

Faculté de philosophie, arts et lettres

Gestures and Speech

The role of learners' gestures in second language acquisition

Auteur : Marie WOUTERS

Promoteur : Prof. Dr. Gaëtanelle GILQUIN

Année académique 2019-2020

Master 120 en langues et lettres modernes, orientation germanique, finalité didactique

Acknowledgments

I would like to express my gratitude to all the people who helped me with the completion of this thesis.

First and foremost, I would like to thank my supervisor, Professor Doctor Gaëtanelle Gilquin, for her support, her time, and her useful and always relevant feedback.

I also wish to thank the following people for helping me finalise this thesis by providing a fresh insight to my thesis and helping me with some statistical analyses: Dominique Warnant, Johan Wouters, Thomas Kintziger and Eloïse Schacht.

Finally, I wish to express my gratitude to my family and my friends for always believing in me and being a source of motivation when I needed it.

Table of contents

Acknowledgments	i
List of tables	iv
List of figures	v
Introduction	1
1. Gestures in communication	3
1.1. Definition of gestures	3
1.2. Classification of gestures.....	7
1.3. Universality of gesturing	11
1.4. Functions of gesturing	13
1.5. Connection between gestures and speech.....	15
2. L2 acquisition	19
2.1. Terminology of language acquisition.....	19
2.2. L1 vs. L2 acquisition processes.....	24
2.3. Language acquisition and gesturing	26
2.3.1. Gesturing in L1.....	26
2.3.2. Gesturing in L2.....	28
2.3.3. Gesturing in the case of bilinguals	30
3. Corpus linguistics	32
3.1. Corpus: definition.....	32
3.2. Corpus linguistics: utility and status.....	33
3.3. Type of data in corpora	36
3.4. A specific case: multimodal corpora	37
3.4.1. Multimodal corpora: some key elements	37
3.4.2. Annotation issues in multimodal corpora.....	39
3.4.3. Annotating gestures in multimodal corpora	41
4. Data and methodology	45
4.1. Research questions and hypotheses.....	45
4.2. The corpus: NESSI.....	47
4.3. The software: ELAN	48
4.4. The selection of data.....	48
4.5. The annotation of data	51

5. Results	61
5.1. Frequency per proficiency level	61
5.2. Gesture types per proficiency level	68
5.3. Relation between gestures and speech	74
5.4. Functions of gestures and frequent patterns	81
Conclusion	91
Bibliography	95

List of tables

Table 1: Conversion from mean score to CEFR proficiency level (based on Dumont 2018: 384). 50	
Table 2: Selected participants' mean score and CEFR level. 51	
Table 3: Difference between self-touching movement and self-adaptor. 52	
Table 4: Difference between resting position and hold..... 56	
Table 5: Ratio between the total gesturing time and the total number of gestures per participant during their speaking turn. 66	
Table 6: Functions of the gesture types observed in this study..... 89	

List of figures

Figure 1: Example of self-touching movement (from participant HK009).....	53
Figure 2: Example of self-adaptor (from participant HK026).	54
Figure 3: Example of resting position (from participant HK015).....	56
Figure 4: Example of hold (from participant HK015).....	56
Figure 5a: Example of [be.1] (from participant HK021).....	59
Figure 5b: Example of [be./rp.1] (from participant HK021).	60
Figure 5c: Example of [be.1/rp.2] (from participant HK021).....	60
Figure 6a: Total number of gestures per participant.	62
Figure 6b: Mean number of gestures per proficiency level.	63
Figure 7a: Percentage of gesturing time per participant.	64
Figure 7b: Mean percentage of gesturing time per proficiency level.	65
Figure 8: Mean ratio per proficiency level	67
Figure 9: Mean duration per gesture type.	68
Figure 10: Proportion of gesture types per participant.....	70
Figure 11: Percentage of gesture types per participant.	71
Figure 12: Mean percentage of frequent gesture types per proficiency level.....	73
Figure 13a: Percentage of gestures-speech matching per participant.	75
Figure 13b: Mean percentage of gestures-speech matching per proficiency level.	76
Figure 14a: Percentage of hold without or with speech per participant.	77
Figure 14b: Mean percentage of hold with or without speech per participant.....	78
Figure 15a: Percentage of gesture types in relation to speech per participant.	79
Figure 15b: Focus on the most used gesture types in relation to speech.	79
Figure 16: Example of beat to punctuate a self-correction (from participant HK037).....	82
Figure 17: Example of gestures and speech repetition ended with a beat (from participant HK034).	83
Figure 18: Example of the limitation for deictic gestures (from participant HK034).....	86
Figure 19: Example of affect display (from participant HK021).....	88

Introduction

Communication goes beyond words. Anyone who has already been in contact with people speaking another language can testify to it. Think for instance of the numerous young people taking part in Erasmus exchange programs: despite a lack of common language due to the various nationalities represented, or sometimes sharing but a few basic words in one common language (often English), it is striking to see how fast they can find ways to communicate and share unforgettable moments together. This example brings to light an interesting element: communication, in this very case, indeed seems to be achievable, but it cannot be performed through speech, or at least not only through speech, as these people do not share the same (spoken) language. Therefore, a hypothesis could be that these people's communication largely occurs via what is called non-verbal communication, including gestures.

Although people tend to mainly—or even only—refer to speech when they think of oral communication, it has actually been shown by researchers like Mehrabian (1972) that this axis only constitutes a small part of the whole communication system. Conversely, body language is argued to rule more than half of the communication. Goldin-Meadow and Alibali (2013: 2) refer more specifically to gestures, which are part of this non-verbal system, as being “integral” to communication. A further interesting aspect of gestures that they point out is that gestures might “serve as a window onto [speaker's] thoughts”, enabling researchers to potentially gain better insight into the processes of thinking and speaking (Goldin-Meadow and Alibali 2013: 18). Many researchers have also acknowledged the essential character of gestures in different areas and some studies have already established the impact they could have on the process of speaking as well as on learning a new language. If, as suggested, gestures are linked to people's speaking, thinking and learning processes and have a central place in the oral communication system, analyses in the field of second language acquisition might benefit from studying their use and their role in the learners' language production. A lot of studies have already been carried out around gestures but most of them focus on gestures in a native (also called L1) context. Therefore, the gestures performed by L2-learners when they have to produce speech and the role these gestures could play in the second language acquisition process is yet to be fully understood.

This thesis analyses the use of gestures, here defined as hand movements produced by the learners during their speaking turn and which appear to influence their communication. The aim of this thesis is to investigate if this use varies depending on the learners' proficiency level and how gesturing potentially impacts the learners' speech production throughout their SLA process. It is believed that gestures might influence the process of second language acquisition. Gestures are expected to play a prominent role in the production of L2 speech in two major ways: supporting and facilitating the thinking process leading to speech production on the one hand and influencing the communicative impact of the produced speech on the other hand. In order to test these hypotheses and to learn more about the use of learners' gestures in relation to their L2 acquisition process, this thesis will be organised as follows. The first part will introduce theories and findings from the literature dealing with different aspects of gestures, of language acquisition and also of corpus linguistics. The current study is indeed corpus-based: this type of analysis, carried out in a pseudo-longitudinal perspective, appeared to be the most relevant and efficient way to compare the evolution of gestures on the basis of the proficiency level achieved in the target language because it allows the researcher to immediately gain insight into the learners' use of gestures across the levels. Moreover, it avoids drawing conclusions which would only be based on the use of one individual. The use of gestures of two participants was observed for each proficiency level from B1 to C2, resulting in the analysis of eight interview excerpts. This corpus-based analysis will be developed in the second part. In this second part, three specific research questions will be addressed to frame the analysis and hypotheses will be constructed. Then, some information will be given about the corpus and the software which have been used to conduct the analysis and a description will be given of the data collection and annotation procedures which were adopted. Both quantitative and qualitative results will then be presented in the third part of this thesis, the former analysing the frequency and types of gestures performed within and between the various proficiency levels as well as their relation to speech, and the latter interpreting the broader context in which the gestures were set and especially the relation they tended to have with the produced speech and speech production. Finally, a conclusion will wrap-up this study. It is expected that this conclusion might lead to further questioning in the field of gestures and language acquisition, but also to a better understanding of how learners use gestures when they have to speak in a L2.

1. Gestures in communication

Before carrying out the present analysis, it is relevant to gain insight into what is currently known about gestures. Gestures have already been studied by various researchers and in many distinct fields. Kendon (1983: 154) traces back the first mention of gestures as an important topic to the Antiquity, with Quintilian's rhetoric. However, the term *gestures* can be misleading, like other a priori common terms whenever they are used for academic purposes. Gestures can indeed be used in a number of fields and, depending on the features of gestures which are relevant to each area, the core definition of gestures will automatically vary, sometimes generating misinterpretations if the term is initially not understood as it was aimed to be. Therefore, it is crucial to offer a concise and clear definition of what will be understood by the term *gestures* in the current analysis. In order to do this, it is necessary to compare what has been said and found about gestures in diverse areas of expertise and to select what could be useful for the present study, analysing the use of gestures made by learners when they have to perform speech in a L2. This chapter will thus clarify what will be considered as a gesture in this study by bringing together existing gesture definitions and classifications and will develop what is said in the literature about the universality of gestures, their functions and their relation to speech.

1.1. Definition of gestures

Defining gestures might seem obvious because everybody gestures in their everyday life. However, researchers who have considered the question and tried to gain valuable insight into how gestures work, where they come from and what they are used for, tend to give contrasting definitions. What appears to be common, however, is that most researchers studying gestures relate them to the concept of communication. As Olivier (2014: 2) summarises it, communication is established along three main axes, namely what can be considered as *verbal* communication (the spoken language with words and sentences), *paraverbal* communication (the voice and the way people transmit their words depending on features like speed or volume) and finally *nonverbal* communication (everything that influences the communication except speech). Gestures are part of this latter axis. Hişmanoglu and Hişmanoglu (2008) go further as they develop a hierarchy between nonverbal communication, gestures and body language, which are often wrongly used as perfect synonyms. For them, nonverbal communication is indeed a whole entity consisting of everything that is used to communicate, besides the words. In this entity, there is, next to aspects such as posture or oculusics (the movement of the eyes), what is considered as body language. Body language is what is communicated through the

body, be it the gaze, the occupation of the space or also the movements of “a part of the body, used instead of or in combination with verbal communication” (Hişmanoglu and Hişmanoglu 2008: 167). This latter component of body language is referred to as gestures (also called *kinesics*). Gestures belong therefore to the body language category, which is part of the whole set of nonverbal communication. This distinction being made, one could wonder if gestures need to be performed with a particular part of the body, and also what specific role they play in communication.

