde la función del gesto es enfatizar y llevar el ritmo del habla, lo que puede no ser tan frecuente en inglés. Consideramos esta una interesante línea para futuras investigaciones.

Cuando analizamos tan solo los gestos que ocurren con disfluencias, encontramos que los gestos icónicos siguen siendo prevalentes en el corpus de la L1 de alumnos de HK, pero también en el de la L2 de españoles nativos. Esto es, parece que hay una tendencia a usar gestos icónicos en inglés y pragmáticos en español. Puede ser que, efectivamente, exista una relación entre la lengua y el tipo de gesto, pero en este caso pensamos que los alumnos de HK usan los gestos pragmáticos, no tanto para enfatizar o llevar el ritmo, sino como apoyos léxicos, para ayudarse a recuperar palabras perdidas. Las observaciones de los gestos indican que suelen ser gestos rápidos, normalmente relacionados con la recuperación de léxico, como puede ser el girar la mano hacia afuera, alejándola del cuerpo. Estos gestos son comunes cuando se conoce la palabra pero no se consigue acceder a ella, mientras que los gestos icónicos suelen representar palabras que no se conocen (Gullberg, 2014). Basándonos en estos resultados, no podemos concluir que el tipo de gesto difiera dependiendo de la lengua, aunque es bastante probable que haya una preferencia a usar gestos pragmáticos en español, para llevar el ritmo y enfatizar. También es probable que el gesto esté relacionado con la disfluencia y que su frecuencia, tipo y fase de producción (si comienza antes o continúa durante la disfluencia) dependa del tipo de disfluencia, al menos en la L2. Para comprobar esta posibilidad fue necesario analizar en detalle las disfluencias que tuvieron lugar con gestos. Se analizaron las interrupciones, alargamientos y marcadores incluyendo sus subcategorías y las pausas asociadas a todas.

Las interrupciones, basándonos en el trabajo de Levelt (1989), fueron divididas en aquellas que tienen lugar después del final de la palabra y las que interrumpen la palabra a la mitad. Las primeras suelen ser más comunes según Levelt, hecho que confirmamos con nuestros resultados del corpus en la L1: el 72 % de todas las interrupciones en el corpus de los nativos españoles ocurre al terminar la palabra, y el 78 % en el de alumnos de Hong Kong. Sin embargo, en la L2 se observó un aumento de estos casos al 91 % en el corpus de nativos españoles, y un descenso a tan solo el 35 % de todas las interrupciones en el corpus de los alumnos de español. Teniendo en cuenta la hipótesis del Retraso de la Interrupción para facilitar la Planificación (Seyfeddinipur, 2006; Seyfeddinipur y Kita, 2001; Seyfeddinipur, Kita e Indefrey, 2008), y dado que el dominio de la lengua es mayor en la L1, tanto a nivel de vocabulario como por la familiaridad con estructuras gramaticales y fonética, esperábamos observar menos pausas y más cortas, con interrupciones a media palabra, así como un mayor número de las mismas en la L1. Efectivamente, los resultados indican que en casi todo el corpus se cumplen las predicciones, excepto en el caso de alumnos de HK usando la L2, donde observamos más interrupciones a media palabra que en la L1. Estos resultados sugieren que los alumnos de HK usando la L2 están más preocupados por la corrección que por la fluidez, lo que les lleva a detener el habla en cuanto perciben un conflicto. Las interrupciones parecen indicar conflictos que necesitan ser reparados, ya que tras un 96 % de todas las interrupciones, el discurso es modificado; la reparación más frecuente es la repetición (39 % de todos los casos de interrupciones). Los gestos más observados con las interrupciones son los icónicos, excepto en alumnos de HK usando la L2, donde se observa una preferencia a usar gestos pragmáticos, la mayoría léxicos, como si el hablante necesitase el gesto para encontrar la palabra.

Las diferencias en el uso de las interrupciones entre nativos españoles y alumnos de HK parecen estar más relacionadas con los diferentes niveles de competencia en la lengua que en las lenguas en sí (español versus inglés). Una de las razones por las que se observó un mayor número de interrupciones a media palabra en la L2 de alumnos de HK, es que este grupo todavía tartamudea en la L2, lo cual no ocurre en niveles de competencia más altos. Sería interesante poder repetir el estudio con hablantes del mismo nivel de diferentes L2, pero también de diferentes niveles (A2, B1, B2, C1 y C2).

Otras disfluencias que se estudiaron en detalle fueron los marcadores, estos se dividieron en dos tipos: 'eh' y 'em' (o sus equivalentes: en inglés 'uh' y 'um', y otros casos como 'ah' y 'am', 'er' y alargamientos como 'e:m' o 'em:').[117] Se anotaron también otro tipo de marcadores como 'pues' o 'and then' en inglés, pero no se usaron en los cálculos ya que no todos se consideraron necesariamente disfluencias. Se observó que los alumnos de HK usan más marcadores que los nativos españoles en sus respectivas L1, lo cual nos sugiere que hay una preferencia a usar marcadores de este tipo en inglés, mientras que en español la preferencia es a usar otras disfluencias, como alargamientos. Sin embargo, es necesario señalar que los casos de marcadores en el corpus no son muchos, con lo que nuestros resultados estadísticos no son muy significativos al pertenecer a una muestra pequeña.

La función de los marcadores ha sido estudiada anteriormente con resultados bastante diversos: Corley, MacGregor y Donaldson (2007) proponen que el hablante usa los marcadores para facilitar la comprensión al oyente; Kjellmer (2003) sugiere que los marcadores son para el hablante, indicando nuevas unidades de pensamiento (de ideas); Clark y Fox Tree (2002) apuntan que el marcador indica un conflicto; según el tipo, el hablante usará 'eh' o 'em' (los conflictos pequeños, rápidos de solucionar van ligados a 'eh', mientras que aquellos más complicados se marcan con 'em'). En ambos casos puede haber pausas antes o después del marcador, pero estas serán más largas con 'em', al necesitarse más tiempo para solucionar el conflicto. Basándonos en estas teorías, las predicciones eran observar un mayor número de 'em' en la L2 y pausas más largas con 'em' que con 'eh'. Nuestros resultados confirman, en parte, las teorías de Clark y Fox Tree, ya que encontramos más 'em' en la L2 en los nativos españoles (pero no es el caso en el corpus de los alumnos de HK). Las pausas anteriores a 'eh' son a menudo más largas que con 'em', pero es posible que sea a causa de interferencias de otras disfluencias, en particular interrupciones, las cuales –tras la pausa– a menudo son seguidas por un marcador. Si el marcador fuese una indicación de una nueva idea, como dice Kejllmer, el gesto (que también se considera una indicación de contenido nuevo, McNeill, 2012) debería estar sincronizado con el 'uh' o 'um', comenzando a la vez, lo cual no es el caso; al contrario, lo más común es que el gesto continúe: en un 35 % de todos los casos de 'eh' (en la L1 y la L2)

[117] Los dos puntos ':' indican el alargamiento.

y en un 22 % de 'em'.[118] Sin embargo esta diferencia (35 % y 22 %) puede estar señalando un diferente uso de 'eh' y 'em', como proponen Clark y Fox Tree, quizás como palabras que indiquen al oyente que se está produciendo un descanso en el discurso. Quizás el marcador esté relacionado con el tipo de conflicto con el que se encuentra el hablante: si procede de dificultades en la memoria de trabajo, el conflicto será menor y se solucionará con mayor rapidez, en cuyo caso el hablante lo indicará con un 'eh'; o si es un conflicto relacionado con la memoria a largo plazo, marcado con un 'em'. Un conflicto relacionado con la memoria de trabajo puede ser el no recordar una palabra que se conoce, mientras que si la palabra no se conoce, o no se ha llegado a procesar la idea que se quiere transmitir, esta dificultad es más probable que esté relacionada con la memoria a largo plazo. Esperamos poder usar el resto del corpus recogido (incluyendo las grabaciones de las sesiones de reflexión) para seguir estudiando estas hipótesis.

En los nativos de español se observó un aumento del uso de marcadores en la L2, una posibilidad es que en español se usen menos marcadores que en inglés, y que nuestros participantes, consciente o subconscientemente, reconozcan este rasgo de la lengua, y lo pongan en práctica, habiendo tenido ocasión de usar la L2 en contextos informales con nativos de esa L2. También se observaron otras diferencias en la frecuencia, el tipo de marcadores y la duración de las pausas relacionadas con ellos, así como en los gestos. No obstante, se decidió que la confianza en los resultados no era suficiente como para usarlos como base para modificar el programa de español de la institución donde estudiaban los alumnos de HK.

Por otro lado, los resultados del análisis de los alargamientos sí han llevado a una revisión del programa, ya que se consideró que su uso por nativos españoles es característico del lenguaje. Rodríguez y Torres (2006) ya reconocían el uso de alargamientos como la disfluencia más común entre nativos españoles, y nuestros resultados confirman los suyos: el 51 % de todas las disfluencias en la L1 de nativos españoles son alargamientos, y el 54 % en la L2. Mientras que en los alumnos de HK esta proporción es tan solo del 24 % en la L1 y del 34 % en la L2, pensamos que este incremento está relacionado con el hecho de que algunos alumnos de HK han tenido contacto directo con nativos españoles en contextos

[118] Estos cálculos incluyen también marcadores sin gestos.

informales. Los nativos españoles tienden a alargar sonidos, uniendo sílabas dentro de una misma palabra o entre palabras (Dauer, 1983), con lo cual se observan menos pausas; en el corpus de ese grupo, un 80 % de los alargamientos suceden sin pausa y el discurso continúa sin reparación, sugiriendo que el alargamiento puede no ser una disfluencia sino sencillamente un rasgo del habla de estos hablantes, todos ellos de Sevilla. Dado que es más probable que el contacto con nativos españoles de los alumnos de HK sea con sevillanos o jiennenses (existen acuerdos de intercambio con esas universidades y la institución de los alumnos de HK), consideramos importante incluir los alargamientos como parte explícita de la formación.

Nuestros resultados indican que existen diferencias en el uso de alargamientos en inglés y español, que es posible que existan también en el uso de marcadores, pero no en interrupciones. Así mismo, el uso de los gestos parece indicar que no todas las disfluencias son iguales: el gesto se detiene con las interrupciones, pero continúa durante el alargamiento; también parecen indicar que no todos los marcadores son iguales y que quizás estos dependan del conflicto que el hablante está tratando de resolver.

Concluyendo:

Hipótesis	Observaciones
H_1 : $r \neq 0$: existe una relación entre el habla y el gesto durante disfluencias.	Parece que no existe una relación causal entre el gesto y la disfluencia en la L1, pero quizás sí en la L2 en hablantes de bajo nivel de competencia lingüística.
H_2: $r \neq 0$: no hay diferencias significativas en las disfluencias de un mismo hablante usando la L1 o la L2.	Existen diferencias en el uso de los alargamientos que parecen estar directamente relacionadas con el tipo de lengua usada por el hablante (inglés o español). Quizás existan diferencias en el uso de los marcadores.

Hipótesis	Observaciones
H_3: r ≠ 0: existe una relación inversa entre el nivel de competencia en la L2 y el número de gestos realizados durante disfluencias.	Se observan más gestos junto con disfluencias en la L2 que en la L1.
H_4: r ≠ 0: existe sincronía entre la disfluencia y el gesto, los dos se detienen durante la disfluencia.	La sincronía entre el habla y el gesto depende de la disfluencia. En general parece que el gesto continúa durante la disfluencia.

Consideramos que la diferencia en el uso de los alargamientos es suficientemente significativa como para incluir su formación explícita en el programa de español como lengua extranjera de la institución de los alumnos de HK. Proponemos seguir estudiando otras disfluencias, sobre todo en hablantes con diferentes niveles de competencia.

Nuestros resultados no son suficientemente concluyentes como para permitirnos confirmar el tipo de relación entre el habla y el gesto o su procesamiento. Parece que en la L1 no existe una relación causal entre el gesto y el habla en el momento de la disfluencia, ya que observamos disfluencias sin gestos y gestos sin disfluencias, lo que sugiere que la función del gesto no es tan solo la de facilitar la palabra. No obstante, en hablantes con un nivel de competencia más bajo en la L2 (un nivel A2 inicial) observamos que la función del gesto puede ser ayudar a la gestión de la estructura del discurso o a recuperar palabras olvidadas. Aunque el uso del gesto no se incluya formalmente en el programa, se tratará como rasgo útil en la comunicación.

A lo largo de este resumen se han mencionado un número de propuestas para futuros trabajos de investigación, tanto en el campo de las disfluencias como en el de gestos, consideramos que una de las áreas más interesantes es la de la relación entre las cargas cognitivas y la disfluencia con el gesto.