Not all body movements are considered as gestures, which implies that researchers should specify which part of the body is concerned and therefore observed in their studies. The definition given by Hişmanoglu and Hişmanoglu (2008: 168) goes as follows: “movements of the head, hands, arms, and legs [which] can be employed to convey specific messages”. Yet, many researchers stress the importance of one part of the body in particular: the hands. Gullberg (1998: 43) indeed explains that she focusses on “hand and/or arm movements”, because it “narrows the scope of the study considerably” and therefore allows for a more detailed and precise analysis. Moreover, Goldin-Meadow and Alibali (2013: 1) put forward the argument that hand movements are ubiquitous in conversations, which justifies to focus on them. A third justification could be traced back to Quintilian, quoted by Kendon (2004: 18) as one of the earliest references to gestures. Quintilian suggests that *gestus* (Latin word for gestures) can be made by various parts of the body, but he gives hand movements a distinct status: “For other portions of the body merely help the speaker, whereas the hands may be almost said to speak” (Kendon 2004: 18). This is why only hand (and arm) movements will be considered as gestures in this study: they seem to be strongly associated to language and speech, they are found in all conversations and limiting the observation on them makes possible to focus on more details and to better analyse their use for learners. Gestures, in this thesis, will systematically be movements from the shoulders to the fingers.

As previously mentioned, gestures are also part of the (nonverbal) communication system, which is why developing the communicative aspects attributed to gestures seems essential. Basing his research on the literature, Gregersen (2007: 52-53) establishes two defining principles to characterise nonverbal communication: it needs to be intentional and interpersonal. Assuming that gestures are a part of nonverbal communication would therefore imply that the body movements need to be both intentional and interpersonal in order to be included in the study. The definition of gestures by Nicoladis et al. (2018: 648) supports the intentional principle as they define gestures as “deliberate hand movements”. However, Beattie

(2003) establishes gesture categories based on the degree to which each gesture is consciously or unconsciously performed, which challenges the intentional pseudo-necessity previously suggested. The interpersonal principle can also be questioned: Goring (2016: 2) acknowledges that “many types of gestures seem to benefit both the listener and the speaker”, and adds that “some types of co-speech gesturing seem to specifically benefit only one side of the communication process”, stressing the fact that gestures sometimes carry meaning for the speaker only. Yet, gestures communicativeness is not denied. Indeed, the unintentionality of performing a gesture, does not automatically impede communicative meaningfulness. Many gestures are actually unconsciously performed, either by habit or as a reflex, but with a direct connection to what is being said. Also, people can unknowingly transmit information through their unintentional gestures. Constantly touching one’s ear or fingers while talking might give information to the interlocutor about one’s emotional state, and eventually, in certain contexts, about the veracity of some statements: these gestures may thus influence the conversation. In the same way, gestures can also be intrapersonal while still being communicative: if communication requires a speaker and a hearer, it remains possible that only one person would endorse both roles, in the process of thinking out loud for example. Only considering gestures made intentionally and at the attention of somebody else hence seems to lack accuracy, especially if analysing the role of gestures in the personal evolutionary process of L2 acquisition. Gestures made by learners as a reflex (hence not intentional) or during their thinking process or word-retrieval (hence not interpersonal) could indeed give valuable information for this study and should not be underrated.

It is nevertheless true that some hand-and-arm movements are not related to communication and would only parasite the analysis if taken into account. The gesture made by a pupil raising his or her hand to scratch his or her head because it is itching, for instance, would have no real communicative impact directly related to the potential discussion and would therefore not be considered as relevant in the study. This example happens to corroborate the theory of Gregersen (2007) because it is not interpersonal nor intentional. The same gesture of a pupil raising his or her hand, placed in a different context, might call up for a completely other interpretation: if this pupil wants to ask or answer a question for example. In that case, the communicative impact appears quite obvious and the gesture would certainly be taken into account, which seems once more to validate the previously exposed approach, the gesture having an interpersonal and intentional character this time. The two principles of Gregersen (2007) are therefore not completely invalid in order to try and define gestures. A third potential

interpretation of the hand raising, however, might attribute a communicative role to a gesture being unintentional and intrapersonal, for instance if the hand is unconsciously raised to the head during the thinking process, reflecting the inner communication going on inside the speaker's head. If the two previously mentioned principles were used to assess the relevance of the observed gestures, this example would be excluded, despite its potentially high source of information. Considering it could therefore be valuable. Gregersen and MacIntyre (2017: 7) apparently have the same reasoning as they show that, despite the interpersonal limitation that many experts have imposed on nonverbal communication, they also wanted to "include the *intrapersonal* function". Gestures will thus not be defined via the interpersonal and intentional parameters previously suggested. Still, hand movements which cannot be related to the speaking and/or the thinking process of the speaker will not be defined as gestures in this study.

The relation to the speaking process is also found in the gesture definition of Gullberg (1998, 2006), who considers that the criterion for inclusion of a gesture in her analysis is its relation to speech. If a body movement is not complementing or replacing speech, it indicates that this gesture has no communicative goal and is therefore useless to her. This definition is appealing and could coincide with the formerly given examples. It could indeed be said that all three situations match this definition: while in the case of questioning the gesture might be used to signify a will of communicating verbally and while in the case of thinking the gesture might make clear (consciously or not) that the person is going to communicate or trying to find a way to communicate verbally in an efficient way, it is true that the scratching of an itching head would not complement nor replace any speech (except if performed while overtly uttering something like "my head is itching"). This definition might consequently be applied to the current study, but it needs to be slightly modified. Gullberg (1998, 2006) indeed explicitly excludes self-adaptors from her definition (i.e. when the speakers "play" with their fingers, their ear, etc...) whereas some of these self-adaptors appear to be of great value in the current study. Observing them seems indeed especially interesting in order to distinguish when they are used and when they are not, as it might reflect some aspects of the speaker's thinking process. The approach of Gullberg (1998, 2006) can serve as a solid basis for the definition of gestures in the current study, but besides the speaking process, the thinking process throughout L2 acquisition must be taken into account and this is why self-adaptors will not be overlooked.

Based on what other researchers have discovered and established about gestures, a few elements appear to be common and relevant to the present purpose. A gesture is a movement made by a person with (a part of) the body. The movements which are taken into account by researchers depend on the perspective of the study: researchers will sometimes only focus on specific types whereas others could decide to consider any body movement. A feature which seems to raise some questions and to cause disagreements among scientists is the relationship between gestures and communication. The fact that scientists discuss what should be considered as a gesture makes quite clear that some body movements should not be called gestures. Once more, it mainly depends on the perspective and the aim of the study. In the current study, for instance, considering the movement of grabbing a real cup of tea and bringing it to one's lips as a gesture would be counter-productive if it is performed to actually drink the tea, but would be very useful if it is performed instead of the word *drinking* in a sentence. A significant gesture, for the current analysis, could be defined as follows: a movement from the shoulders to the fingers having an impact, conscious or not, on the communication process, i.e. the speaking and/or thinking process, for the speaker and/or the listener.

1.2. Classification of gestures

As the core definition of what can be considered as a gesture varies, depending on the researcher or the focus of the study, it becomes evident that the classification of these relevant gestures will also vary. Many linguists have advanced typologies with the aim of categorizing gestures on the basis of distinct features. These classifications often either present slight variations within a specific type of gesture they agree on, or focus on completely different features, resulting in diverging gesture types. Kendon (2004: 84) already concluded from this that “no attempt should be made to develop a single, unified classification scheme, since so many different dimensions of comparison are possible”. The classification suggested here will thus not be presented as the best and unique classification to follow, but will bring together different previously established classifications to be as comprehensive and useful as possible for the present analysis.

An interesting point to raise is that experts have sometimes labelled gesture types with ‘new’ terms, whereas their definition of this ‘new category’ actually strongly resembled existing ones. It is thus important to notice that what Rosborough (2014) labels *Italianate* is also called *conventional*, by Nicoladis et al. (2018) for instance, and is usually referred to as *emblem* in the literature. Similarly, Golding-Meadow and Alibali (2013) refer in their study to

pointing gestures, which have actually been referred to as *deictic* elsewhere. Another example of terminologies which could create confusion exists with the term *iconic gestures*, which has been used by many linguists, but not always with the same nuances. For example, while Beattie (2003) equals the term *iconic* to the term *illustrator*, Gregersen and McIntyre (2017) consider *illustrators* as a larger category which can be subdivided into various subcategories, namely *iconic*, *metaphoric*, *deictic* and *beats*. Capone and McGregor (2005) use the term *iconic* interchangeably with another term, which is *representational*, but other linguists such as McNeill (1992), Rosborough (2014) or Goring (2016) consider *representational* gestures as a general term which embodies both *iconic* and *metaphorical* gestures. Moreover, *iconic* gestures are also referred to as *lexical* gestures by Driskell (2003). These examples justify the necessity to precisely explain what is meant by all the terms which will be used in this study to avoid any confusion.

Different classifications found in the literature have been compiled and compared in order to establish a gesture typology which would be useful for the current analysis. Eight gesture types were considered as relevant because they could possibly be used by the learners during their speaking turn: affect displays, beats, butterworths, deictic gestures, emblems, representational gestures, regulators and self-adaptors. This typology closely matches the one proposed by Gregersen and MacIntyre (2017: 27-29) except that they consider iconic and metaphoric gestures as two distinct categories (instead of one here under the term *representational*) and they do not include self-adaptors nor butterworths. The classification of Ekman and Friesen (1969) has been of great use too, mostly for the inclusion of the self-adaptor category. They do not include the butterworths but the seven other types are referred to, even though they include more categories and subcategories than what is considered here. The butterworths have been used in different studies of Tellier and Stam (Tellier and Stam 2010; Tellier et al. 2012; Stam 2012) and directly refer to the work of Brian Butterworth. As Kong et al. (2015: 97) say, many researchers have already warned and illustrated that “one particular form of gesture can serve one or more communicative functions”, which is why one gesture can fall into more than one category: each category is non-exclusive, which means that being in one does not automatically exclude all the others.