11 BIBLIOGRAPHY[119]

Aesop/Esopo (n.d). *El león y el ratón*. Retrieved from http://www.pequered.com/cuento-el-leon-y-el-raton/

Aesop/Esopo (n.d). *The lion and the mouse*. Retrieved from www.cpalms.org/Uploads/resources/.../TeachingLionandtheMouse.docx

Albright, L., Cohen, A. I., Malloy, T. E., Christ, T., & Bromgard, G. (2004). Judgments of communicative intent in conversation. *Journal of Experimental Social Psychology, 40*, 290-302.

Alibali, M. W., & Kita, S. (2010). Gesture highlights perceptually present information for speakers. *Gesture, 10(1*), 3-28.

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15*, 593-613.

Andersen, J. F., Andersen, P. A., & Jensen, A. D. (1979). The measurement of nonverbal immediacy. *Journal of Applied Communication Research*, *7*, 153-180.

Andric, M., Solodkin, A., Buccino, G., Goldin-Meadow, S., Rizzolatti, G., & Small, S. L. (2013). Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia, 51*(8), 1619-1629.

Arbib, M. A. (2012). *How the brain got language*. Oxford, UK: Oxford University Press.

Armstrong, D. (1999). *Original Signs, gesture, sign and the sources of language*. Washington, DC: Gallaudet University Press.

Austin, J. L. (1962/1975). *How to do things with words*. Oxford, UK: Oxford University Press.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science 15*, 415-419.

[119] According to APA 6th edition.

Banich, M. T., & Compton, R. J. (2011). *Cognitive Neuroscience* (3rd ed.). Belmont, CA: Wadsworth.

Bavelas, J. B. (1994). Gestures as part of speech: Methodological implications. *Research on Language and Social Interaction*, *27*, 201-221.

Bavelas, J. B. (2007). Face-to-face dialogue as a micro-social context. In S. D. Duncan, J. Cassell & E. T. Levy (Eds.), *Gesture and the dynamic dimension of language. Essays in honor of David McNeill* (pp. 127-146). Amsterdam, Netherlands: John Benjamins.

Bavelas, J. B., & Chovil, N. (2006). Hand gestures and facial displays as part of language use in face-to-face dialogue. In V. Manusov & M. Patterson (Eds.), *Handbook of nonverbal communication* (pp. 97-115). Thousand Oaks, CA: Sage.

Bavelas, J. B., & Gerwing, J. (2007). Conversational hand gestures and facial displays in face-to-face dialogue. In K. Fiedler (Ed.), *Social communication* (pp. 283-308). New York, NY: Psychology Press.

Bavelas, J. B., & Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. *International Journal of Listening, 25*(3), 178-198.

Bavelas, J. B., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing. A review. *Gesture*, *13*, 63-92.

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, *15*, 469-489.

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, *52*, 566-580.

Bavelas, J. B., Gerwing, G., Allison, M., & Sutton, C. (2011). Dyadic evidence for grounding with abstract deictic gestures. In G. Stam, M. Ishino & R. Ashley (Eds.), *Integrating gestures: The interdisciplinary nature of gesture* (pp. 49-60). Amsterdam, Netherlands: John Benjamins.

Bavelas, J. B., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495-520.

Bavelas, J. B. (n.d.) *Microanalysis of dialogue.* Retrieved from http://web.uvic.ca/psyc/bavelas/Micro_dialogue.html

Benus, S., Gravano, A., & Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics, 43*(12), 3001-3027.

Bergen, B. (2012). *Louder Than words: The new science of how the mind makes meaning*. New York, NY: Basic Books.

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition, 39*(3), 173-194.

Blas Arroyo, J. L. (1995). La interjección como marcador discursivo. *Anuario de Lingüística Hispánica, 11*, 81-118.

Bortfeld, H., Leon, S., Bloom, J. E., Schober, M., & Brennan, S. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*(2), 123–147.

Brown, A. (2010). Gesture viewpoint in Japanese and English: Cross-linguistic interactions between two languages in one speaker. In M. Gullberg & K. de Bot (Eds.), *Gestures in Language Development* (pp. 113-134). Amsterdam, Netherlands: John Benjamins.

Brown, A., Quiros, L., Limon, F., Aspuerez, K., Church, R. B., & Mahootian, S. (2014). Cognitive load indicates simultaneous activation of two language systems for bilinguals: Alternative explanations. Poster presented at *6th Conference of the International Society for Gesture Studies*, University of California, San Diego, CA.

Brown, D. (1988). *Understanding research in second language learning*. Cambridge, MA: Cambridge University Press.

Burgoon, J. K., Buller, D. B., & Woodall, W. G. (1989). *Nonverbal communication: The unspoken dialogue.* New York, NY: HarperCollins/Greyden Press.

Burke, P. (1991). The language of gesture in early modern Italy. In J. Bremmer & H. Roodenburg (Eds.), *A cultural history of gesture* (pp. 71-83). Oxford, UK: Polity Press.

Butterworth, B., & Hadar, U. (1989). Gesture, speech and computational stages: A reply to McNeill. *Psychological Review, 96*, 168-74.

Caffarel, C. (Ed.). (2008). *Plan curricular del Instituto Cervantes. Niveles de referencia para el español A1/A2*. Madrid, Spain: Instituto Cervantes Biblioteca Nueva.

Campione, E., & Véronis, J. (2002). A Large-Scale Multilingual Study of Silent Pause Duration. In B. Bel & I. Marlien (Eds.), Proceedings of the *Speech Prosody 2002 Conference* (pp. 199-202). Aixen Provence, France: Laboratoire Parole et. Langage.

Cassell, J. (2007). Body language: Lessons from the near-human. In J. Riskin (Ed.), Genesis redux: *Essays in the history and philosophy of artificial intelligence* (pp. 346-374). Chicago, IL: University of Chicago Press.

Cassell, J., Kopp, S., Tepper, P., Ferriman, K., & Striegnitz, K . (2007). Trading spaces: How humans and humanoids use speech and gesture to give directions. In T. Nishida (Ed.), *Conversational Informatics* (pp. 133-160). New York, NY: John Wiley & Sons.

Choi, S. J., & Lantolf, J. P. (2008). The representation and embodiment of meaning in L2 communication: Motion events in speech and gesture in L2 Korean and L2 English speakers. *Studies in Second Language Acquisition, 30*, 191-224.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: The MIT Press.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General, 140(1*), 102-115.

Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics, 37*, 871–887.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Clark, H. H. (2002). Speaking in time. *Speech Communication, 36*, 5-13.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speech. *Cognition*, *84*, 73-111.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37*, 201-242.

Colletta, J. M., Pellenq, C., & Cefidekhanie, A. H. (2014). Age related changes in the processing of bimodal production between 3 and 11 years. Paper presentedat the *6th Conference of the International Society for Gesture Studies*. University of California, San Diego, CA.

Consejo de Europa (2002). *Marco Europeo común de referencia para las lenguas: aprendizaje, enseñanza, evaluación.* Madrid: Ministerio de Educación, Cultura y Deporte.

Cook, S. W., Yip, T., & Goldin-Meadow, S. (2010). Gesture makes memories that last. *Journal of Memory and Language*, *63*, 465-475.

Corballis, M. (2011). *The recursive mind. The origins of human language, thought, and civilization*. Princeton, NJ: Princeton University Press.

Corley, M., & Hartsuiker, R. J. (2011). Why um helps auditory word recognition: The temporal delay hypothesis. *PLoS ONE, 6*, e19792.

Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37*, 162-175.

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

Cutler, A. (1983). Speakers' conceptions of the function of prosody. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 79-91). Heidelberg: Springer.

Dauer, R. M. (1983). Stress timing and syllable timing reanalyzed. *Journal of Phonetics, 11,* 51 – 62.

De Fornel, M. (1992). The return gesture: Some remarks on context, inference, and iconic gesture. In P. A. & A. Di Luzio (Eds.), *The contextualization of language* (pp. 159-176). Amsterdam, Netherlands: John Benjamins.

De Ruiter, J. P. (1998). *Gesture and speech production.* Unpublished doctoral dissertation at Catholic University of Nijmegen, Netherlands.

De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284–312). Cambridge, NY: Cambridge University Press.

De Ruiter, J. P. (2007). Postcards from the mind: the relationship between speech, imagistic gesture, and thought. *Gesture*, *7*(1), 21-38.

De Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the Tradeoff Hypothesis. *Topics in Cognitive Science, 4*, 232-248.

Decety, J., & Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences, 3*(5), 172-178.

Dell, G. S., & Oppenheim, G. M. (2012). Insights for speech production planning from errors in inner speech. In M. Redford (Ed.), *Handbook of Speech Production* (n.p.). New York, NY: Wiley-Blackwell. Retrieved from http://pages.bangor.ac.uk/~pss238/pubs/dellOppenheim_revised_innerSlipsChapter_201208.pdf

Dell, G., & Reich, P. (1980). Toward a unified model of slips of the tongue. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 273-286). New York. NY: Academic Press.

Deterding, D., Wong, J., & Kirkpatrick, A. (2008). The pronunciation of Hong Kong English. *English World-Wide, 29*(2), 148–175.

Duncan, S. (2005). Gesture in signing: A case study in Taiwan Sign Language. *Language and Linguistics, 6*(2), 279-318.

Duncan, S. (2006). Co-expressivity of speech and gesture: manner of motion in Spanish, English, and Chinese. *Proceedings of the 27th Berkeley Linguistic Society Annual Meeting* (pp. 353-370). Berkeley, CA: Berkeley University Press.

Duncan, S. Jr. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, *23*(2), 283-292.

Duncan, S., & Pedelty, L. (2007). Discourse focus, gesture, and disfluent aphasia. In S. D. Duncan, J. Cassell & E. T. Levy (Eds.), *Gesture and the dynamic dimension of language* (pp. 269-284). Amsterdam, Netherlands: John Benjamins.

Edmunds, P. (2006). Buen- Buena gente: Repair in the Spanish of the Southwest. In T. L. Face & C. A. Klee (Eds.), *Selected Proceedings of the 8th Hispanic Linguistics Symposium* (pp. 204-213). Somerville, MA: Cascadilla Proceedings Project.

Efron, D. (1941). *Gesture and Environment*. Morningside Heights, NY: King's Crown Press.

Egbert, M. (1997). Some interactional achievements of other-initiated repair in multi-person conversation. *Journal of Pragmatics, 27*, 611-34.

Egbert, M. (2004). Other-initiated repair and membership categorization—some conversational events that trigger linguistic and regional membership categorization. *Journal of Pragmatics, 36*(8), 1467-1498.

Ekman, P., & Friesen, W. V. (1969). The Repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica, 1*(1), 49-98.

Ellis, N. (2005). Principles of instructed language learning. In L. Ortega (Ed.). (2011), *Second Language Acquisition. Critical Concepts in Linguistics*, vol. VI (pp. 291- 307). New York, NY: Routledge.

Erard, M. (2007). U*m…: Slips, stumbles, and verbal blunders, and what they mean*. New York, NY: Knopf Doubleday Publishing Group.

Erasmus, D. (1530/2003). On good manners. In E. Rummel (Ed.), *The Erasmus reader* (pp. 101-118). Toronto: University of Toronto Press.

Fedorenko, E., & Kanwisher, N. (2009). Neuroimaging of language: Why hasn't a clearer picture emerged? *Language and Linguistics Compass, 3*, 839-65.

Feyereisen, P. (1987). Gestures and speech, interactions and separations. A reply to David McNeill (1985). *Psychological Review, 94*(4), 493-498.

Feyereisen, P. (2003). Are human gestures in the present time a mere vestige of a former sign language? Probably not. *Behavioral and Brain Sciences, 26*(02), 220 - 221.

Finlayson, S., Forrest, V., Lickley , R., & Mackenzie Beck, J. (2003). Effects of the restriction of hand gestures on disfluency. Proceedings of DISS, Gothenburg Papers in *Theoretical Linguistics*, *90,* 21-24.

Foster, N. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In L. Ortega (Ed.). (2011), *Second Language Acquisition. Critical Concepts in Linguistics*, vol. VI (pp. 308- 317). New York, NY: Routledge.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition 62*(2), 151-167.

Fromkin, V. (1971). The Non-anomalous nature of anomalous utterances. In V. Fromkin (Ed.). (1973), *Speech errors as linguistic evidence* (pp. 215-242). The Hague: Mouton & Co.

Fromkin, V. (Ed.). (1973). *Speech errors as linguistic evidence*. The Hague, Netherlands: Mouton & Co.

García Mouton, P. (1994). *Lenguas y dialectos de España*. Madrid: Arco Libros.

García-García, M. (2009). *El Lugar de la comunicación no verbal en la clase de ELE kinésica contrastiva*. Unpublished doctoral dissertation. Salamanca University, Spain.

Garrett, M. (1975). The analysis of sentence production. *Psychology of Learning and Motivation, 9*, 133-177.

Garrett, M. (1992). Disorders of lexical selection. *Cognition, 42*(1), 143 -180.

Gelbes, S. (2003). The particle "eh" and the Theory of Relevance. An example of procedural content. *Estudios Filológicos, 38,* 157-177. Retrieved from http://mingaonline.uach.cl/scielo.php?script=sci_arttext&pid=S0071-17132003000100010&lng=en&nrm=iso

Goldin-Meadow, S. (2001). Giving the mind a hand: The role of gesture in cognitive change. In J. McClelland & R. S. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives* (pp. 5-31). Mahwah, NJ: Earlbaum Associates.

Goldin-Meadow, S. (2009). How gesture promotes learning throughout childhood. *Child Development Perspectives, 3*, 106-111.

Goldin-Meadow, S. (2010). When gesture does and does not promote learning. *Language and Cognition, 2*, 1-19.

Goldin-Meadow, S., & McNeill, D. (1999). The role of gesture and mimetic representation in making language the province of speech. In M. C. Corballis & S. E. G. Lea (Eds.), *The descent of mind: Psychological perspectives on hominid evolution* (pp. 155-172). New York, NY: Oxford University Press.

Goldin-Meadow, S., Nusbaum, H., Kelly, S., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, *12*, 516-522.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.

Goodwin, C. (1979). The Interactive Construction of a Sentence in Natural Conversation. In G. Psathas (Ed.), *Everyday language: Studies in Ethnomethodology* (pp. 97-121). New York, NY: Irvington.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32,* 1489-522.

Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica, 62*(1-2), 51-75.

Graziano, M., & Gullberg, M. (2013). Gesture production and speech fluency in competent speakers and language learners. Paper delivered at *TIGER*, Tilburg University, Netherlands.

Graziano, M., & Gullberg, M. (2014). When speech stops, gesture stops: Evidence from crosslinguistic and developmental comparisons. Paper delivered at *6th Conference of the International Society for Gesture Studies*, University of California, San Diego, CA.

Grice, H. P. (1968). Utterer's meaning, sentence meaning, and word meaning, foundations of language. In H. P. Grice (1989), *Studies in the way of words* (pp. 117–137). Harvard, MA: Harvard University Press.

Gullberg, M. (1998). *Gesture as a Communication Strategy in Second Language Discourse: A Study of Learners of French and Swedish*. Travaux de l'institut de linguistique de Lund, 35. Linguistics and Phonetics. Unpublished doctoral dissertation. Retrieved from http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=18684&fileOId=3912717

Gullberg, M. (1999). Gestures in spatial descriptions. *Working Papers*, 47. Dept. of Linguistics, Lund University.

Gullberg, M. (2003). Gestures, referents, and anaphoric linkage in learner varieties. In C. Dimroth & M. Starren (Eds.), *Information structure and the dynamics of language acquisition* (pp. 311-328). Amsterdam, Netherlands: John Benjamins.

Gullberg, M. (2006). Handling discourse: Gestures, reference tracking, and communication strategies in early L2. *Language Learning*, *56*(1), 155-196.

Gullberg, M. (2008). Gestures and second language acquisition. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 276-305). New York, NY: Routledge.

Gullberg, M. (2010). Methodological reflections on gesture analysis in SLA and bilingualism research. *Second Language Research, 26*, 75-102.

Gullberg, M. (2014). Why gestures are not (only) a compensatory device – evidence from language learners. Paper delivered at *6th Conference of the International Society for Gesture Studies*, University of California, San Diego, CA.

Gullberg, M., & Kita, S. (2009). Modulating addressees' attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior, 33(4),* 251-277.

Gullberg, M., & McCafferty, S. (2008). Introduction to gesture and SLA: Toward an integrated approach. In S. McCafferty, & M. Gullberg, Gesture and SLA: Toward an integrated approach (pp. 133-146). *Special Issue Studies in Second Language Acquisition, 30*(2).

Gullberg, M., de Bot, K., & Volterra, V. (2010). Gestures and some key issues in the study of language development in M. Gullberg & K. de Bot (Eds.), *Gestures in Language Development (pp.* 3-35). Amsterdam, Netherlands: John Benjamins.

Hadar, U., & Butterworth, B. (1997). Iconic gesture, imagery and word retrieval in speech. *Semiotica, 115*, 147-172.

Hagoort, P., & Levelt, W. J. M. (2009). The speaking brain. *Science, 326*, 372-373. Retrieved from http://www.mpi.nl/world/materials/publications/Levelt/Hagoort_Levelt_Perspectiv es _SpeakingBrain-Science.pdf

Hardway, C. L., & Duncan, S. D. (2005). Me first! Structure and dynamics of a four-way family conflict. In L. Anolli, S. D. Duncan, Jr. & M. Magnusson (Eds.), *The hidden structure of social interaction* (pp. 210-224). Amsterdam, Netherlands: IOS Press.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology, 42*, 113-157.

Hartsuiker, R. J., Pickering, M. J., & De Jong, N. (2005). Semantic and phonological context effects in speech error repair. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 921-932.

Hartsuiker, R. J., Roelin, B., Postma, A., & Wijnen, F. (2005). Phonological encoding and monitoring in normal and pathological speech. In R. J. Hartsuiker, R. Bastiaanse, A. Postma & F. Wijnen (Eds.), *Phonological Encoding and monitoring in normal and pathological speech* (pp. 1-14). New York, NY: Psychology Press.

Hauk, O., Johnsrude, I., Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41,* 301-307.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555-568.

Hernández Alcaide, C. (2013). *Grammar in the classroom. Learning grammar structures to know them, or to use them?* Workshop presented at Hong Kong University, Hong Kong.

Hostetter, A. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137*, 297-315.

Ibáñez, F. (2007). *Mortadelo y Filemón. ¡El dopaje qué potaje!* Barcelona: Ediciones B.

Iverson, J., & Goldin-Meadow, S. (1997). What's communication got to do with it? Gesture in children blind from birth. *Developmental Psychology, 33*, 453-467.

Jack, R., Caldara, R., & Schyns, P. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General*, *141*(1), 19-25.

Jefferson, G. (1974). Error correction as an interactional resource. *Language in Society, 3*(2), 181-199.

Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation'. In D. Roger & P. Bull (Eds.), *Conversation: an interdisciplinary perspective* (pp. 166-196). Clevedon, UK: Multilingual Matters.

Jokinen, K., Nishida, M., & Yamamoto, S. (2011). *On eye-gaze and turn-taking*. Retrieved from http://links.cse.msu.edu:8000/iui/papers/jokinen_turn_taking.pdf

Josse, G., & Tzourio-Mazoyer, N. (2003). Hemispheric specialization for language. *Brain Research Reviews, 44*(1), 1-12.

Keenan, E. O. (1976). On the universality of conversational postulates. *Language in Society, 5*(1), 67–80.

Kellerman, E., & Van Hoof, A. M. (2003). Manual accents. *International Review of Applied Linguistics, 41,* 251-69.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*, 253-260.

Kemmerer, D. (2010). How words capture visual experience: The perspective from cognitive neuroscience. In B. C. Malt & P. Wolff (Eds.), *Words and the mind. How words capture human experience* (pp. 287-327). New York, NY: Oxford University Press.

Kendall, T. (2013). *Speech rate, pause and sociolinguistic variation: Studies in corpus sociophonetics*. Basingstoke, UK: Palgrave Macmillan.

Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica, 26,* 22-63.

Kendon, A. (1972). Some relationships between body motion and speech. In A. Seigman & B. Pope (Eds.), *Studies in Dyadic Communication* (pp. 177-216). Elmsford, NY: Pergamon Press.

Kendon, A. (1983). Gesture and Speech: How they interact. In Wiemann & R. Harrison (Eds.), *Nonverbal interaction: Sage annual reviews of communication*, Vol. 11, (pp. 13-46), Beverly Hills, CA: Sage Publications.

Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction, 27*(3), 175-200.

Kendon, A. (2004). *Gesture: Visible action as utterance.* New York, NY: Cambridge University Press.

Kida, T. (2008). Does gesture aid discourse comprehension in second language? In S. G. McCafferty & G. Stam (Eds.), *Gestures and second language acquisition* (pp. 131-156). Mahwah, NJ: Erlbaum.

Kimbara, I. (2007). Indexing locations in gesture: Recalled stimulus image and interspeaker coordination as factors influencing gesture form. In S. Duncan, J. Cassel & E. Levy (Eds.), *Gesture and the dynamic dimension of language* (pp. 213-220). Amsterdam, Netherlands: John Benjamins.

Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., Konrad, K., & Green, A. (2008). Neural interaction of speech and gesture: differential activations of metaphoric co-verbal gestures. *Neuropsychologia, 47*(1), 169-79.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture: Window into thought and action* (pp. 162–185). New York, NY: Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*, 16-32.

Kita, S., & Özyürek, A. (2007). How does spoken language shape iconic gestures? In S. Duncan, J. Cassel & E. Levy (Eds.), *Gesture and the dynamic dimension of language* (pp. 67-74). Amsterdam, Netherlands: John Benjamins.

Kjellmer, G. (2003). Hesitation in defence of er and erm. *English Studies*, 84(2), 170-198.

Knapp, M. (2006). A historical overview of nonverbal research. In V. Manusov & M. L. Patterson (Eds.), T*he SAGE handbook of nonverbal communication* (pp. 3-20). Thousand Oaks, CA: SAGE Publications, Inc.

Krauss, R., Chen, Y., & Gottesman, R. (2000). Lexical gestures and lexical access: a process model. In McNeill, D. (Ed.), *Language and gesture* (pp. 261-283). Cambridge, NY: Cambridge University Press.

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review* (n.p.). Retrieved from http://www.psychology.sunysb.edu/sbrennan-/papers/confed.pdf

Lakoff, G. (2013). *George Lakoff on Embodied Cognition and Language*. Retrieved from https://www.youtube.com/watch?v=XWYaoAoijdQ

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition 14*, 41-104.

Levelt, W. J. M. (1984). Spontaneous self-repairs in speech: Processes and representations. In M. P. R. van den Broecke & A. Cohen (Eds.), Proceedings of the *10th International Congress of Phonetic Sciences* (pp. 105-117). Dordrecht: Foris.

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

Levelt, W. J. M., Cutler, A. (1983). Prosodic marking in speech repair. *Journal of Semantics, 2*, 205-217.

López-Ozieblo, R. (2012). *Manifestaciones sociolingüísticas para dar y tomar la palabra en la interacción oral de estudiantes hongkoneses de español*. Unpublished Master thesis. Jaén University, Spain.

López-Ozieblo, R. (2013a). Reflections on the use of nonverbal communication in teaching Spanish to Sino-speakers. *Revista de Lenguas para Fines Específicos, 19*, 321-340.

López-Ozieblo, R. (2013b). Using hand gestures to understand second language speech processing. Paper presented at *the 9th International Symposium on Teaching English at Tertiary Level*. Zhejiang University, Hangzhou, China.

López-Ozieblo, R., & McNeill, D. (forthcoming, 2015). Exchange on why we gesture. In R. B. Church & M. Alibali. Philadelphia, PA: John Benjamins.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia, 48*, 3982-3992.

MacKay, D. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York, NY: Springer.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*, 19-44.

Madrid, D., Pueyo, S., & Hockley, N. (n.d.). Observación e Investigación en el Aula (study notes from the Master programme of *Lingüística Aplicada a la Enseñanza del Español como Lengua Extranjera*. FUNIBER. Jaén University, Spain.

Magnusson, M. S. (2000). Discovering Hidden Time Patterns in Behavior: T-Patterns and their Detection. *Behavior Research Methods, Instruments and Computers, 32*(1), 93-110.

Magnusson, M. S. (2005). Understanding Social Interaction: Discovering Hidden Structure with Model and Algorithms. In L. Anolli, S. Duncan Jr., M. S. Magnusson & G. Riva (Eds.), *The hidden structure of social interaction* (pp. 4-22). Amsterdam, Netherlands: IOS Press.

Magnusson, M. S. (2006). Structure and communication in interaction. In G. Riva, M. T. Anguera, B. K. Wiederhold & F. Mantovani (Eds.), F*rom communication to presence: Cognition, emotions and culture towards the ultimate communicative experience* (pp. 125-144). Útgefandi: IOS Press.

Maheux-Pelletier, G., & Golato, A. (2008). Repair in membership categorization in French. *Language in Society, 37*(5), 689-712.

Mahl, G. F. (1987). *Explorations in nonverbal and vocal behavior*. Hillsdale, NJ: Erlbaum.

McCroskey, J., & McCroskey, L. (2006). Nonverbal communication in instructional contexts. In V. Manusov & M. L. Patterson (Eds.), T*he SAGE handbook of nonverbal communication* (pp. 421-436). Thousand Oaks, CA: SAGE Publications, Inc.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

McNeill, D. (2000). Catchments and contexts: Non-modular factors in speech and gesture production. In D. McNeill (Ed.), *Language and gesture: Window into thought and action* (pp. 312-328). New York, NY: Cambridge University Press.