The eight gesture types retained in the present study are the following:

- **Affect displays** tend to occur more with facial expressions, like when smiling, but can also be prompted via gestures. They are used, often unconsciously, to express emotions. An example could be a speaker moving a hand to his or her heart while talking about an event in which he or she was emotionally involved.
- **Beat** gestures emphasize and punctuate the speech (and probably the thinking process) of the speaker. This type of gesture is not related to any specific conceptual meaning or referent, but has a strong connection with speech, as it stresses some words, elements or ideas, matching (and perhaps regulating) the speech flow. An example would be to slightly move one's fingers on the stressed syllable *por* when saying the word *important*.
- **Butterworths** are gestures which are strongly related to the thinking process and the "lexical planning" (Butterworth and Beattie 1978: 358). They are often circular wrist movements with the palms facing up and they are made when the speakers cannot immediately find the word they want to say. These are spontaneous gestures most people could think of when they imagine a scene of someone trying to remember the name of an actor, for instance, eventually adding words like "you know this guy, erm". They are expected to be used by learners, especially when they lack vocabulary in their L2.
- **Deictic** gestures highly depend on the immediate context and may specify some information which is harder to transmit only verbally. It can be linked to verbal features like 'this' or 'that', which only carry meaning when brought into context. A speaker could for example point at an object while referring to it. This specific type of gesture is expected to be highly used by beginners in a L2, as is the case for babies (see Section 2.2.1), when the spatial context enables it. This type of gesture can sometimes nearly completely replace the verbal utterance: for example, showing something even without naming it allows the speaker to talk about this thing, despite a potential lack of vocabulary to designate it.
- **Emblems** are culturally dependent gestures which carry an intrinsic meaning and are therefore strongly independent of speech. They are apparently consciously made and understood. An example would be the thumb up, which calls up the meaning of something being approved by the speaker, of something being good, or positive.

- **Representational** gestures are used to illustrate (elements of) what is being said. It can be subdivided into iconic and metaphorical gestures. These gestures directly represent speech and are therefore relatively free from cultural connotations. They can be conscious or not. If what is illustrated is something more concrete, the gesture will often be referred to as *iconic*, whereas for a more abstract concept, the gesture tends to be called *metaphorical*. For example, a speaker could talk about a ‘narrow road’ and, by saying so, move his or her hands closer to each other, as to illustrate how narrow the road was. This would be an example of iconic gesture. An example of metaphorical gesture could be to move one’s hand behind one’s shoulder to talk about the past: this would be more abstract. In the current study, it seems important to acknowledge the distinction between iconic and metaphorical gestures, but both subtypes will be analysed as *representational* for practical reasons, as these subcategories might be harder to differentiate in the practice.

- **Regulators** have a social role in communication. They are used to regulate the conversation and especially the turn-taking between the different characters interacting in the discussion. For example, the speaker could start moving his or her hand towards some potential interlocutor, inviting him or her to take the word and keep the conversation going.

- **Self-adaptors** have often been excluded from studies because they “lack semantic attachment” (Kong et al. 2015: 100). However, Ekman and Friesen (1969: 85) have suggested that self-adaptors are “first learned to facilitate or block sound-making and speech”. Therefore, including self-adaptors in this study –observing the potential impact of gestures on the production of speech in L2 and throughout the acquisition process – seems relevant. This type of gestures is also thought to help the learners relax and/or focus on the L2 speech they need to produce. An example of self-adaptor could be a speaker constantly rolling a pencil between his or her fingers while talking.

1.3. Universality of gesturing

If analysing the types of gestures used is important in the current study, it is also crucial to deal with the question of universality of gestures. This universal character has partly been pointed out by Goldin-Meadow and Alibali (2013: 1) as they say that everybody gestures, even a blind person who has never *seen* other people gesturing, for instance. This universal character is developed in many other analyses and conceptions of gestures. Gullberg (1998: 61) explains for example that gestures were considered in the Antiquity as “a universal language understood by all”. Kendon (2004: 3) also suggests that this universal and natural character given to gestures is the reason why people were interested in them. He adds that when people are prevented from speaking for one or another reason, they can often communicate via gestures and this becomes, then, “a form of language all by itself” (Kendon 2004: 3). An interesting theory is to be found in the work of Darwin (1872), who compares how humans and animals express their emotions. He draws many interesting conclusions from his observations, but the main point regarding the universality of gestures could be summarised as follows: some gestures are universal and innate whereas others are culturally inherited. Darwin (1872: 352-353) points out that the intuition people may have about the universality or not of certain gestures is often misguided as some gestures which are thought to be specific to a community appear to have a universal character whereas other gestures which seem innate appear to be culturally bounded. He explains that the innate nature of a gesture can be induced from the fact that it is performed by humans all over the world regardless of their culture and also by babies and by blind people. On the other hand, some gestures appear to be learnt as any language must be. Darwin (1872: 356) also insists on the fact that some gestures are not innate but have acquired a universal status “through the power of conscious and unconscious imitation”. Saying that gestures form a universal language is thus overgeneralising and not totally accurate. Nespoulous and Lecours (1986: 51) even talk about “the myth” of gestures’ universality. Several studies actually lead to the conclusion that *using* gestures is universal but that not all gestures are cross-culturally shared.

The culture in which the speaker has grown seems to play a role both in the choice and in the frequency of the gestures which are made. Nespoulous and Lecours (1986: 51) say that “cultural differences are far from negligible”. A good and simple example given by Fast (1970: 14) is the expression of *yes* and *no* in the Western and Indian cultures. In the Western world, people are used to nodding up and down in order to signify their agreement. Conversely, moving their head from right to left represents the negation. It might appear as innate and

universal to most speakers in the Western world because these words are so often used that their production is unconscious and seems logical. However, in the Indian world, people shake their head in the other direction to mean the same thing: they shake their head up and down to show their disagreement and from side to side to express an affirmation. If somebody from one culture goes to the other one without any knowledge of this difference and starts gesturing without any cultural adaptation, there is a great risk of misunderstanding. Different interpretations for a same gesture often lead to misunderstanding but the consequences can be even worse as some a priori neutral or even positive gestures might have an obscene or upsetting connotation in other foreign cultures. The example of people who “form a circle with fingers to indicate “O.K.”” is given by Hişmanoglu and Hişmanoglu (2008: 172). If this gesture is usually seen as positive, it can be seen as very obscene by Brazilian and German people. Also, depending on the context, this gesture sometimes means “zero” or “worthless” for French-speaking people, or “money” for Japanese-speaking people. This is one of the many examples in which one gesture calls for unrelated and even contradictory interpretations depending on the cultural setting. Besides variations in the interpretation of gestures, there is also an apparent difference in the frequency of performed gestures. The notion of “high-gesturing” and “low-gesturing” languages, which is to be found in the work of Goring (2016: 5), suggests that some societies tend to gesture more often than others. An Italian speaker, for instance, arguably tends to gesture more than a British speaker. Gullberg (1998: 57) nevertheless stresses that the differences in gesturing are not always bound to a culture “but also to such factors as region and socio-economic status”. The reality of these cultural variations leads researchers like Hişmanoglu and Hişmanoglu (2008: 168) or Gregersen (2007) to insist on the necessity of teaching at least some non-verbal cultural aspects too when teaching a new language.

If some differences can be observed between cultures, some variations between people sharing the same culture have also been noticed. Nicoladis et al. (2018: 644) explain that the personality of speakers can impact their use of gesture. It is not that extrovert people always gesture more than introverts, but their degree of introversion or extroversion might have an influence on their reaction and their use of gesture in a specific context. The researchers explain that, in a situation in which the speakers are not sure that their interlocutor understands what they are saying, for instance, extrovert people tend to gesture more than introvert people. This distinction between the gesturing habits and tendencies of extrovert and introvert people is also present in the work of Gullberg (1998: 55). Based on previous studies, she also concludes that the extrovert character of a person will not have any direct and significant influence on the

frequency of gestures or on their duration, but that it will impact the way a person uses gestures. For instance, she suggests that introvert and anxious people have been proved to “use more self-adaptors” or “self-touching gestures” (Gullberg 1998: 55). Interestingly, she adds that this higher use of self-adaptors and self-touching gestures also seems present among people who do not (fully) master a language. Another element which appears to influence the use of gestures, next to the speaker’s cultural background and personality, is the topic of the conversation. Gullberg (1998: 54-55) claims that people use gestures differently depending on their familiarity with the topic and depending on the content of the topic. The context in which gestures are performed plays thus a role in the way the speakers will use them. Naturally extrovert people might become more introvert or anxious when facing an unknown or unmastered situation and it is perhaps why the speakers’ use of gestures varies with their proficiency level in the spoken language. No absolute generalisation should be made about the use of gestures because of the multiple cultural, individual and contextual variables which influence it, but it seems obvious that gestures must have a crucial function because everyone uses them, even if they do so differently.

1.4. Functions of gesturing

Scholars have presented many theories round the potential functions of gesturing, both from the listener’s point of view and from the speaker’s point of view. Stam (2012: 55) lists seven main functions which summarise the findings of most gesture studies: gestures can add information in the same way as speech does; they can support speech by adding visual and conceptual clues; they can reduce cognitive load to improve the final production; they can punctuate transitions in the thinking and speaking processes; they can give tacit indications to rule the conversation between speakers and listeners; they can indicate some troubles in the speech production; and they can make the thinking process and especially word retrieval easier. Many researchers have tried to attribute functions to gestures in general, like him, but some others have decided to focus on functions which seem to be more intrinsically related to specific gesture types. Consequently, as explained in Section 1.2, affect displays will serve to express emotions whereas beats will punctuate speech, and possibly facilitate the speaker’s thinking process. Butterworths will directly be linked to the thinking process and word-retrieval. Deictic gestures will add details to speech or will serve as an alternative to it, such as emblems which will have more culturally inherent meanings, and representational gestures which will illustrate speech. Regulators will have more control on the social aspects of communication and self-adaptors will comfort the speaker and a priori facilitate the speech production.