McNeill, D. (2005). *Gesture and thought*. New York, NY: Cambridge University Press.

McNeill, D. (2009). Imagery for speaking. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura & Ş. Özçalişkan (Eds.), *Crosslinguistic approaches to the psychology of language* (pp. 517-530). New York, NY: Psychology Press.

McNeill, D. (2012). *How language began: Gesture and speech in human evolution*. New York, NY: Cambridge University Press.

McNeill, D., & Duncan, S. D. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture: Window into thought and action* (pp. 141-161). New York, NY: Cambridge University Press.

McNeill, D., Duncan, S., Franklin, A., Goss, J., Kimbara, I., Parrill, F., Welji, H., Chen, L., Harper, M., Quek, F., Rose, T., & Tuttle, R. (2010). Mind Merging. In E. Morsella (Ed.), *Expressing oneself/expressing one's self: Communication, cognition, language, and identity* (pp. 143-164). New York, NY: Psychology Press.

Mehrabian, A. (1972). *Nonverbal communication*. Chicago, IL: Aldine-Atherton.

Meister, I. G., Boroojerdi, B., Foltys, H., Sparing, R., Huber, W., & Topper, R. (2003). Motor cortex hand area and speech: Implications for the development of language. *Neuropsychologia*, *41*, 401–406.

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture, 4*, 119-141.

Mondada, L. (2012). Interacting bodies: Multimodal resources for the organization of social interaction. Paper presented at the *6th International Conference on Multimodality*. Institute of Education, London, UK.

Moniz, H., Batista, F., Mata, A., & Trancoso, I. (2012). Analysis of disfluencies in a corpus of university lectures. *Proceedings of the International Conference of Experimental Linguistics. ExLing 2012*, 27-29 August 2012, Athens, Greece.

Monterubbianesi, M. G. (2013). La comunicación no verbal en los manuales de E/LE. *Revista Electrónica de Didáctica del Español Lengua Extranjera*, *25*. Retrieved from http://www.mecd.gob.es/dctm/redele/MaterialRedEle/Revista/2013/2013_redELE_25_03Monterubbianesi.pdf?documentId=0901e72b8159e598

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts the temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 615-623.

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *American Journal of Psychology, 117*(3), 414-424.

Negueruela, E. J., Lantolf, J. P., Reh Jordan, S., & Gelabert, J. (2004). The "private function" of gesture in second language communicative activity: A study on motion verbs and gesturing in English and Spanish. *International Journal of Applied Linguistics, 14*, 113-147.

Nemeth, Z. (2012). Recycling and replacement repairs as self-initiated same-turn self-repair strategies in Hungarian. *Journal of Pragmatics, 44*(14), 2022–2034.

Neu, J. (1990). Assessing the role of nonverbal communication in the acquisition of communicative competence in L2. In R. Scarcella, E. Andersen & S. D. Krashen (Eds.), *Developing communicative competence in a second language* (pp. 121-138). New York, NY: Newbury House Publishers.

Neuliep, J. W. (2006). *Intercultural communication: A contextual approach* (2nd ed.). New York, NY: Houghton Mifflin.

Nicoladis, E. (2007). The effect of bilingualism on the use of manual gestures. *Applied Psycholinguistics*, *28*, 441-454.

Nicoladis, E., Pika, S., Yin, H., & Marentette, P. (2007). Gesture use in story recall by Chinese-English bilinguals. *Applied Psycholinguistics, 28*, 721-735.

Nooteboom, S. (1980). Speaking and unspeaking: detection and correction of phonological and lexical errors in spontaneous speech. In V. Fromkin (Ed.), *Errors in linguistic performance* (pp. 215-240). New York, NY: Academic Press.

Norris, J., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. In L. Ortega (Ed.). (2011), *Second Language*

Acquisition. Critical concepts in linguistics, vol. VI (pp. 249-288). New York, NY: Routledge.

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language, 9*, 19–35.

Özyürek, A. (2000). Differences in spatial conceptualization in Turkish and English discourse: Evidence from both speech and gesture. In A. Goksel & C. Kerslake (Eds.), *Studies on Turkish and Turkic languages* (pp. 263-272). Wiesbaden: Harrassowitz.

Özyürek, A., & Kita, S. (1999). Expressing manner and path in English and Turkish: Differences in speech, gesture, and conceptualization. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society* (pp. 507-512). London: Erlbaum.

Özyürek, A., Kita, S., Allen, S., Furman, R., & Brown, A. (2005). How does linguistic framing of events influence co-speech gestures? Insights from crosslinguistic variations and similarities. *Gesture, 5*(1), 215-237.

Pika, S., Nicoladis, E., & Marentette, P. (2006). A cross cultural study on the use of gestures: Evidence for cross-linguistic transfer. *Bilingualism, Language and Cognition, 9*(3), 319-327.

Portoles Lazaro, J., & Martin Zorraquino, A. M. (1999). In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (pp. 4051-4211). Madrid, Spain: R.A.E.

Postma, A. (2000). Detection of errors during speech production. A review of speech monitoring models. *Cognition, 77*, 97-131.

Postma, A., & Oomen, C. (2005). Critical issues in speech monitoring. In R. J. Hartsuiker, R. Bastiaanse, A. Postma & F. Wijnen (Eds.), *Phonological Encoding and Monitoring in Normal and Pathological Speech* (pp. 157-166). New York, NY: Psychology Press.

Poyatos, F. (1992). The audible-visual approach to speech as basic to nonverbal commination research. In F. Poyatos (Ed.), *Advances in nonverbal communication* (pp. 3-24). New York, NY: John Benjamins.

Poyatos, F. (1993). *Paralanguage: A linguistic and interdisciplinary approach to interactive speech and sound*. Amsterdam, Netherlands: John Benjamins.

Poyatos, F. (n.d.). *Paralingüística y kinésica: para una teoría del sistema comunicativo In el hablante español*. Retrieved from http://dialnet.unirioja.es/servlet/articulo?codigo=1221471

Queck, F., McNeill, D., Bryll, R., Ma, X. F., Kirbas, C., McCullough, K., & Ausari, R. (2002). Multimodal human discourse: Gesture and speech. *ACM Transaction on Computer-Human Interaction, 9*(3), 171-193.

Quintilian (90AD/1920). *Institutio Oratoria*. (H.E. Butler Trans.). Harvard, MA: Harvard University Press.

Quinting, G. (1971). *Hesitation Phenomena in Adult Aphasic and Normal Speech*. The Hague, Netherlands: Mouton & Co.

Raffler-Engel, W. (1980). *Aspects of nonverbal communication*. Amsterdam, Netherlands: Swets and Zeitlinger.

Ramachandran, V. S. (2012). *The tell-tale brain*. London, UK: Windmill Books.

Rauscher, F. B., Krauss R. M., & Chen Y. (1996). Gesture, speech and lexical access: The role of lexical movements in speech production. *Psychological Science, 7,* 226–230.

Rezaei, S., Mozaffari, F., & Hatef, A. (2011). Corrective Feedback in SLA: Classroom Practice and Future Directions. *International Journal of English Linguistics 1*(1), 21–29. Retrieved from www.ccsenet.org/journal/index.php/ijel/article/download/8719/7021

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. Annual Review of Neuroscience *27*, 169-192. Retrieved from http://www.annualreviews.org/doi/abs/10.1146/annurev.neuro.27.070203.144230

Rodríguez, L. J., & Torres, M. I. (2006). Spontaneous speech events in two speech databases of human-computer and human-human dialogs in Spanish. *Language and Speech, 49*(3), 333-366.

Rossini, N. (2012). *Reinterpreting gesture as language*. Amsterdam, Netherlands: IOS Press.

Rühlemann, C., Bagoutdinov, A., & O'Donnell, M. (2011). Windows on the mind: Pauses in conversational narrative. *International Journal of Corpus Linguistics 16*(2), 198-232.

Sacks, O. (1985). *The man who mistook his wife for a hat*. London: Duckworth; New York, NY: Summit Books.

Sacks, O. (1990). *Seeing voices: A journey into the world of the deaf*. New York, NY: Harper Collins Publishers.

Saville-Troike, M. (2006). *Introducing second language acquisition* (2nd ed.). Cambridge, UK: Cambridge University Press.

Schegloff, E. (1984). On some gestures' relation to talk. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 266-296). Cambridge, MA: Cambridge University Press.

Schegloff, E. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn taking organisation. In G. Button & J. R. E. Lee (Eds.). *Talk and social organisation* (pp 70-85), Clevedon, NJ: Multilingual Matters.

Schegloff, E. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society, 29*, 1-63.

Schegloff, E., Jefferson G., & Sacks, H. (1977). The Preference for Self- Correction in the Organization of Repair in Conversation. *Language, 53(*2), 361-382.

Schmitt, J. C. (1991). The rationale of gesture in the West: Third to thirteenth centuries. In J. Bremmer & H. Roodenburg (Eds.), *A cultural history of gesture* (pp. 59-70). Oxford, UK: Polity Press.

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience, 10*(4), 295-302.

Seo, M. S., & Koshik, I. (2010). A conversation analytic study of gestures that engender repair in ESL conversational tutoring. *Journal of Pragmatics, 42*(8), 2219-2239.

Setter, J. (2012). Speech rhythm in World Englishes: The case of Hong Kong. *TESOL Quarterly, 40*(4), 763-782.

Seyfeddinipur, M. (2006). *Disfluency: Interrupting speech and gesture*. Doctoral dissertation, Radboud University Nijmegen, Nijmegen. Retrieved from http://hdl.handle.net/11858/00-001M-0000-0013-1B6F-F

Seyfeddinipur, M., & Kita, S. (2001). Gestures and self-monitoring in speech production. Proceedings of the *Annual Meeting of the Berkeley Linguistics Society, 27*(1). Retrieved from http://elanguage.net/journals/bls/ article/view/717/576

Seyfeddinipur, M., Kita, S., & Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition, 108(3),* 837-842.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language, 101*(3), 260-277.

Slezak-Swiat, A. (2010). Observable Strategizing: On limbic communication in advance users of language. In J. Arabski & A. Wojtaszek (Eds.), *Neurolinguistic and psycholinguistic perspectives on SLA* (pp. 64-78). Bristol, UK: Multilingual Matters.

Slobin, D. (1996). Two ways to travel: Verbs of motion in English and Spanish. In M. Shibatani & S. A. Thompson (Eds.), *Grammatical constructions: Their form and meaning* (pp. 195–219). Oxford, UK: Oxford University Press.

So, W. C., Demir, O. E., & Goldin-Meadow, S. (2010). When speech is ambiguous gesture steps in: Sensitivity to discourse-pragmatic principles in early childhood. *Applied Psycholinguistics, 31,* 209-224.

Stam, G. (2006). Thinking for speaking about motion: L1 and L2 speech and gesture. *International Review of Applied Linguistics, 44*(2), 146-171.

Stam, G. (2012). Gestes et recherche de mots en langue seconde. In R. Vion, A. Giacomi & C. Vargas (Eds.), *La corporalité du langage: Multimodalité, discours et écriture* (pp. 55-71). Provence, France: Presses Universitaires de Provence.

Stam, G. (2014). A second language learner's thinking for speaking in her L1 and L2 after fourteen years: Verb framed, satellite-framed or in between? Paper delivered at the *6th Conference of the International Society for Gesture Studies*, University of California, San Diego.

Stenström, A. (1994). *An Introduction to Spoken Interaction*. London and New York, NY: Longman.

Talmy, L. (2009). Main verb properties and equipollent framing. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura & Ş. Özçalişkan (Eds.), *Crosslinguistic approaches to the psychology of language* (pp. 193-208). New York, NY: Psychology Press.

Tellier, M. (2010). The effect of gestures on second language memorisation by young children. In M. Gullberg & K. de Bot (Eds.), *Gestures in Language Development* (pp. 75-92). Amsterdam, Netherlands: John Benjamins.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., & Scifo, P. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience, 17*(2), 273-281.

Tottie, G. (2011). Uh and um as sociolinguistic markers in British English. *The International Journal of Corpus Linguistics, 16*, 173–196.

Tydgat, I., Stevens, M., Hartsuiker, R., & Pickering, M. (2011). Deciding when to stop speaking. *Journal Of Memory And Language, 64*(4), 359-380.

Van Patten, B., & Cadierno, T. (1993). Input processing and second language acquisition: a role for instruction. In L. Ortega (Ed.). (2011), *Second Language Acquisition. Critical concepts in linguistics*, vol. VI (pp. 101-120). New York, NY: Routledge.

Van Wijk, C., & Kempen, G. (1987). *From sentence structure to intonation contour* (n.p). Retrieved from http://pubman.mpdl.mpg.de/pubman/item/escidoc:1147606:8/component/escidoc:1271607/VanWijk_Kempen--FromSentenceStructure_1985.pdf

VARK (n.d.). Retrieved from http://www.vark-learn.com/english/index.asp

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage, 30*(4), 1414-1432.

Warner Bros. (1950). *Canary Row*. Retrieved from: http://vimeo.com/74506505

Weatherall, A. (2002). *Gender, language and discourse*. Hove England: New York, NY: Routledge.

Weisberg, H., & Emmorey, K. (2014). Neural responses during perception of naturally produced, meaningful co-speech gestures. Paper delivered at the *6th Conference of the International Society for Gesture Studies*, University of California, San Diego, CA.

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language, 101*(3), 278-289.

Wong, J. (2000). Delayed next turn repair initiation in native-nonnative speaker English conversation. *Applied Linguistics, 21*(2), 244-267.

Wong, J., & Olsher, D. (2000). Reflections on conversation analysis and nonnative speaker talk: An interview with Emanuel A. Schegloff. *Issues in Applied Linguistics, 11*(1), 111-128.

12 APPENDICES

12.1 Appendix 1 – Written, Aural And Video Input for the Retellings

Participants were asked to retell a story they had read, heard or watched. The three stories, the input, are transcribed below (in the case of the video input a summary of the story is provided).

12.1.1 Input in English

Initially participants retold the written, aural and half of the video stories in their mother tongue (English or Spanish). After retelling three stories in their mother tongue, participants were asked to retell the second half of the video in the L2. HK students were given the written and aural input in English.

12.1.1.1 Written input

The lion and the mouse[120] - Aesop/Esopo (n.d.). Number of words: 381.

Once, as a lion lay sleeping in his den, a naughty little mouse ran up his tail, and onto his back and up his mane and danced and jumped on his head ......so that the lion woke up.

The lion grabbed the mouse and, holding him in his large claws, roared in anger. 'How dare you wake me up! Don't you know that I am King of the Beasts? Anyone who disturbs my rest deserves to die! I shall kill you and eat you!'

The terrified mouse, shaking and trembling, begged the lion to let him go. 'Please don't eat me Your Majesty! I did not mean to wake you, it was a mistake. I was only playing. Please let me go - and I promise I will be your friend forever. Who knows but one day I could save your life?'

[120] Recovered from www.cpalms.org/Uploads/resources/.../TeachingLionandtheMouse.docx.

The lion looked at the tiny mouse and laughed. 'You save my life? What an absurd idea!' he said scornfully. 'But you have made me laugh, and put me into a good mood again, so I shall let you go.' And the lion opened his claws and let the mouse go free.

'Oh thank you, your majesty,' squeaked the mouse, and scurried away as fast as he could.

A few days later the lion was caught in a hunter's snare. Struggle as he might, he couldn't break free and became even more entangled in the net of ropes. He let out a roar of anger that shook the forest. Every animal heard it, including the tiny mouse.

My friend the lion is in trouble,' cried the mouse. He ran as fast as he could in the direction of the lion's roar, and soon found the lion trapped in the hunter's snare. 'Hold still, Your Majesty,' squeaked the mouse. 'I'll have you out of there in a jiffy!' And without further delay, the mouse began nibbling through the ropes with his sharp little teeth. Very soon the lion was free.

'I did not believe that you could be of use to me, little mouse, but today you saved my life,' said the lion humbly.

'It was my turn to help you, Sire,' answered the mouse.

Even the weak and small may be of help to those much mightier than themselves.

12.1.1.2 Aural input

Tweety and the worm, created for this study. Length 1’31” Number of words: 228.

So one day, Tweety Bird was swinging away in his cage, when he looked out of the window and saw a big fat juicy warm sunning himself in the garden. And so Tweety Bird decided to fly downstairs and fly out into the garden to try to catch the warm to eat. So Tweety Bird opened the gate of the cage, flew down very, very quickly downstairs (I don’t know why he flew downstairs and not just out of the window maybe the window was closed), flew downstairs and flew out of the window to catch the juicy worm to eat. But what Tweety Bird didn’t know was that just outside the house and under the window, Sylvester was waiting for him. And as Tweety flew out Sylvester jumped up and he jumped up and he caught Tweety but what Sylvester didn't know was that the Granny, Tweety's owner, was ironing in the downstairs room and as he jumped up to grab Tweety, Granny saw him and she reached out of the window and wacked him with the iron on the head. And the force of the whack with the iron was such, that Tweety, not Tweety, I beg your pardon, Sylvester, released Tweety and rolled down the slope the grassy slope to the bottom where there was a pond and he went into the water head first.

12.1.1.3 Video input

Canary Row, Warner Brothers (1950)[121]

First part (to be retold in the L1). Length 3'44" (including opening titles).

Sylvester is looking for Tweety through a set of binoculars and Tweety, in the building opposite, looks back at Sylvester and as he their eyes meet Tweety is horrified. Sylvester then tries four times to catch Tweety:

The first time he walks into the building but is thrown out as the building does not allow cats or dogs. The second time he climbs up through the drain pipe to Tweety's window, but Tweety sees him, flees his cage, gets Granny, his owner, and she comes and whacks Sylvester with an umbrella. The third time Sylvester climbs up through the drain pipe but Tweety sees him and throws a bowling ball down the drain. Sylvester, still inside the drain, swallows the ball and rolls down the street to a bowling alley. The fourth time Sylvester steals the clothes from a street performing monkey and enters Tweety's apartment. He starts to search for Tweety under the carpet and Granny's skirt. She seems to be taken by the ruse until Sylvester lifts off his cap to thank her for a coin and she whacks him again with an umbrella.

Second part (to be retold in the L2). Length: 3" (including closing titles).

Sylvester is hiding in a pigeon hole behind reception listening to the conversation between the receptionist and Granny. It seems that Granny wants to leave and so Sylvester

[121] Recovered from http://vimeo.com/74506505.

dresses up as a bell boy and goes upstairs to pick up her bags. He takes a suitcase and the bird cage, disposes of the first and takes the cage to the back alley where he uncovers it to discover Granny, who jumps out of the cage and proceeds to whack him with an umbrella.

The cat's next try is to stand at the end of a fulcrum, he throws a heavy weight onto the other end and is propelled to Tweety's window, he catches him mid-air and is back on the ground when the weight comes down on top of his head. Next Sylvester is in the building opposite calculating the angle necessary to swing from his window to Tweety's, he swings across and stamps himself against the wall, sliding to the ground. Finally he notices the tram cables leading to Tweety's window, he climbs up a pole and walks along the cable only to be chased by a tram, driven by Tweety and Granny, which keeps giving him electric shocks.

12.1.2 Input in Spanish

Participants from Spain were given the text and aural inputs in Spanish, their mother tongue, and asked to retell them in Spanish.

12.1.2.1 Written input

El León y el Ratón,[122] Aesop/Esopo (n. d.). Number of words: 340.

Érase una vez en una tarde muy calurosa, un león dormitaba en una cueva fría y oscura. Estaba a punto de dormirse del todo cuando un ratón se subió a su cola y se puso a corretear sobre el lomo del león, luego por su melena hasta llegar al hocico.

Con un rugido iracundo, el león levantó su pata y aplastó al ratón contra el suelo. -¿Cómo te atreves a despertarme? Gruñó. Te voy a espachurrar. Oh, por favor, por favor, perdóname la vida chilló el ratón atemorizado. Prometo ayudarle algún día si me deja marchar.

[122] Recovered from http://www.pequered.com/cuento-el-leon-y-el-raton.

-¿Quieres tomarme el pelo? Dijo el León; -¿Cómo podría un ratoncillo como tú ayudar a un león grande y fuerte como yo? Se echó a reír con ganas. Se reía tanto que un descuido deslizó su pata y el ratón escapó.

Unos días más tarde el león salió de caza por la jungla. Estaba justamente pensando en su próxima comida cuando tropezó con una cuerda estirada en medio del sendero. Una red enorme se abatió sobre él y pese a toda su fuerza, no consiguió liberarse. Cuanto más, se removía y se revolvía, más se enredaba y más se tensaba la red en torno a él.

El león empezó a rugir tan fuerte que todos los animales le oían, pues sus rugidos llegaban hasta los mismos confines de la jungla. Uno de esos animales era el ratoncillo, que se encontraba royendo un grano de maíz. Soltó inmediatamente el grano y corrió hasta el león.

-Oh, poderoso león -chilló- Si me hicieras el favor de quedarte quieto un ratito podría ayudarte a escapar. El león se sentía ya tan exhausto que permaneció tumbado mirando como el ratón roía las cuerdas de la red.

Apenas podía creerlo cuando, al cabo de un rato, se dio cuenta que estaba libre. Me salvaste la vida, ratoncillo dijo.

Nunca volveré a burlarme de las promesas hechas por los amigos pequeños.

Moraleja

No importa como de débil y pequeña sea la criatura puede ser de ayuda si llega el momento.

12.1.2.2 Aural input

Piolín y la lombriz, (created for this study) - Length 1’25”. Number of words: 165.

Estaba el pájaro Piolín tranquilamente descansando en su jaula cuando miró por la ventana y vio que en el jardín había una hermosa lombriz tostándose al sol. Sin pensárselo dos veces Piolín salió de la jaula y voló hasta el piso de abajo, donde estaba la ventana abierta, para salir al jardín a comerse a la lombriz. Lo que Piolín no sabía es que el gato Silvestre, le estaba espiando y sabiendo que saldría por la ventana, se agazapó para esperar a que saliese, saltar y cogerle. Pero, lo que Silvestre no sabía es que la abuela estaba planchando en la habitación. Y se había dado cuenta de que el gato estaba escondido bajo la ventana. Así que cuando Piolín salió por la ventana, el gato Silvestre, saltó, lo agarró y a la vez la abuela cogió la plancha y le pegó tal planchazo al gato que este inmediatamente soltó al pobre pájaro y salió rodando césped hacia abajo hasta caer en el estanque al final del jardín.

12.2 Appendix 2 – Record of Extracts Analysed in the Preliminary Study

No.	File name	Language	Input	Length	Date recorded	Date analyzed	Proficiency level	Number of disfluencies
1	C_c_es_1_wmv	Spanish	photo	08.3 seconds	16/03/2012	23/03/2013	A2 low	17
2	i_r_esd_1_wmv	Spanish	dialogue	49 seconds	16/03/2012	28/03/2013	A2 low	13
3	b_k_esd_1_wmv	Spanish	dialogue	20 seconds	21/12/2011	04/04/2013	A2 low	3
4	k_c_es_1_wmv	Spanish	photo	11.7 seconds	18/12/2011	08/04/2013	A2	2
5	k_c_es_6_wmv	Spanish	photo	21.5 seconds	18/12/2011	08/04/2013	A2	5
6	k_c_es_2_wmv	Spanish	photo	42.2 seconds	18/12/2011	08/04/2013	A1	11
7	k_c_es_4_wmv	Spanish	photo	9.2 seconds	18/12/2011	10/04/2013	A1	5
8	k_c_es_5_wmv	Spanish	photo	07.6 seconds	18/12/2011	10/04/2013	A2	5
9	s_j_esd_1_wmv	Spanish	dialogue	07.6 seconds	18/12/2011	10/04/2013	A2 low	11
10	s_j_esd_2_wmv	Spanish	dialogue	10.1 seconds	18/12/2011	10/04/2013	A2 low	7
11	s_j_esd_1_wmv	Spanish	dialogue	9.8 seconds	18/12/2011	10/04/2013	A1	4
12	s_j_esd_4_wmv	Spanish	dialogue	2.1 seconds	18/12/2011	10/04/2013	A2 low	1
13	s_j_esd_5_wmv	Spanish	dialogue	09.1 seconds	18/12/2011	10/04/2013	A2 low	4
14	s_j_esd_6_wmv	Spanish	dialogue	31.1 seconds	18/12/2011	10/04/2013	A1	19
15	s_j_esd_7_wmv	Spanish	dialogue	8.8 seconds	18/12/2011	12/04/2013	A1	5
16	o_y_esd_2_wmv	Spanish	dialogue	38.8 seconds	18/12/2011	15/04/2013	A1	13
17	o_y_esd_3_wmv	Spanish	dialogue	13.1 seconds	18/12/2011	16/04/2013	A1 high	4
18	b_k_esd_2_wmv	Spanish	dialogue	19.2 seconds	18/12/2011	17/04/2013	A1	13
19	b_k_esd_5_wmv	Spanish	dialogue	17 seconds	18/12/2011	17/04/2013	A1	11
20	m_c_esd_1_wmv	Spanish	dialogue	9.3 seconds	08/06/2012	18/04/2013	native	5
21	m_c_esd_1_wmv	Spanish	dialogue	13.4 seconds	08/06/2012	18/04/2013	native	5
22	j_k_esd_1_wmv	Spanish	dialogue	7.5 seconds	16/03/2012	22/04/2013	A1 low	4
23	j_k_esd_2_wmv	Spanish	dialogue	6.3 seconds	16/03/2012	22/04/2013	A1 low	2
24	j_k_esd_3_wmv	Spanish	dialogue	7.9 seconds	16/03/2012	24/04/2013	A1 low	3
25	j_k_esd_4_wmv	Spanish	dialogue	7.9 seconds	16/03/2012	24/04/2013	A1 low	6
26	d_l_esd_1_wmv	Spanish	dialogue	24.7 seconds	16/03/2012	25/04/2013	A1 low	4
27	d_l_esd_2_wmv	Spanish	dialogue	24.7 seconds	16/03/2012	25/04/2013	A1 low	7

12.3 Appendix 3 - Transcription Annotation

The Jefferson Transcript Notation system was the base of the speech transcription used. However, as it was necessary to integrate information on disfluencies and gestures, we adapted her original annotation system to include gesture annotation (adapted from McNeill's lab annotation system) and disfluencies (adapted from Clark) as follows:

Symbol	Name	Use
[text]	Brackets	Indicates the start and end points of a gesture phrase.
text	Underlined text	Indicates where the gesture stroke takes place. This was only recorded in a number of cases as it is not key to this study.
(# of seconds)	Timed Pause	A number in parenthesis indicates the time, in seconds, of a pause in speech. This was only used in the transcripts for the preliminary study.
-	Hyphen	Indicates an abrupt halt or interruption in utterance, a cut-off.
ALL CAPS	Capitalized text	Indicates stressed speech (more emphatic). This was only recorded in a number of cases as it is not key to this study.
:	Colon	Indicates prolongation of sound, elongation or lengthening.
(*exhales*)		Audible exhalation or inhalation.
(*coughs*)		Cough
(*laughs*)		Laughter
(*other notes*)	Other notes which might refer to speech which is unclear or in doubt in the transcript.	