Many researchers have also tried to prove that gestures could have a communicative impact. Goldin-Meadow and Alibali (2013: 5) have supported this theory via different experiments which enabled them to conclude that some gestures are made on purpose by speakers, with the specific aim of enhancing their communication to optimise the listeners' comprehension. The work of Gregersen and MacIntyre (2017) seems to prove them right as they conclude that gestures made by speakers enable the listeners to gather pieces of information from gestures which might have been missing if they had only had access to speech. They hypothesise that listeners "synthesize the auditory and visual channels to figure out the meaning of a given situation" and thereby confirm the positive, and even crucial effect that gestures have in the communication system (Gregersen and MacIntyre 2017: 45). But many researchers stress that, if gestures have an impact on the listeners' perception, they also have an impact on the speakers' production of a message. Driskell (2013: 451-452) indeed shows that forbidding speakers to use gestures had a negative impact on their speech production as well as on the listeners' comprehension. Conversely, listeners could better understand a conversation when gestures were added to speech and speakers could better produce speech when they were not prevented from producing gestures. He also adds that the influence of gestures on the listeners' speech comprehension and on the speakers' speech production have been proved independently in his study: it is not because speakers' speech is enhanced by gestures that listeners understand the speech better but it is because they can see the gestures which are performed with speech. Therefore, gestures made by the speakers seem to positively impact both the speakers and the listeners.

Next to the communicative role which is often developed in gesture studies, some pedagogical implications have been suggested. Gullberg (1998: 70) says that "gestures appear to facilitate interpretation, understanding and retention of the message". She confirms this pedagogical function by developing that gestures made by the learners when they learn as well as gestures made by teachers when they teach positively influence the learning results (Gullberg, 2014: 1871-1872). Capone and McGregor (2005) support her findings when they suggest that children tend to achieve better performances when they gesture while learning new skills. Golding-Meadow and Alibali (2013) add that the impact of gestures in the enhancement of learning will be even more optimal in the case of language learning. The idea of "cognitive crutch" presented by Goring (2016: 7) also justifies the positive effect of gesturing in the learning process, especially regarding language acquisition, as she develops the compensatory character of gestures which provide cognitive support to the working memory when it is too

slow or damaged. The use of gestures in relation to L1 and L2 will be further developed (see Section 2.2) but it is interesting to notice that in the context of learning, both learners and teachers will tend to gesture in specific ways and at specific moments. These findings indicate that there must be a correlation between gestures and the acquisition of a language.

1.5. Connection between gestures and speech

As previously mentioned, despite the wide range of gesture definitions found in the literature, a point on which researchers do agree is that gesturing plays a role in communication. It has already been suggested that communication can broadly be divided between verbal and nonverbal (sometimes also adding the para-verbal category) and that while speech would fit in the first category, gestures would form a part of the second one. However, many researchers have shown in various ways that these categories, far from being completely distinct, seem to be intertwined, which implies that there must automatically be a certain connection between gestures and speech. Beattie (2003) states that this connection between gestures and speech is precisely what distinguishes gestures from the other sorts of body language. Gullberg et al. (2008: 6) provide a few further arguments, based on the literature, to prove the complementary character of both communication means, one of them being that “gestures and speech develop together in (first) language acquisition (...) and that they break down together in disfluency, in aphasia, etc. (...)”. The nature of this relationship, however, seems to be more complex and more disputed among experts. This might be explained by the fact that researchers do not always take the same starting point to discuss the question of this connection. Some prefer to investigate on the basis of historical evolutions, orienting their questions towards the finding of the origins, wondering which of the two means of communication came first and might therefore be prior to the other one. Some others base their observations on experiments around people talking and/or gesturing, in order to try and find evidence to establish the relation, and potentially the hierarchy, which might exist between the two means of communication. The conclusions which are drawn may vary depending on the investigators and the method of analysis, but they all seem to agree on the fact that there is an undeniable but intricate connection between speech and gestures.

From a historical perspective, it is generally accepted in the research community that gestures were used prior to speech, temporally speaking. Dortier (2003) points out that gestures would indeed have been much more appropriate to the environmental situation of humans in prehistoric times: making too much noise while being surrounded by savage animals and

predators could be a death penalty. Corballis (2010) considers practical reasons which might have led human beings to switch from a purely, or essentially, gestural communication to vocalization: he hypothesises that human beings using more and more manual tools, they started to need their hands to carry these objects and therefore needed to find another way to communicate, not (only) using hands. He also bases his conclusions on several studies, such as for instance the analysis of the neuronal activity of monkeys responding to different movements (manual and facial), to prove that vocalization, and therefore speech, emerged after body movements. Moreover, he justifies his assumption that gestures and speech are linked by highlighting the fact that despite the addition of speech to gesture, apparently to optimize the communication between human beings, it has not completely taken over from gestures as gestures are still used nowadays, together with speech. He adds that children are able to learn spoken types of language in the same way as they can learn sign languages. According to him, the evolution from a primary gestural communication to a vocal one must have been the result of a long process: it is hardly possible, if not completely impossible, that this change happened abruptly (Corballis 2010: 3-5). Therefore, it might be concluded that words were certainly progressively added to gestures, mostly for practical reasons, but that they cannot replace them (or at least not completely) as gestures are still used nowadays together with speech, and are sometimes even still carrying the core of the message themselves.

Next to the historical evolution of gestures and speech, some researchers have focussed on the way human beings have considered and interpreted them. Gullberg (1998) develops the evolution of these conceptions through the ages, explaining that gestures have often been defined via speech. She explains that in the Antiquity, gesturing was considered “as a language in itself”, and even as “a universal language understood by all” (Gullberg 1998: 61). The universal character of gestures is a question to which many researchers have tried to answer (see Section 1.3). Researchers have then gradually started orienting their debate towards the relationship there could be between gestures and speech, regardless of the potential universality of gestures. Three main hypotheses have been proposed: gestures have either been seen as being “governed by speech” or as being “primary” whereas language was considered as “subordinate” and finally some have concluded that neither gestures nor speech had a higher status, but they were “interdependent” (Gullberg 1998: 61-62). These three positions could coincide with the results of various studies, although the last one seems to reconcile the other two and thus to be applicable to most research.

The first view, considering gestures as a supporting device, could be used to interpret the following examples of gesture studies. For instance, Morett et al. (2012: 5) insist on the fact that, although gestures can add valuable information, speech remains irreplaceable. Theories summarised by Gullberg (2006) could justify this point of view when they observe the facilitating nature of gesture in the speech process, be it to find appropriate words or to verbalise a concept, for example: gestures could be used by the speakers to serve speech. To determine the relationship between gestures and speech, another interesting approach is the one of Hoetjes et al. (2014: 264-265), who compared speech performed when speakers were allowed to use gestures and when they were forbidden to gesture. They explain that they could not find significant differences between the speech used with or without gestures (especially in terms of fluency) and they draw two potential conclusions. The first possibility, according to them, is that gestures were still intended though not actually made, and they thus indirectly helped the speech production. The second possibility is that speech can be performed without gestures as gestures only support speech but are not indispensable. Gestures could be seen in these examples as helping (or not) the speakers to transmit their message verbally: speech might therefore be considered as the main communicative mean and gestures as its assistants.

The second view, considering that gestures are more important than speech, is to be found in various studies, mainly in the fields of communication, management or marketing. In that way, some recent results obtained by Cornelissen, which are described in an article by Torres (2019), make him conclude that gestures “matter more than words”. Through the interview, Torres (2019) presents Cornelissen’s late findings, namely that gestures have been proved to be significantly much more impacting for the listeners than any other aspects related to language when speakers “pitch an idea”. The impact is thought to be even bigger when the speakers are trying to convince their audience or when something new (like a new product, for instance) is being introduced to the hearers, making this new thing more concrete. In these situations, it has been proved that gestures highly impact the reaction and understanding of the hearers. The explanation of Goldin-Meadow and Alibali (2013) could also support this view to some extent. They argue that gesturing, besides its supporting character towards the creation of speech, also sometimes reveals a certain knowledge or understanding which is not (yet) expressible through speech. In that way, gestures would make it possible to communicate either together with speech, or without it. In these examples, gestures could be concluded to be more important than speech, because messages can be carried by gestures only and speech would therefore not be absolutely required.

It is however important to notice that the two previously mentioned interpretations could be reconciled under the third interpretation. This last view, considering gestures and speech as two coexisting channels which both confer meaning in a specific way, seems indeed to be convenient for most researchers nowadays and all the previously given examples can be included and explained via this conception. Indeed, in the previously mentioned studies of Morett et al. (2012: 5) and of Goldin-Meadow and Alibali (2013), the researchers actually go further than this a priori hierarchy between speech and gestures. Besides highlighting the irreplaceability of speech, Morett et al. (2012: 5) observe that speakers tend to modify their use of gestures depending on the direct presence or not of an interlocutor: people tend to gesture more when they can see the person they are talking to and conversely to specify their speech and to gesture less when they cannot see them. She interprets this as evidence that speakers are “aware of the communicative properties of each modality”, hence insisting on the existence and importance of both channels. Goldin-Meadow and Alibali (2013) also go further than the hierarchy which could be induced in their study: they claim that gestures can indeed be used without speech, carrying meaning on their own, but instead of considering gestures as intrinsically more important, they develop speech and gestures as two different but parallel communicative means. They point out that both means convey meaning in different ways and they give them an equal though not similar status. Holler and Bettie (2003) go deeper in this theory of double channel as they analyse speech and gestures as two distinct means of communication which operate in different contexts and consequently serve diverse roles. They argue that understanding both channels separately would enable a better understanding of their interaction when taken together. Kendon (2004: 111) supports this viewpoint and considers “gesture and speech as two different kinds of *expressive resource* available to speakers”. He presents an interesting observation which might justify this view: “Where the utterance was repeated, for instance, or where it was corrected, both components [gestures and speech] were repeated or both components were corrected as if they were equally important partners in a single unit” (Kendon 2004: 156). Based on the argumentation of most researchers who embrace this ‘double channel theory’, gestures and speech could be seen as two cognitive and communicative tools in humans’ interactions, both highly important in the study of language and language evolution.