Notes

- Transcribed text did not use capital letters or punctuation (as these could be symbols used in annotation).
- Exact transcriptions were made of the utterances, including any errors speakers made.
- If there was a pronunciation error this was captured (whenever possible without making the utterance impossible to understand).
- Whenever translations are provided in the main text these try to keep the disfluency and the meaning of the utterance even if creating an unlikely utterance in the translated language.
- Volume, pitch and intonation were not noted.
- Speed of speech was not noted although the start and end times were given for utterances in the transcripts of the main study.
- Jefferson Transcript Notation, Jefferson (1984)

12.4 Appendix 4 – Gesture Categorization

Description of Gestures

- Imagistic:

    Iconic (gestures with iconicity): used to represent some possible concrete object or action.

    Metaphoric (gestures with metaphoricity): used to as a means of representing an image of an abstract idea.

- Non-imagistic:

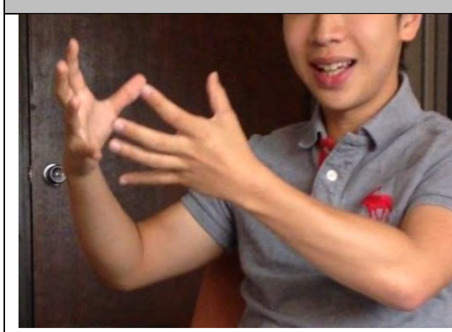
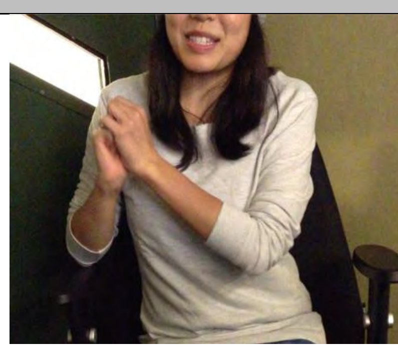
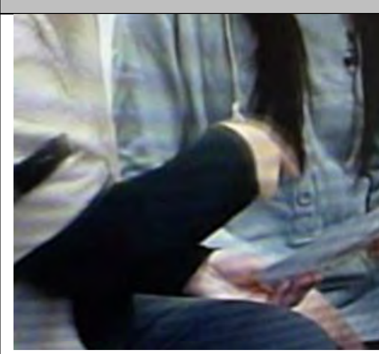
    Batons: no discernible meaning other than follow the rhythm or give emphasis.

    Deictic: pointing movements.

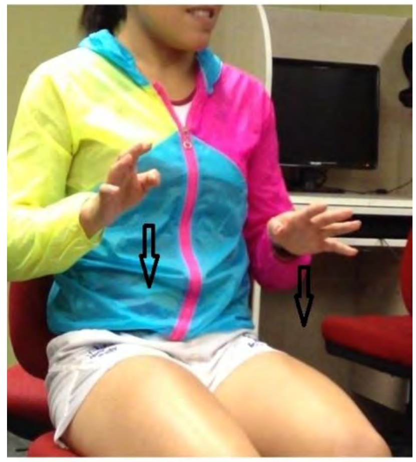
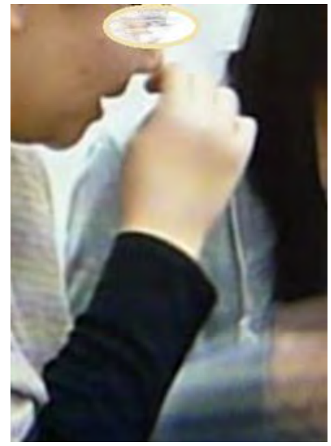
- Lexical: those used to aid word recovery.

- Adaptors: those used to attend a bodily need (scratch, adjust a piece of clothing, etc.)

Adapted from Kendon (2004, p. 100)

Gestures with metaphoricity	Gestures with iconicity	Gesture of a deictic nature
		
Hands indicate the container was full[123]	*Hands illustrate the action of whacking the cat with an umbrella.*	*Hand points at a photo.*

[123] For description purposes only: The photo is from our corpus but the speech co-occurring with it was not referring to a container.

Lexical gesture	Baton gesture	Adaptor gesture
		
The right hand circles above the left, in an action not related to the speech.	*Hands used to keep the rhythm.*	*Hand used to scratch the nose.*

12.5 Appendix 5 – Table of Acronyms Used in the Preliminary Study

	1	2	3	4	5	6	7	8	9
	Repair	**Error**	**Interruption**	**Repair**	**Gestures**	**Gesture on initiation of repair**	**Gesture as repair continues**	**Synchrony**	**Synchrony on resumption**
1	E = speech error	I = interruption	W = mid-word	C = carry on after repeat	i = iconic	a0 =none	b0 = on holdn	a = gesture stops first	s = synchronous
2	A = appropriateness	P = pause after word	A = after the word has been completed	S = substitute word	m = metaphoric	a1 = preparer	b1 = new gesture	d = gesture stops after speech	gbs = gesture before speech
3	O = Other	L= lengthening		B = correct a word/sound	d = deictic and b =baton	a2 = hold	b2 = continues	s = synchronous	sbg =speech before gesture
4		R = repeat		A = add	a = adaptor	a3 = stroke	b3 = retracted	l = new gesture	asy = asynchronous (speech and gesture not related)

	1	2	3	4	5	6	7	8	9
	Repair	**Error**	**Interruption**	**Repair**	**Gestures**	**Gesture on initiation of repair**	**Gesture as repair continues**	**Synchrony**	**Synchrony on resumption**
5		M = filler		R = restart	l = lexical	a4 = hold	b4 = repeated	0 = no gesture	ngs = no movement or no speech
6				St = stuck	u = unclear	a5 = retract	b5 = no visible gesture		
7					n = no visible gesture	a6 = whole phrase			

Notes:

- Ambiguous cases were not analysed (unless the ambiguity is in the type of gesture).
- If speech continues (as in fillers) and gesture also continues no note was made whether the gesture is synchronized with the repair.
- Adaptors were transcribed but were not analysed as gestures.

12.6 Appendix 6 – Example of Results from Extract of Conversation No. 1 in the Preliminary Study

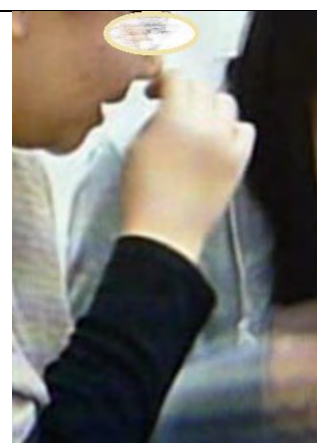
N°	File name	Language	Input	Length	Recorded	Analysed	No. disfluencies
1	C_c_so_1_wmv	Spanish	Photo	00:08.3 (8.3 seconds)	7/12/11	23/3/13	6

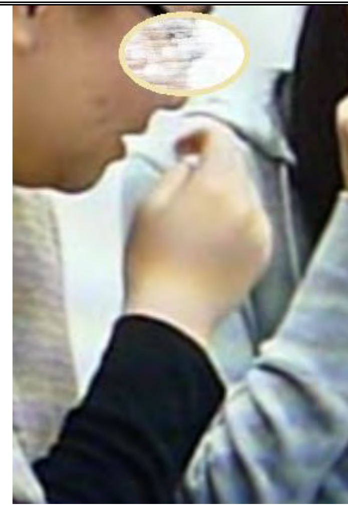
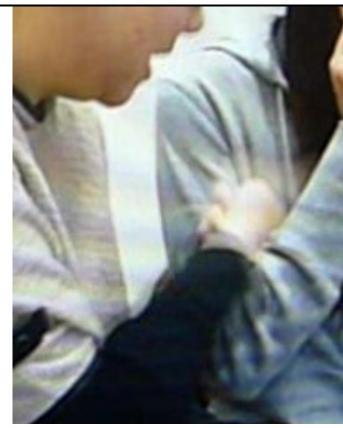
Participants are describing a photo (A4 size) of several male models.

A and B are seated side by side, bodies turned slightly towards each other.

They hold the photo with four, three or two hands.

Speaker	Line No.	Length of utterance seconds)	Transcript	1	2	3	4	5	6	7	8	9
A	1	01.4	[¿<u>pien</u>[:-	O	I, L	W	R	d	2	1	s	asyn
A	2	02.1	pien:-	O	R, L	W	R	l	4	4	na	sbg
A	3	02.8	[pien:-	O	R, L	W	R	l	1	3	s	s
A	4	03.5	pien<u>sas</u> [/ e:-	O	L		C	d	2	2	s	ngs
A	5	05.0	es-	O	I	W	C	d	2	0	s	ngs
A	6	05.6	tes hombres [son <u>gua</u>-/	O	I	W	C	d	2	0	s	sbg
A	7	07.6	pos?]]]]									

Line No.	Description of gesture	Photo
0	The interlocutors are seated side by side holding the photo. *A* holds the right side of the photo with her right hand.	
1	With the articulation of ‘pien:-‘ the right hand leaves the photo, it goes slightly up as a fist with the index finger pointing right and down, seemingly pointing at something in the photo (not clear from the recording). As the elongation of the phoneme /n/ commences the hand loses its shape with the index finger being retrieved into the fist and the hand moves towards the mouth/nose, interrupting the original gesture and becoming a new one.	
2	The hand is still by the mouth/nose in a lose fist, covering the mouth.	

3	The hand is still by the mouth/nose in a loose fist, covering the mouth but it begins to lower as if the speaker was ready to carry on with the speech.	
4	Before starting the repair the hand starts to drop forming a fist, with fingers closed towards the speaker and the index finger extended pointing to the right first, and then to B and finally towards the photo, the hand reaches its destination quickly, stopping and pointing to a point in the photo that coincides with the end of the word ‘piensas’. Still pointing the hand is retracted and moves to another point in the photo during the pause between ‘piensas’ and ‘e’ it start to point elsewhere within the photo coming closer to it and moving with the elongation of the phoneme / e /.	
5	The finger is already on a point in the photo when the word ‘estes’ begins to be articulated. The gesture is put on hold with the cut-off ‘es’.	

6	The gesture continues with the syllable 'tes'. The hand moves towards the left pointing at other men in the photo. With "son" the hand stops at a particular man and with 'guapos' it begins to move again towards the left to point at the addressee. The hand is still in a fist with the index finger extended towards the photo and B, the fingers pointing towards the speaker herself. With the cut-off at 'gua-' the gesture is put on hold.	
7	At the beginning of 'pos' the hand begins the retrieval phase, goes down and it is retracted.	

12.6.1 Analysis

A wants to ask something about the men in the photo, she has an idea that we see reflected in the relatively rapid onset of the hand movement releasing the photo and moving quickly to a point in the photo. But she encounters a conflict in her speech, producing a word cut-off after a lengthening and she is still not able to come up with the right words. The initial gesture is interrupted at its peak (the stroke) and is replaced by another (hand to mouth / nose - this could be an adaptor or an aid to lexical retrieval (as sometimes it is used when speakers are deep in thought) which takes place together with the lengthening. In this case the disfluency is synchronized with the gesture: the original gesture is put on hold with the cut off.