2. L2 acquisition

The previous chapter has developed how gestures are considered in the literature and has established the possible connection which exists between gestures and speech. It has also been pointed out that gestures have been attributed various functions, some of them being directly linked to language production. For instance, Gullberg (1998: 64) concludes that gestures can be “seen as part of the speech production process, as part of global communicative intent, or as an integral reflection of thought”. Next to the communicative function, she and other researchers claim that gestures can be attributed a certain pedagogical function, improving the learning process (see Section 1.4). These two elements justify the relevance of analysing gestures in the frame of language acquisition. As the current study focuses on gestures made by language learners when they have to perform speech, it might be useful to observe what has been said and what is known about language acquisition, about gesture acquisition and about the potential relationship between both. In the literature, it is claimed that the acquisition of a L1 (also *first language* or *mother tongue*) differs from the acquisition of a L2 (*foreign* or *second language*). A first important step is to clearly define the terms related to the field of language acquisition which will be used here. An overview of the (spoken) language acquisition in the different contexts will be given afterwards, highlighting the similarities and the distinctions which are worth being made between the L1 and the L2 acquisition processes. As the focus of this study is the use of gestures, the last point will develop what has been found in the literature about gesturing in L1 acquisition, in L2 acquisition and in the more specific case of bilinguals.

2.1. Terminology of language acquisition

In order to understand what is meant by *second language acquisition*, some terms are important to clarify. A distinction will be made between the terms *first language* and *mother tongue* and between the terms *second language* and *foreign language*. Because the terms *L1* and *L2* have been used in the current study, it appears essential to clarify what is meant by these two words. Besides the general L1-L2 distinction, which can be applied to any language, some information will be given about two models which have been created specifically for the English language: Kachru’s Model of the Three Concentric Circles and Schneider’s Dynamic Model of postcolonial Englishes. These models are relevant to develop because the participants observed in this study come from Hong Kong, a former British colony, and have English as their L2. Also, as will be explained, the distinction between the terms *second* and *foreign* languages tends to be ambiguous. This is why clarifying the status that English has for the observed students seems essential.

The term *first language* or *L1* is used when referring to the first language which is spoken by a baby and which, usually, will be the primary language used throughout the person's life. The term *mother tongue* is often used in everyday language as a synonym of L1. However, Punchihetti (2013: 3) points out that this terminology is quite out-of-date in this globalised and multicultural world: the first language a child masters is not always the language of his or her mother anymore. In a situation where the two parents do not speak the same L1, they could choose to teach their child either the mother's L1 or the father's L1 or they could choose to teach both simultaneously, hence giving their child two first languages. Another situation is that of immigrants, for instance, who could prefer to teach the language of their new country to their child as first language instead of theirs. Many other examples could confirm this statement that relating the first language (only) to the mother's language is not accurate anymore, even if the idea behind the term *mother tongue* in everyday language still corresponds to the concept of L1.

Fewer and fewer people are strictly monolingual, which means they learn at least one other language in addition to their first language. To mark the contrast with this L1, linguists often refer to *L2*, but it should be mentioned that this term refers to two concepts, namely *second* and *foreign language*. The distinction between these two terms is sometimes ambiguous. In order to clarify it, Punchihetti (2013: 3-5) develops the contexts in which languages can be acquired. He explains that there exist three different contexts of language acquisition: endolingual contexts, semi-endolingual contexts and exolingual contexts. Languages are acquired in an endolingual context when they are accessible and used all around in everyday life, like English in the UK, as the researcher suggests; languages are acquired in a semi-endolingual context when they are present in everyday life but only to a certain extent, like French in the UK, as he says; languages are acquired in an exolingual context when they are not present in everyday life outside the classroom, like Sinhala in the UK (Punchihetti 2013:5). According to the Puchihetti (2013: 3-5), a major distinction between second and foreign languages is that second languages, like first languages, are acquired in endolingual (or sometimes semi-endolingual) contexts only, whereas foreign languages can be acquired in any context. Indeed, he suggests that people who learn a second language always have many occasions to use this language on an everyday basis because they are surrounded by this language in their environment, be it their country or, on a smaller scale, their family or social circle. On the other hand, foreign languages tend to be learnt for "personal" or "for academic or professional reasons" and are therefore usually learnt in exolingual contexts like at school or

university, but they can also be learnt in (semi-)endolingual contexts, in the case of an Erasmus experience for instance (Punchihetti 2013: 4).

Littlewood (1984: 2) summarises the distinction between the two terms by saying that “a ‘second’ language has social functions within the community where it is learnt [...] whereas a ‘foreign’ language is learnt primarily for contact outside one’s own community”. According to Punchihetti (2013: 3), it is indeed a question of “historical, geographical, [and] socio-economical ties that the country shares with the country of origin of the non-native language”. Sometimes, these ties can be very strong precisely because the learner’s country and the “country of origin of the non-native language” are one and the same country, for instance in countries which recognise and use more than one language as their official language. Punchihetti (2013: 5) also suggests that a distinction has to be made between the “social or personal environment” of the learner and the context in which the learners acquire the language: as already mentioned, second and foreign languages could both be acquired in a (semi-)endolingual context, with direct contact to the language and the associated culture. What must be observed to distinguish second and foreign languages is thus not the context in which the language is learnt but the relation this language has (or not) with the learner’s origin: a second language is strongly linked to the speaker’s community whereas a foreign language is not. However, this theoretical distinction is sometimes much more complicated to see in the practice. The English which is spoken in Hong Kong, for instance, could be argued to be both a second and a foreign language. This is why the term L2 is often preferred in the literature as it refers to both second and foreign languages.

Another terminology which has been proposed in order to reduce the ambiguity between the terms *second* and *foreign* language is found in the Model of the Three Concentric Circles suggested by Kachru (1985). His model has been created and based on the English language. Indeed, the international character which has been acquired by this language has led to the emergence of many (sub)varieties. In his model, instead of talking about English as a first or native language (ENL) and English as a second or foreign language (ESL or EFL), Kachru (1985: 12-13) suggests using the terms *inner circle*, *outer (or extended) circle* and *expanding circle*. The three circles could broadly be associated with the previously discussed conception: the inner circle could represent the native speakers of English, i.e. L1 speakers, the outer circle could represent second language learners and the expanding circle could represent people learning English as a foreign language. Still, this new terminology (and conception) lessens the ambiguity which can be found with the other terms. The countries which form the inner circle

are countries where English is the primary language, such as the USA, the UK, or Australia. The outer circle includes multilingual countries in which English is one main language among others, with an official or semi-official status. Some examples are India, Singapore or Sri Lanka. Most countries in this circle have once been colonized by English-speaking peoples. The countries in the expanding circle are countries which “do not necessarily have a history of colonization by the users of the inner circle” but in which English is widely used, due to its universal status (Kachru 1985: 13). Some examples are the Netherlands or Japan.

A consequence of the model of the three concentric circles is that linguists can use it to develop the plurality of English(es) instead of considering one variety as the only existing standard. The emergence of the concept of New Englishes has led Schneider (2007; 2009) to establish another model: The Dynamic Model of postcolonial Englishes. According to him, all varieties of English undergo some changes over time and the evolutionary process is always the same (Schneider 2009: 284-287). There is first the *foundation*, during which English is settled as a lingua franca. Then, there is the *exonormative stabilization*, during which the colonial status is clearly established. Thirdly, there is the *nativization*. This is a crucial step as this is the moment when indigenous start to willingly include the culture and language of the colonisers, resulting in a lot of contact between the ethnic groups. The fourth phase is the *endonormative stabilization*, which is when the new English variety starts to be used in the country, with new norms being accepted and codified. The fifth and last phase in his evolutionary process is the *differentiation*, which occurs when there is no external pressure from the colonisers anymore, but the new English variety is used by the citizens and is part of their identity. Each variety of English which is spoken around the world can be placed on this continuum. The idea that some of the varieties of English could become institutionalised, which means they acquire a different status and a sort of independence or recognition, is already mentioned by Kachru (1992: 55-56) when he develops his Model of the Three Concentric Circles. He explains that accepting a new variety as a norm is a slow process but that, once institutionalised, this variety can be characterised as having a certain “emotional attachment”, “functional importance” and “sociolinguistic status” (Kachru 1992: 55). His explanation coincides with the various phases which are developed by Schneider in his Dynamic Model.

In order to have a clear view on the position of English in Hong Kong, retracing the history of this place is essential. Bolton (2002) gives a complete overview of the use of English in Hong Kong throughout the years. He explains that “the role that English plays in contemporary Hong Kong society has been shaped by specific historical processes, including

British colonialism (...) from 1842-1997” (Bolton 2002: 32). He also stresses that the period of late British colonialism, between the mid-1960s and the 1990s, has been essential in the construction of the city, and therefore in the construction of the identity of its people (ibid.). The attitude of the population towards English can also be observed through the evolution of the languages used and imposed in the education system. Indeed, Bolton (2002: 34) points out that, in the 1970s, neither Chinese nor English could be chosen as the only language of instruction at school, which, according to him, mirrors the difficulty of the government to position themselves between the two languages. During the following decades, the English language was used differently in the country, with some schools using only English, other schools using Cantonese (though the handbooks were in English for everyone) and finally a lot of schools using a mix of both languages (Bolton 2002: 35). Schneider (2007: 136-137) suggests that the introduction of English at school created a “mass bilingualism”, which means that bilingualism was not restricted to an elite anymore. He adds that the international status of English combined with the internationalisation ambitions of the city also boosted people in Hong Kong to use English more and more. The evolution of English in Hong Kong, Bolton (2002: 44) explains, gave it an “autonomous (or semiautonomous)” status. The Hong Kong English variety can be observed for instance in some students’ mails, who seem to frequently use “a ‘mixed’ form of written English and Chinese” (Bolton 2002: 43).

Still, as Schneider (2007: 137) points out, “the status of the variety itself has been disputed”. He adds that the constant and noticeable evolution of English in Hong Kong confirms his Dynamic Model. As he explains, “the constituent features of phase 3 fit together remarkably well in the case of the Hong Kong of the last third of the twentieth century”, but the variety seems to evolve towards the fourth phase of his model (Schneider 2007: 137). A similar conclusion could be drawn regarding the Model of the Three Concentric Circles of Kachru as the situation of Hong Kong seems to oscillate between the outer and the expanding circle: Hong Kong was once a British-colony but nowadays, English seems to be mostly used because of the international advantages it brings. Moreover, the status of Hong Kong English is not official (yet), even though some linguistic variations appear to be shared among the speakers, mixing codes from English and Chinese. Although both models are applicable to the use of English in Hong Kong, it seems more neutral and therefore more relevant to use the term *L2* in this study, especially because it covers both English as a second and as a foreign language (ESL and EFL) and is therefore unambiguous.

2.2. L1 vs. L2 acquisition processes

There are various aspects in which the acquisition of a language seems to vary depending on whether the learner is dealing with a first language or whether he or she already masters a mother tongue. The main difference which might be highlighted between the acquisition of a L1 and a L2 is indeed the fact that learners who work on a L2 already master at least one other language whereas the acquisition of a L1 starts from nothing. The context of acquisition can also play an important role in the acquisition process and will usually be directly linked to the nature of the L1 or the L2 (see Section 2.1). The condition in which L1 and L2 are acquired and developed varies indeed on numerous aspects. As Punchetti (2013: 5) points out, L1 are always acquired in (semi-)endolingual contexts, whereas L2 can be acquired in endolingual or in exolingual contexts. This is confirmed by Ipek (2009: 160) when he explains that L1 can be acquired “only in a natural context” whereas L2 can be acquired “in various contexts”. It is indeed possible to learn a L2 in a much more conscious and active way, through grammar rules, structures and specific vocabulary which are taught to the child or the adult, for instance. This artificial context of acquisition is impossible with a L1. Moreover, Mettewie (2017-2018) adds that all babies are able to (and do) acquire a L1, in an unconscious and passive way, by first and only having access to speech, whereas not everyone learns a L2, but those who do usually immediately get access to this new language both in its spoken and its written form. A L2 can thus be acquired in a context similar to the L1 context, but it can also be very different.

In addition to the distinctions in the way learners get access to the language they aim to master, there are distinctions in the way they are going to evolve within the acquisition process. Ipek (2009: 155-156) explains that there are specific patterns, in the development of L1 as well as in the development of L2, which are usually followed by the learners. These patterns appear to share similar characteristics, but there are also characteristics which are more specific to the one or the other. He divides language acquisition into three main stages: the silent period, during which the learners mostly listen to the target language, the formulaic speech period, during which learners start using blocks of words, and the structural and semantic simplifications, during which learners start forming complete sentences but in a simple way. He concludes that the only nuance which exists between L1 and L2 acquisition is the “urge” of L2-learners “to skip the silent period” (Ipek 2009: 156). The importance of this silent period and of the “input” which is then gained is supported by many researchers, like Stam (2012: 56), and seems crucial in the language acquisition processes of both L1 and L2.

Although the three stages of language acquisition he proposes coincide with most other language acquisition divisions, Ipek (2009) insists on the utility of another concept, namely *Interlanguage*. Interlanguage stadia are the different stadia L2 learners have to go through in order to improve their proficiency. These stadia are not present in L1 acquisition as they are mostly based on the fact that L1 acquisition is not based on any priori language knowledge, unlike L2 acquisition. This concept of interlanguage also stresses the “systematic, dynamic and constantly evolving” character of language acquisition (Ipek 2009: 155). This idea that language acquisition is a dynamic process is also explained by Mettewie (2017-2018: 26), who insists on constant potentiality to learn, but also to *unlearn* the target language which is being acquired. When the concept of interlanguage is included in the L2 acquisition process, a particular phenomenon can be observed which does not exist in L1 acquisition, i.e. fossilization. Fossilisation is the idea that some incorrect elements might rise during the L2 acquisition process, mainly due to negative transfers from the learner’s mother tongue. If these elements are not corrected, they become fixed (fossilised) and very hard for the learner to get rid of. As Ipek (2009: 160) points out, various causes can be found to this phenomenon: it can be due to the age of the learners (it has been shown that linguistic acquisition is easier at a certain period of time and that past that period it becomes harder – though not impossible – to learn and master a new language), the communicative pressure (when learners are asked to produce language which they cannot correctly form yet due to their proficiency level), a lack of learning opportunity and/or feedback and finally, a lack of motivation and/or a negative attitude towards the target language. These are elements which can influence acquisition of a L2 but which do not play any role in L1 acquisition.

In addition to the general conclusions which are drawn around the acquisition processes of L1 and L2, researchers have also turned their attention to individual differences (also called inter-learner variation), which appear to influence the acquisition of languages. Some of these factors are given by Stam (2012: 56): the age of the learners (before or after the critical period), the motivation they have to learn the target language (why they learn it) and the attitude they have towards this L2 and its culture (which can be positive or negative), the context in which they learn it (teaching methods, endolingual or exolingual context), the contact with the target language (the more input the learners get, the better they will tend to perform), and some elements directly related to the learners personality and/or language learning facilities. Mettewie (2017-2018: 28) also adds that the distance between the mother tongue and the target language can play a role because transfers are often done, and can be either positive when an

aspect of the target language is similar to the L1, or negative when they differ. If general conclusions can be drawn regarding the acquisition of L2, it must nevertheless be kept in mind that many individual factors can influence this process, either making it easier or more difficult.

This being said, one could conclude that the acquisition processes of L1 and L2 share a similar structure but that some particularities are to be found in SLA. If the characteristics of L1 acquisition can sometimes be found in some situations of L2 acquisition, the contrary is not always true. This mostly has to do with the fact that L2 acquisition is automatically based on L1 knowledge whereas L1 is, as suggested in the term, the first language which is acquired and can therefore not be based on any prior (language) knowledge. Also, L2 acquisition is usually consciously learned and can be learnt in a non-natural context like in a language class at school for instance, which is never the case for L1. This implies the importance of interpersonal variations in SLA such as the learners' motivation and attitude towards the target language, which are not present in L1 acquisition, and also the learners' linguistic and cognitive faculties, which might strongly be influenced by their age. The input is of great importance too, in both L1 and L2 acquisition, and giving feedback seems also essential in both contexts, especially in SLA in order to avoid fossilisation. L1 and L2 acquisition processes are similar but not exactly the same.

2.3. Language acquisition and gesturing

It has been suggested that there is a connection between speech and gestures: both could be defined as parallel but distinct communicative means which convey a message. In the first section of this chapter, it has been shown that (spoken) language acquisition follows a certain pattern which shares similar elements with both L1 and L2 acquisition, despite some features being specific to the learning of a L2. The main difference is that the learning of a first language is not based on any former linguistic knowledge whereas the learning of new languages at a later stage in life is often based at least on the mother tongue, which is already mastered. It sounds interesting to study the acquisition of gestures in both cases, in order to see whether gestures play the same role in each case or whether certain variations can be observed specifically in the case of SLA, adding new roles to gesturing for instance.

2.3.1. Gesturing in L1

Many researchers have pointed out that babies' first gesture emerges before their first uttered word. Esteve-Gilbert and Prieto (2013: 308) have established that the first recorded

gestures performed by the babies are made without speech. Around the age of 1 year old, children will tend to produce more and more gesture-word combinations, until they finally start to gesture more with speech than without. Goldin-Meadow and Alibali (2013: 5-6) go further when they observe that babies are first prompt to produce gesture-word combinations before word-word combinations, which makes them hypothesise that gesturing influences the emergence of linguistic capacities. This hypothesis has also been formulated and supported by Capone and McGregor (2005: 1470), stating that the gesturing of children heightens their linguistic performance and language acquisition (see Section 1.4). It has already been suggested in the previous chapter that gestures and speech should be considered as two intertwined channels, each of them carrying meaning in a certain manner and enabling communication. Gullberg (1998: 74) insists on this conception too and says that when studying the evolution of speech and gesture acquisition processes, it is noticeable that gestures are not “replaced by speech”, as adults gesture as much as children do, but that a complementarity can be found in the evolution of the two channels. She adds that gestures performed by adults can be characterised as “more complex and refined” than the ones performed by children, which can remind one of the way vocabulary and grammar structures complexion evolves through life (Gullberg 1998: 76). Based on their analysis, Esteve-Gilbert and Prieto (2013: 312-313) conclude that the production of gestures is intentional in the same way as the production of speech and that the first gestures which seem to be made, namely the deictic ones, show a “temporal coordination” with speech, which stresses once more the connection though independence between both channels. Gesture acquisition appears to be similar to what has been said of language acquisition since there is a possible connection in the acquisition of the two channels and also a parallelism in the way gestures evolve and vary through different stages in life.

The different stages of the acquisition of gestures have been introduced among others by Gullberg et al. (2008). As previously mentioned, the first gestures which babies make seem to be the deictic ones and they usually start to be made before the child’s first birthday. The researchers associate these gestures with the role of “drawing the interlocutor’s attention to something in the environment” (Gullberg et al. 2008: 7). They say that through these first gestures, babies can either request, show, give or point to something specific, without the need for words, which might not be mastered yet. After their first year on Earth, babies will apparently start producing gestures which carry a type of information closer to the one conveyed by words, and they will start to be able to communicate about things which are not

directly related to their immediate environment. At that stage, they will also start combining their gestures with words and, once more, there will be a connection between the two channels: the topics communicated via speech and via gestures are indeed the same. The gestures (and words) which are performed at that time will represent some basic needs which are present in babies' daily routine such as eating or dressing up. The older children get, the less restricted to the immediate context and the more varied their gestures will get (Gullberg et al. 2008: 7-8). Besides the general gesture evolution, some more specific features are to be found depending on individual and cultural variations (Gullberg et al. 2008: 9). These cultural and personal variations have indeed been pointed out in many gesture studies, also when observing adults (see Section 1.3). It can be concluded that gestures appear to evolve similarly for all babies, although the evolution concerning the gesture types and the frequency of gestures used when children grow up possibly get culturally and individually specialised.