A tries twice more to externalize the speech, we can guess that the first try is not going to be successful by looking at her hand which doesn't move, almost as if the speaker knew her attempt would fail. However the second speech attempt is performed with what could be the continuation of the original gesture and the speech is uttered, however *A* is not content with it and both modalities are put on hold again. It seemed that the gesture was ready to externalize the idea but the speech wasn't.

Finally *A* resolves the speech conflict and even before articulation begins we can see the hand moving, fast and confident, first towards *B* (to ask her opinion) and then towards the photo. It is such a quick movement (pointing towards *B*) that she cannot interpret it to be a turn change. The confidence in the gesture suggests that *A* is ready to continue her speech. She does so but encounters new difficulties; the gesture again seems to be faster than speech so it is required to wait for the speech before pointing at the various men in the photograph. We can guess at the change in turn when *A* says 'guapos' and moves her hand towards *B*, indicating to her addressee that she has to answer the question. As the speech is slower than the gesture we see a pause between the two syllables which is reflected in a hold of the gesture. When the turn is completed the hand retrieves confirming the turn change.

12.6.1.1 Gesture phrases

We can see five different phrases in this utterance, as the gestures merge one into the next without the hand going back to the rest position in between, these are nested within each other (Kendon, 2004, pp. 111-112). We suggest three Growth Points (GP) are present: phrases 1 and 3 probably share the same GP; phrase 2 is harder to confirm as a GP, as it could be an adaptor; phrase 4 is another and 5 the last one.

[¿<u>pien</u>[124]	**[:- pien:-**	**[pien:- pien<u>sas</u>**	**[/ e:- es-<u>tes</u> hombres**	**[son <u>gua-</u>/pos?]]]]**
Phrase 1	Phrase 2 (adaptor?)	Phrase 3	Phrase 4	Phrase 5
[[125]do you thi	[:- th:-	[th:- think	[/the:- these men	[are good look-/ing

[124] All transcriptions capture all errors (including phonetical errors when these could be transcribe without making comprehension impossible.

[125] Translations try to capture the sense of the disfluencies, although these might be unlikely disfluencies in English, with cut-offs or elongations occurring at unlikely places.

At the end of this gestural unit (end of phrase 5) the hand goes back to the rest position.

Phrase 4 can be divided into the following phases:

[/ e:-	**es-/**	**<u>tes</u>**	**hombres**
The index finger starts to move during the pause to the point on the photo	*Gesture on hold*	*The index finger moves over the photo from left to right pointing to various items on the photo.*	*Gesture on hold*
Preparation	Pre-stroke hold	stroke	Post stroke hold

Full gesture transcription will indicate when the preparation and holds occur, as this information is not necessary for the analysis we undertook in most transcription we have only marked where the stroke takes place.

12.7 Appendix 7 – Table of Participants in the Main Study

									English			Spanish			L1
	sex	Mother tongue	Second Language	Name in study	Age	Occupation	Reflection Learning preference	Issues L2 proficiency*	written	audio	video	written	audio	video	Debriefing
1	f	Chinese	English	---	40+	professional	visual		x	x	x				
2	f	Chinese	English IELTS 8	HK2	under 25	Student	visual	A2 (Spanish)	x	x	x			x	x
3	m	Spanish	English	---	under 25	student	visual	Unfamiliar with researcher			x	x	x	x	
4	f	Chinese	English	---	under 25	student	visual		x	x	x				x
5	m	Dutch	English	---	under 25	student	visual	Unfamiliar with researcher	x	x	x			x	
6	m	Chinese	English	---	25-39	professional	words	Unfamiliar with researcher	x	x	x				
7	f	Chinese	English	---	under 25	student	words	Unfamiliar with researcher	x	x	x				x
8	m	Spanish	English	---	25-39	professional	visual	cold	x	x	x			x	
9	f	English	English	---	40+	professional	words	cold	x	x	x			x	
10	m	English	English	---	40+	professional	?	cold	x	x	x				
11	f	Chinese	English	---	40+	professional	?	cold	x	x	x				
12	f	Chinese	English	---	25-39	professional	visual		x	x	x				

									English			Spanish			L1
	sex	**Mother tongue**	**Second Language**	**Name in study**	**Age**	**Occupation**	**Reflection** **Learning preference**	**Issues** **L2 proficiency***	**written**	**audio**	**video**	**written**	**audio**	**video**	**Debriefing**
13	f	Spanish	English	---	under 25	student	word		x	x	x				x
14	f	Chinese	English IELTS 8	HK1	under 25	Student	Not done	A1/A2 (Spanish)	x	x	x			x	x
15	m	Chinese	English IELTS 8.5	HK3	under 25	Student	Not done	A1/A2 (Spanish)	x	x	x			x	x
16	f	Spanish	English	Sp1	under 25	student	Not done	Unfamiliar with researcher B1/B2 (English)			x	x	x	x	x
17	f	Spanish	English	Sp2	under 25	student	Results not known	Unfamiliar with researcher (B2 English)			x	x	x	x	x
18	m	Spanish	English	Sp3	under 25	student	Results not known	Unfamiliar with researcher (B2 English)			x	x	x	x	x
19	m	Spanish	English	Sp4	under 25	student	Results not known	Unfamiliar with researcher (B2 English)			x	x	x	x	x

									English			Spanish			L1
	sex	**Mother tongue**	**Second Language**	**Name in study**	**Age**	**Occupation**	**Reflection** **Learning preference**	**Issues** **L2 proficiency***	**written**	**audio**	**video**	**written**	**audio**	**video**	**Debriefing**
20	m	Spanish	English	Sp5	under 25	student	Results not known	Unfamiliar with researcher			x	x	x	x	x
21	m	Spanish	English	Sp6	under 25	student	Results not known	Unfamiliar with researcher			x	x	x	x	x
22	f	Spanish	English	Sp7	25-39	student	Results not known	Unfamiliar with researcher			x	x	x	x	x
23	f	Spanish	English	Sp8	under 25	student	Results not known	Unfamiliar with researcher			x	x	x	x	x
24	f	Spanish	English	Sp9	under 25	Student, mildly autistic	Results not known	Unfamiliar with researcher			x	x	x	x	x
25	f	Spanish	English	Sp10	under 25	student	Results not known	Unfamiliar with researcher			x	x	x	x	x
26	f	Spanish	English	Sp11	under 25	student	Results not known	Unfamiliar with researcher			x	x	x	x	x
27	f	Spanish	English	Sp12	under 25	student	Results not known	Unfamiliar with researcher			x	x	x	x	x

									English			Spanish			L1
	sex	**Mother tongue**	**Second Language**	**Name in study**	**Age**	**Occupation**	**Reflection** **Learning preference**	**Issues** **L2 proficiency***	**written**	**audio**	**video**	**written**	**audio**	**video**	**Debriefing**
28	m	Chinese	English IELTS 8	HK4	under 25	Student	Results not known	A1/A2 (Spanish)	x	x	x			x	x

* HK participants' proficiency based on the curriculum and tests done in class. Spanish natives' proficiency based on a conversation carried out prior to the experiment.

12.8 Appendix 8 – Consent Forms and Ethics Committee Approval Confirmation

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Information Sheet

Use of verbal and non verbal language in cases of self repair

(I) Information Sheet

1. Basic Information of the Project

1.1 Title of the study

Use of verbal and non verbal language in cases of self repair

1.2 Purpose of the study

The aim of the study is to find whether there is a link between the gesture and the speech in cases of self-repair (when there are pauses, repetitions, hesitations, etc, in the speech). If a link is found the nature of this link will be studied further by analyzing the type of gesture and of self-repair.

In particular this study will focus on second language speakers, to identify means of aiding second or foreign language teaching. It is expected that a gesture-speech link will be found. A clear understanding of the gesture might give an insight of the cognitive conflicts the speaker might be experiencing and thus an experienced teacher might be able to help solve the conflict.

1.3 Description of procedures, purpose and length of time required

Subjects will be asked to sit down and talk in their second/foreign language. The topic of the talk will vary on the proficiency of the speaker. They will either be asked to recount three stories (one read, one heard and one watched) or to describe a photo or just to talk to a friend about a common topic. In this last case the sessions will be done with pairs of subjects, otherwise it will be the subject and the researcher only.

Each session will last about 20 minutes. The students will be told the general aim of the study, what they are required to do and then asked to give their consent and sign the consent form.

The recordings will then be transcribed looking for cases of self repair and a transcription of the gesture will be added to these. These cases will be compared looking for patterns.

The purpose of the study is as mentioned above.

2. Involvement of the Subject

2.1 Why the subject was selected

Volunteers will be asked if they wish to participate and have time to do so. Anyone willing to participate who speaks a second or foreign language will be invited to participate.

2.2 How will the subject be involved

See point 1.3

2.3 Discomforts, inconveniences expected

None

2.4 Benefits and / or risks, if any

Subjects will get the change to practice their second/foreign language

2.5 Standard treatment or alternative treatment that may be withheld

Not relevant

2.6 Compensation to be expected, if any

None

3. Confidentiality Issues

3.1 Classes of persons to whom the subject's personal data may be transferred to

Not applicable

3.2 How confidentiality, anonymity and privacy will be maintained

Subject's personal information will not be recorded

Researcher will identify subjects by codes

Any video to be used in presentations will be edited to ensure the subjects can't be recognized.

All data will be kept in the PolyU computer or lap-top of the researcher

4. Rights of the Subject

4.1 Right to refuse to participate or withdraw at any time for any reason without penalty of any kind

The subject will have the right to refuse to participate or withdraw at any time for any reason without penalty of any kind

4.2 Right to request access to and correction of the personal data supplied for the project

The subject will have the right to request access to the data supplied for the project (including copies of the video recordings)

4.3 Sources for information and assurances that the researcher will provide further and ongoing information (e.g., name and contact phone no. of the researcher)

The subject will be given the contact details of the researcher.

Please see a copy of the Information Sheet to be given to subjects before they sign their consent. The above information will also be available to them if requested.

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

INFORMATION SHEET

Use of verbal and non verbal language in cases of self repair

You are invited to participate on a study conducted by Ms. Lopez, who is a staff member of the Department of English in The Hong Kong Polytechnic University. The project has been approved by the Human Subjects Ethics Sub-committee (HSESC) (or its Delegate) of The Hong Kong Polytechnic University (HSESC Reference Number: HSEARS20131213001).

The aim of this study is to identify whether there is a link between verbal and non verbal communication. The study will involve retelling a story that you will read, one that you will hear and one that you will see on video or describing a photo or talking to a friend about a topic you have in common. You will be video recorded (sound and audio) and this recording will be analyzed. It is hoped that this information will help to understand the link between verbal and non-verbal communication to aid teachers of second or foreign languages. Further information will be provided upon request.

The testing should not result in any discomfort, but you will be videotaped. All information related to you will remain confidential, and will be identifiable by codes only known to the researcher.

You have every right to withdrawn from the study at any point without penalty of any kind. The session will last about 20 minutes.

If you would like to get more information about this study, please contact Ms. Lopez on renia.lopez@polyu.edu.hk or at AG433.

If you have any complaints about the conduct of this research study, please do not hesitate to contact Dr Virginia Cheng, Secretary of the Human Subjects Ethics Sub-Committee of The Hong Kong Polytechnic University in writing (c/o Research Office of the University) stating clearly the responsible person and department of this study.

Thank you for your interest in participating in this study.

Ms.Lopez

Principal Investigator

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

CONSENT TO PARTICIPATE IN RESEARCH

Use of verbal and non verbal language in cases of self repair

I ________________________ hereby consent to participate in the captioned research conducted by Ms. Lopez of the Department of English Hong Kong Polytechnic University`.

I understand that information obtained from this research may be used in future research and published. However, my right to privacy will be retained, i.e. my personal details will not be revealed.

The procedure as set out in the attached information sheet has been fully explained. I understand the benefit and risks involved. My participation in the project is voluntary.

I acknowledge that I have the right to question any part of the procedure and can withdraw at any time without penalty of any kind.

Name of participant ____

Signature of participant ____

Name of researcher Ms. Renia Lopez (renia.lopez@polyu.edu.hk; Office AG433)

Signature of researcher ____

Date ____

Hung Hom Kowloon Hong Kong 香港 九龍 紅磡
Tel 電話 (852) 2766 5111 Fax 傳真 (852) 2784 3374
Email 電郵 polyu@polyu.edu.hk
Website 網址 www.polyu.edu.hk

12.9 Appendix 9 – Detailed Case Study for C.S. 2HK - Written Input.

Participant 2HK volunteered to help in this study and allowed us to record her. She was a student of Spanish as a foreign language at the Hong Kong Polytechnic University. She had spent two months in Spain studying Spanish. Her mother tongue was Cantonese and her English was fluent, having achieved an 8 in the IELTS (English proficiency) exam. Her proficiency level in Spanish was A2.