2.3.2. Gesturing in L2

It has been shown that gestures seem to have a positive impact on the learning process, and this when gestures are made by teachers when they teach as well as when gestures are made by learners when they learn (see Section 1.4). Concerning the gestures made by teachers, Hişmanoglu and Hişmanoglu (2008: 170, 175) explain that they “help students to understand, learn or remember new information” and that “the teacher should [therefore] make broad hand and arm gestures, which stimulates students to participate in learning and encourages them to feel safe in the classroom”. Goldin-Meadow & Alibali (2013: 17) support this idea too as they declare that the “gestures teachers produce during their lessons matter for students' learning”. Rosborough (2014: 243) concludes from his observations that gestures are indeed important to analyse from a pedagogical perspective, and he adds that it is as true for the teacher as it is for learners: through gestures, they optimise each other's understanding and communicative possibilities in the target language. Gregersen and MacIntyre (2017) draw the same conclusion as they specify that “gestural input” (i.e. teacher's gestures) is as crucial as “gestural output” (i.e. learner's gestures) and this for two main reasons: input helps the learner feel involved and output regulates the learner's communication. Concerning the gestures made by learners, researchers have been focussing especially on two aspects. The first aspect indicates that gestures could support and positively impact the L2 acquisition process and the second aspect suggests that the use of gestures in a context of L2 acquisition could be influenced by various factors.

Gesture studies have shown that learners' use of gestures is essential for them to acquire the target language. According to Macedonia and von Kriegstein (2012: 7), this can be due to the fact that combining a gesture with a word when learning it apparently leaves a "motor trace" in the brain, engraving the word in the memory and enabling a faster retrieval whenever the learner wants to use this word again. According to Gregersen and MacIntyre (2017: 41), learners gesture more when they are talking in the language which they are acquiring than when they are talking in their mother tongue. A reason for this is to be found in the work of Gullberg (1998), who suggests that gestures can compensate a lack of verbal knowledge, which would justify the higher frequency of gestures in L2 than in L1. Gesturing would help the learners convey pieces of information which they cannot (yet) express through words, hence reducing the potential communication problems which arise from a lack of verbal knowledge in the target language. Gullberg et al. (2008: 22) also mention that "gestures help reduce cognitive load". This notion of cognitive load has been proposed by many researchers and refers to the fact that when multiple tasks need to be simultaneously performed, some can burden the brain. The cognitive load theory is applied to the learning process by Kirschner (2016), who explains that better understanding how the brain works and which activities help and/or prevent the learning process can lead to an optimisation of the learning activity: some types of load can indeed be useful as they lead for instance to automatization, but some can make the learning process harder by adding irrelevant or misleading information. Various studies claim that performing gestures can be one of those aspect that relieves the cognitive system, hence optimising the learning activity and thus probably also L2 acquisition.

Some analyses around gestures have also pointed out that, like in the language acquisition processes, the use of gestures could vary when performed in a L2 context. It seems indeed that many variables can impact the way people gesture in a L1 or a L2 and some individual and cultural characteristics appear to be of great influence. While Gullberg (1998: 229) could not find any significant variation in the frequency of gestures made by learners from two different cultures (Sweden and French people), she could nonetheless point out that the types of gestures used by the two groups of participants presented cultural variation. Gullberg et al. (2008: 13) add that some patterns specific to the mother tongue of most learners are frequently found in the way they use gestures in their L2: the gesturing habits that learners have acquired with their L1 seem to influence the way they use gestures in their L2. This transfer from L1 gestures to L2 gestures recalls what has been said of language acquisition (see Section 2.2), namely that the first language can, positively or negatively, influence the acquisition of a

L2. Studies have also shown that the use of gestures in L1 and in L2 in general is different. For instance, Gullberg (1998: 299) observes that gestures are “more speech-associated” when learners talk in the target language than in their mother tongue. Graziano and Gullberg (2018: 13-14) also suggest that the L2 learners tend to gesture differently from the L1 participants to which they were compared. For instance, they apparently tend to use a lot of referential gestures “to elicit lexical help from the interlocutor” (Graziano and Gullberg 2018: 13). These gestures which are made by learners when they do not find the correct word in the L2 are also observed by Stam (2001: 271), who says that learners can either gesture to retrieve the word on their own or they can gesture to manifest their need for external help. Depending on the intention of the learners, the type of gesture they use can vary. Learners tend to use a lot of iconic and beat gestures, but also many butterworths when they are in this situation of word-retrieval (Stam 2001: 272; Stam 2012: 62). Extending the observation to the gestures made throughout SLA and not only when words are lacking, Gullberg (1998: 229) adds that learners tend to use a large number of metaphorical and deictic gestures. She mentions nevertheless that some individual variations tend to be present in the frequency and in the types of gestures the learners use. She also stresses that the proficiency level of the learners apparently influences the use of gestures (Gullberg 1998: 230). All these findings seem to corroborate the idea that gestures need to be learnt and adapted in the same way as the verbal aspects when acquiring a new language.

2.3.3. Gesturing in the case of bilinguals

A specific case, bridging in some ways L1 and L2 acquisition, is that of bilinguals. According to Goring (2016: 4-5), bilinguals tend to use more gestures than monolinguals, no matter the type of bilingualism, be it acquired as a child or later as an adult. Still, she adds that drawing generalisations about the way they gesture in the one or the other language is more complicated as some bilinguals seem to gesture more in their L1 and others in their L2. She suggests however that the type of language combination could influence the way bilinguals use gestures. Indeed, she explains that some languages have been defined as high- or low-gesture languages (see Section 1.3) and that the way people use gestures in their L1 apparently influences the gesturing habits in L2 (see Section 2.3.2), which could justify why some bilinguals tend to use more gestures in their L1 or in their L2 (Goring 2016: 4-5). The results of Nicoladis et al. (2018: 659) confirm that the languages which are mastered by a bilingual person have an impact on the way this person gestures. Indeed, they show that “Chinese-English and Hindi-English bilinguals gestured less often than French-English or Spanish-English bilinguals”. Once more, individual and cultural factors seem to play a role in the use of gestures.

Despite the existing variations between the bilinguals, all of them proportionally use more gestures than monolinguals. An explanation is suggested by Goring (2016): combining the hypothesis that gesturing enables word-retrieval and the hypothesis that bilinguals constantly have access to two languages (and therefore all their words), she suggests that bilinguals have a greater panel of words between which they need to choose when producing an utterance. This might complexify the process of word-retrieval and accordingly explain their higher need for gesturing. Bilinguals indeed master two languages at a level which enables them to constantly have access to both languages and all the respective features. The potentially negative aspect of this is that they therefore constantly need to make their brain work double in order to choose the accurate word from twice as many possibilities as is the case for monolinguals. Analysing the use of gestures in the case of bilingual speakers illustrates in itself the importance of gestures in both L1 and L2 acquisition and also exemplifies the crucial role of gestures both related to the communication process and to the thinking process.

3. Corpus linguistics

The current study will be based on an analysis of videos which have been downloaded from NESSI (or *New Englishes Student Interviews*), a multimodal corpus compiling “spoken data produced by young speakers of New Englishes”¹. It seems thus important to concretely explain what corpora are, and more specifically, what corpus linguistics consists in. Once the different types of data which can be found in corpora will have been established, a separate and more detailed section will be dedicated to multimodal corpora. In this last point, the distinctive character of these corpora will be developed. As they contain other types of data than the more traditional corpora in which researchers search for written words (be it an originally written text, such as a news article or a story, or a transcript of originally spoken data), the methods of annotation are often different too. Some of the annotation issues which are faced when analysing multimodal corpora will be pointed out. Finally, the annotation systems which have been suggested by some researchers who have been working on multimodal corpora will be compared and will serve as a basis for the current analysis.

3.1. Corpus: definition

The word ‘corpus’ is originally a Latin word – which explains why the plural is ‘corpora’ – and literally means ‘body’. Nowadays, corpora are widely used in linguistics. McEnery et al. (2006: 4-5), after defining a corpus as “a body of naturally occurring language”, identifies four key features: machine-readability, authenticity, sampling and representativeness. In order to be used for linguistic purposes, a corpus must be machine-readable, which means the analysed data should be in an electronic format and should be analysed via software specifically designed for corpus analyses. This has not always been the case, however. Stubbs (2004: 109) suggests that “corpus study goes back hundreds of years” and McEnery et al. (2006: 3) trace the evolution of corpora, from the “shoebox” corpora to the first modern corpus (the Brown corpus in the 1960s) and afterwards. Both writers point out the prominent role that technological advances (computers and specialised software) have played in this evolution, making it possible to collect a high number of data and to make the analysis much easier and faster. This is why, nowadays, machine-readability is a key feature of corpora. The second key feature is authenticity. What is interesting in corpus analyses is that researchers have the opportunity to work with *real* language. This is what differentiates corpus linguistics from theoretical linguistics, according to Abuczki and Ghazaleh (2013: 86): studies using corpora

¹ <https://uclouvain.be/en/research-institutes/ilc/cecl/nessi.html>

focus on the performance more than on the competence, on what is actually performed more than on what should or is thought to be done by language users. Researchers compile and base their observations on examples of real language use when they work with corpora, which enables an empirical approach (McEnery & Gabrielatos, 2006: 34). The third aspect is that a corpus is a sample, which means that it does not capture the whole set of language use in the world. Thanks to the use of computers, corpora can contain a considerable amount of data – hundreds of millions of words from thousands of texts, according to Stubbs (2004: 106) – but a corpus is automatically limited as it is strictly impossible to record and stock every language use occurrence in the world, no matter how large the corpus is. The sample will therefore be collected following certain criteria, often depending on the immediate goal of the researcher. This sample should be representative of a certain group and/or population so that linguists can draw relevant conclusions and generalisations from their observation. Representativeness is also the last of the four criteria which are used to define corpora. As Allwood (2014: 214) explains, the aim of the study will determine which type of data will be used: some researchers might want to study or compare the language use of a specific part of the population whereas others might prefer to draw conclusions for the whole population. Because corpora can be used for more than one study, it is important to collect data from various types of people so that researchers can choose what they focus on for their own observation. This means that a corpus should contain language evidence produced by people of different genders (not only men, for instance), different ages (not only teenagers), different socio-economic backgrounds, and many other potential variables. To summarise these criteria, we can say that a corpus is a large but limited collection of computerised data representing the authentic language use (written or spoken) of a varied spectrum of users.