The input in this experiment was the text with the story of *The lion and the mouse*. This is the transcription without annotations or pauses:[126]

there is a lion and a mouse and: the lion was always despising the the mouse because he thinks that he's too small and not being able to help or anything like that and then one day um he caugh- the lion caught the mouse and was about to kill him and just eat him and then the mouse said um if you please just leave me alo- uh just let me go this once and I'll be sure to repay you in the future then the lion let him go and then uh there is one instance that the lion was caught by a hunter and then he was bound by this rope and everything and then the mouse recognized his roar and came by and used his small sharp teeth and grind the rope and saved and released the the lion and saved him and then the mouse finally said see so now even a small mouse like me could confer benefits to you the lion

Duration	Number of words	Speed of speech	Gestural Phrases
62.7 seconds	168 words including fillers	2.68 words per second (161 words per minute)	15 0.25 phrases per second 0.09 phrases per word

[126] For the full annotated transcription refer to Dropbox: https://www.dropbox.com/sh/yuljvsf40kta510/AAB_IFWFN3OgyigyNIPBYVpMa?dl=0.

12.9.1 Disfluencies

The initial manual analysis identified 20 disfluencies, when the *Praat* pause analysis was run, it recorded 29 pauses. In five instances it was necessary to adjust the pause length as they were shorter than the waveform and intensity-pitch graph indicated. The average pause was 0.358 seconds; the shortest recorded one was 0.024 seconds and the longest 0.695 seconds. In addition there were also two other disruptions that were not marked with a pause. When we looked at pure pauses, with no other audible disruption suggesting covert disfluencies, in the manual analysis we found 12, and *Praat* identified 13.

Initially the discrepancy was overlooked – explained by a human inability to perceive pauses with the same degree of accuracy as the software – only after analysing the video input transcription it was realised that further information could be gathered from seemingly pure pauses that hid instead another filler: 'and then'. Thus the pauses were re-analysed to clean the data further. Out of the 29 pauses that were recorded by Praat there were 13 cases of pure pauses, with an average length of 0.404 seconds – this included 4 cases where 'and' followed a pause. The table below details the results:

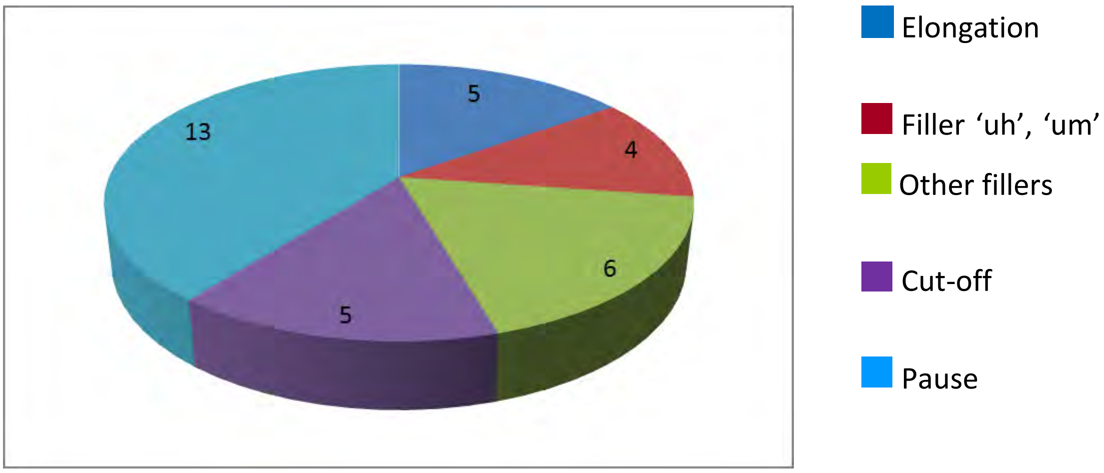
Table 46 Appendix. - Case study 2HK (written input): Data on all disfluencies.

Frequency	Average length (seconds)	Description
29	0.358	All pauses
13	0.404	Pure pauses between seemingly fluent speech: continuation (including 'and')
0	na	before filler 'uh'
3	0.303	before filler 'um'
0	na	after filler 'uh'
1	0.57	after filler 'um'
4	0.444	before 'and then'
1	0.272	after 'and then'
4	0.153	after cut-off
4	0.42	after elongation

Note: There were no non-reductions noted.

The breakdown by disfluency is shown below in Figure 1 Appendix.

Figure 56 Appendix. - Case study 2HK (written input): Number of disfluencies.



12.9.2 Fillers

There were only four fillers ‘uh’ and ‘um’ initially identified in this text, after realizing that ‘and then’ was also being used as a filer the total number rose to 10 (see Table below).

Table 47 Appendix. - Case study 2HK (written input): Data on fillers.

	uh		um		and then	
Total*	1		3		6	
Preceded by a pause (%)	0	0%	3	100%	4	67%
Average pause duration	na		0.303		0.444	
Followed by a pause (%)	0	0%	1	33%	1	17%
Average pause duration	na		0.573		0.272	
Neither followed nor proceeded by a pause (%)	1	100%	0	0%	2	33%

* Some fillers were both preceded and followed by a pause.

12.9.3 Word cut-offs

The next step was to analyse the interruptions or cut-offs. Cut-offs are of two types, after a word has been completed or mid-word. Speech preceding a repetition was considered a cut-off where the word had been completed.

Table 48 Appendix. - Case study 2HK (written input): Data on cut-offs.

Cut off	Within word		After word completion	
	Not followed by a pause	Followed by a pause	Not followed by a pause	Followed by a pause
Number	1	1	0	3
Average length of pause (seconds)		0.226		0.13

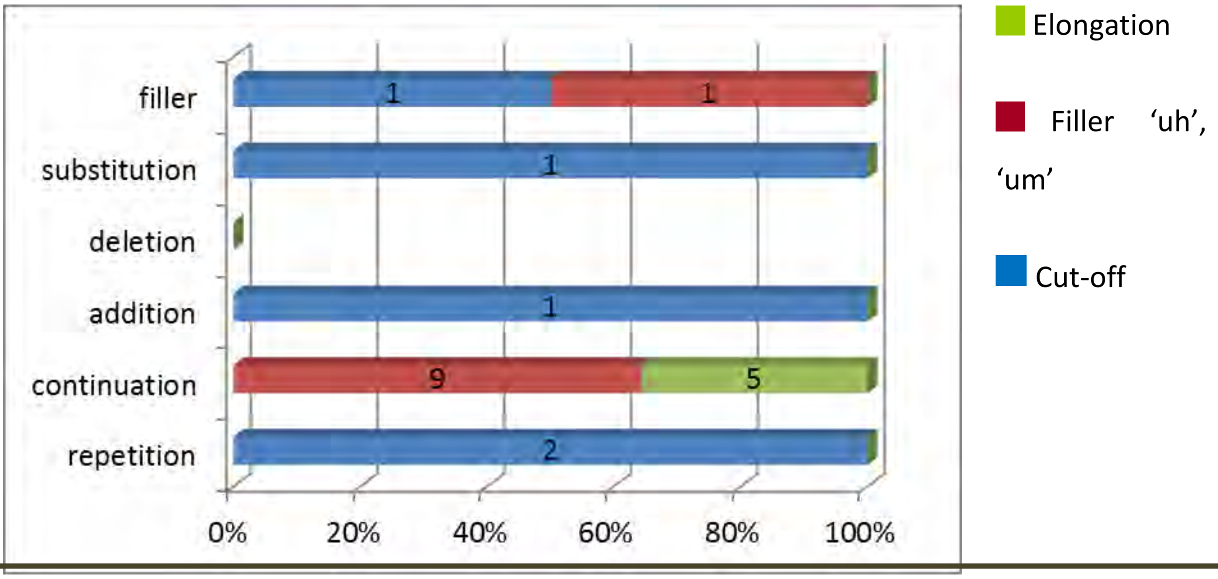
12.9.4 Elongations

There were six elongations, five of them followed by a pause averaging 0.353 seconds.

12.9.5 Resumption

The type of resumption was annotated after overt disfluencies. Pure pauses were taken as covert disfluencies (see Figure below).

Figure 57 Appendix. – Case study 2HK (written input): Types of resumptions after overt disfluencies.



12.9.6 Gestures

This was the first text recollected by the speaker, based on the written text. The full gestural phrases transcript can be found in *Dropbox*.[127] The most relevant points are discussed below. The speaker was sitting in a chair with armrests but did not use them. The rest position was hands on her lap, palms up with the left hand cupping the right one.

There were 15 gestural phrases. Not much additional information seems to be transmitted through the gestures, most are slight gestures that seem to just stress particular information, newsworthy information. There is however one observation worth making pertaining anaphoric gestures, these are the ones referring back to other ideas or objects already mentioned. In this text the speaker starts by introducing the two main characters:

[127] Link to the Dropbox file: https://www.dropbox.com/sh/yuljvsf40kta510/AAB_IFWFN3OgyigyNIPBYVpMa?dl=0.

Example – C.S. 2HK written input (see Figure 5)

Line No.	Start (seconds)	End (seconds)	Duration (seconds)	Speech Transcription	Gesture Description
1	2.122	4.635	2.513	there is [a <u>li</u>on] and [a <u>mou</u>se]	*Left hand (LH) up and down; right hand (RH) up and down, in both cases palms open facing in.*

The lion, is thus represented for most of the remaining text by the left hand, and the mouse by the right hand or by the space on the right side.

In English the gesture stroke often coincides with verbs. In this case we observed that 15 strokes fell within a verb and 13 within other words (if the stroke covered both a verb and non-verb it was counted in both categories).

When analysing the disfluencies and the gestures we found only one filler (we include as fillers: 'uh', 'um' and 'and then') falling within a gestural phrase; a filler 'uh' which in this case follows a cut-off within a word. The gesture is placed on hold following the cut-off, and the filler is uttered with a held gesture. There are other cases of gestures during cut-offs (a total of 4 cases out of the 5 cut-offs recorded in this text). In two other cut-off cases the hold led to a new stroke and in one case it just detained it and was consequently resumed.

There is an exceptional case recorded where a cut-off is followed by a glottal sound. It is the only case in which we observed a gesture continuing during a disfluency or a pause.

Example – C.S. 2HK written input.

Line No.	Start (seconds)	End (seconds)	Duration (seconds)	Speech Transcription	Gesture Description
17	17.382	17.844	0.462	[he caugh- *(glottal sound)*	*LH comes up and with the end of 'caugh-', a glottal sound is produced during which the hand continues to move up*
18	17.844	18.070	0.226		
19	18.070	21.554	3.483	the lion <u>caught</u> [the <u>mous</u>e [and was about to <u>kill</u> him and just eat <u>him</u>]]]	*hand comes down with 'lion', the hand into a loose cup, palm facing up, with 'caught' it comes down, adding stress, with 'mouse' the hand moves to the right, as above, to indicate the position of the mouse, and then it comes up and to the left, fingertips down and touching, and slightly down on 'kill' and 'eat', to stress*

12.9.7 Analysis

By moving the left hand, also the one raised when introducing the lion at the beginning of the story, it seems that the speaker is trying to make clear the subject of the sentence, which might not have been clear to the addressee. On beginning the gesture it seems that the speaker realizes that the speech reference is not clear and tries to add it before concluding the word 'caught' resulting in a glottal sound that is unintelligible. In this case it could be argued that the gesture might have caused the disfluency by adding information the speech modality did not consider clear. The presence of the continuing gesture could be explained if this instance is classified as an overlap of the words, ‘caught’ and ‘the’, and so continuous speech, instead of a cut-off followed by a glottal sound.

12.10 Appendix 10 Mortadelo y Filemon (Example of Gesture Exercise)

Mortadelo y Filemón: gestos y expresiones coloquiales
ENGL356

Relaciona las expresiones en inglés con las expresiones en español y encuentra una viñeta donde se podrían usar.

Match the English expressions with the Spanish ones and find a cartoon where they could be used.

	But, but…how?	Estooooooo…..
	Down with him!	¡Use la cabeza!
	Hey! You!	Eso / Vale
	I did it!	¡Eh, tú!
	I fear problems! I'm not sure about this!	Las vamos a pasar canutas
	I have an idea!	¡Mierda!
	It's not a big deal!	¡Psché! Eso no es nada
	Oh shit!	¡Soy un hacha!
	Oh, I see!	¡Lo conseguí!
	Ops, sorry!	¡Abajo con él!
	Sorted!	¿Y cómo lo hago?
	Stupid man!	¡Ah! Hombre, haberlo dicho
	That's it!	¡Será tonto!
	That's nothing!	¡Espera un momento!
	Think!	¡Vamos, no es para tanto!
	Wait, wait!	¡Tengo una idea!

Students are given the following cartoons without text.

Mortadelo y Filemón. ¡El dopaje qué potaje! (Ibáñez, 2007).