3.2. Corpus linguistics: utility and status

There have been pros and cons regarding the use of corpora for linguistic purposes, but as Gilquin et al. (2007: 319) suggest, it has “secured a key role in most language-related fields”. McEnery and Gabrielatos (2006: 33) come to the same conclusion when they acknowledge that the use of corpora in linguistics creates discussions and debates among researchers, but that all of them agree that corpora have enabled great discoveries in various linguistic domains, especially in English, because many corpora which have been created so far collect evidence of this language in particular. Davidse et al. (2014: 1) explain this success of corpus linguistics by the fact that it brings to light “new types of information”. This new approach to linguistics enables researchers to focus on quantitative evidence, especially observing language

phenomena in terms of frequency (McEnery and Gabrielatos 2006: 35). This does not mean that corpus linguistics prevents researchers from conducting qualitative studies, but it enables them to draw relevant conclusions based on quantitative aspects of the language too. The use of corpora in linguistics has proved relevant and even innovating and interesting in various fields: it enables linguists to discover and to verify some patterns and/or models related to a certain or several language(s); it enables linguists to gain better insight into a language in order to optimise dictionaries or wordlists and language descriptions; it can serve as a basis to create better teaching books, manuals or techniques, especially in the teaching of L2; it can be used to study the changes of a language through time; it enables new approaches to literary criticism (Abuczki and Ghazaleh 2013: 86; Gilquin et al. 2007: 319; McEnery and Gabrielatos 2006: 33; Thompson 2006). As Allwood (2014: 220) summarises: “the limits are set by the creativity and insight of the researchers”.

There are also negative aspects to the use of corpora. Fillmore (1991: 35) highlights some of them when he compares what he calls “armchair” linguists to corpus linguists. He caricatures the old-school armchair linguist, thinking about the language he masters and deducing interesting elements from it, and imagines him talking to the caricatured corpus linguist, focussing on statistics and frequencies to draw reliable though uninteresting conclusions about language. The imagined dialog he creates between the two caricatures may be exaggerated to some extent, but it proves that corpus linguistics has not been adopted by everybody. McEnery et al (2006: 1) also explain that the use of corpora in linguistics has not been praised by all linguists, especially in the early days due to a lack of effectiveness, later compensated by the arrival of computers and specific kinds of software. A famous linguist who criticised the use of corpora was Chomsky. In an interview, he says that “corpus linguistics doesn’t mean anything” and he draws a parallel with other sciences like physics and chemistry, arguing that if they simply collected videos “of things happening in the world” to make generalisations instead of realising experiments themselves, they would not be considered as scientists (Andor 2004: 97). Fillmore (1991: 35) also considered himself as an opponent to corpus linguistics in the beginning of his career, placing himself in the category of “armchair linguists”, but instead of completely rejecting corpus linguistics, he decided to observe it. Two elements arise from his observation: corpora are limited and therefore cannot give information about the whole language, but they also make possible to discover facts which would otherwise have remained unnoticed and unknown. This second observations leads him to conclude that corpus linguistics is useful. He adds however that it should not be the only way of studying

language: the best way is to combine the two different approaches. The same conclusion is to be found in the work of Inaki and Okita (2006: 293) who aim at proving the “effectiveness and validity” of combining corpus-based and non-corpus-based approaches. Granger (2002: 4) also sees corpora as “one source of evidence among many, complementing rather than replacing other data sources”. Despite it not being the only approach to linguistics, using corpora is indeed now seen by the majority of linguists as valuable and relevant.

Although most researchers have acknowledged that corpora can be of great value in various – if not all – fields related to language, the status of corpus linguistics is sometimes still unclear. Some prefer considering it as a theory, namely a distinct discipline or a new branch of linguistics, whereas some consider it as a methodology, namely a tool or an innovative way of ‘doing’ linguistics. Stubbs (2004: 107) opts for the latter when he says that corpus linguistics is “not a branch of linguistics” but “an empirical approach”. The same conclusion is drawn by Granger (2002: 4), who clearly defines it as “a linguistic methodology”, specifying that in spite of it not being a theory, it is “a particularly powerful methodology”. The point of view of McEnery and Gabrielatos (2006: 35) is similar as they also stress the methodological character of corpus linguistics. They add, however, that the “methodological practices adopted by corpus linguists are not uniform” and that corpus linguistics is still, implicitly or explicitly, “influenced by the theoretical orientation of the researchers” (McEnery and Gabrielatos 2006: 44). Corpus linguistics is thus widely considered as an approach to linguistics, a method in order to provide new and valuable information in the existing (sub)disciplines. There are two main axes which can be followed in order to link this methodology to linguistic theories: the corpus-based approach and the corpus-driven approach. From the explanation of McEnery and Gabrielatos (2006: 35-36), it can be summarised that the corpus-based approach is more deductive, trying to verify a hypothesis or a theory via the observation of corpus data, whereas the corpus-driven approach is more inductive, starting from the observation of corpus data in order to develop a hypothesis or a theory. Corpus linguistics can therefore be seen as a major breakthrough in linguistics, providing a new way of observing and studying languages in order to verify and/or develop reliable theories.

3.3. Type of data in corpora

If all corpora share the same basic features, some differences exist between them. Corpora can contain data in only one language (monolingual) or in several languages (multilingual); language elements collected during one specific period of time (synchronic) or through time (diachronic); language elements produced by native speakers or by learners; language elements specifically used in a particular area such as the medical jargon (specialised) or used in various contexts (general). It is also possible to compile a corpus with a fixed amount of collected data (static) or to open its limits by constantly adding new data (dynamic). Another interesting aspect which needs to be observed is the type of material which is collected. The corpus can be a collection of originally written texts, from one specific genre or from different genres, but it can also be a collection of originally spoken data, transcribed with or without access to the original audio data. It can also be a mix of originally written texts and transcriptions of originally spoken evidence of language use. A third interesting type of corpus is what is called *multimodal corpus* (see Section 3.4), which consists in a collection of verbal and non-verbal data, often videos of people talking. This diversity in the types of corpora which are available needs to be taken into account when researchers start their study: the choice of the corpus is a crucial step in a corpus study process. The objective linguists follow should be clear and fixed before they start their analysis, so that they can relevantly select what kind of data they need to observe and, consequently, what kind of corpus they will need to use. Depending on what information is needed for the study, researchers can decide to use an existing corpus or to combine several of them. It is also possible to compile a new corpus specifically dedicated to the research, but it is often impossible due to time and/or money shortage, which leads many researchers to use existing corpora to conduct their studies.

When or before analysing the data from the selected corpus or corpora, researchers may want to know more about what they are observing. This is why another element of great value has to be added regarding corpus studies, i.e. metadata. Metadata give information about the context in which the data have been produced. It is, in other words, ‘data about data’ and it is seen as crucial by many linguists. McEnery et al (2006: 22) explain that metadata can give information about three main aspects: language product, language producer and language production. Examples of information related to the language product which is usually found in the metadata can be the text genre or the publication date. Regarding the language producer, valuable information could be the gender or the age of the writer or speaker. Finally, elements can be specified in the metadata about the language production itself, be it about the degree of

formality or the language variety. There are many possibilities and while this kind of data has not always been regarded as relevant, it has often proved useful. One of the reasons why metadata should be collected and made available to researchers lies in the very definition of a corpus. It has indeed been previously discussed that two defining aspects of corpora are sampling and representativeness. In order to select data on a reliable basis, some information about who produced the language evidence or where and when the data has been collected may therefore be decisive, and this kind of information is what linguists can find in the metadata. However, data and metadata should not be made available in the same place in the corpus. McEnery et al. (2006: 22) indeed insist on the importance of collecting and providing metadata and data simultaneously, but “separately from the corpus data *per se*”. Providing information about data via metadata can be considered as important as the data themselves.

3.4. A specific case: multimodal corpora

3.4.1. Multimodal corpora: some key elements

It is obvious from the term *multimodal* itself that this kind of corpus refers to something containing more than one modality. However, as Vandelanotte and Dancygier (2017: 1) report, “multimodality remains an elusive concept” for many as it “keeps on being re-defined”. Allwood (2014: 207-208) states that multimodality occurs when two modalities are simultaneously present: these modalities can belong to the “sensory modality” class (namely the five senses) or to the “production modality” class (he gives inter alia the examples of speech and gestures). He also insists on the distinction between modality and medium, which implies the distinction between multimodality and ‘multimediality’. He exemplifies this distinction by defining “a ‘communication medium’ as the physical carrier of multimodal information” (Allwood 2014: 208). In other words, communication is made possible by different physical media (sound or light waves among others) which physically carry the information made available through different modalities (here hearing or seeing). The very nature of human communication lies in this multimodality, according to Abuczki and Ghazaleh (2013: 87), both for the producer and for the receiver. Multimodality is therefore a feature which appears inherent in human communication and this is why it should not be overlooked when studying languages and language use.

One of the main aspects which distinguishes multimodal corpora from the others lies in the nature of the data they contain. Most of the other corpora contain evidence of written language. It is an important distinction because, as Allwood (2014: 212) suggests, speech and

gestures have an ephemeral nature whereas written texts are permanent. Written traces are tangible, which means they remain observable once they have been produced. This is not the case for gesture nor speech since once a word has been uttered or a gesture has been performed, it disappears. Therefore, researchers need to find a way to fix evidence of spoken human interaction, and to keep a trace of it. This is why researchers need to capture the ephemeral (and inherently multimodal) character of face-to-face human communication and the easiest way to do so is by recording human interactions on videos. Once the data have been recorded, the ephemeral nature of face-to-face human interactions disappears, allowing linguists to observe them and learn more about all the aspects of language.

Many linguists agree on the importance of multimodality and it is also why multimodal corpora have been created. Allwood (2014: 218) states that, even if theoretically all modalities could be used conjointly to create multimodality, only two are frequently used in the context of corpora: hearing and vision (sensory modality), or speech and gestures (production modality). The easiest way to compile these two modalities is to film people, making it possible to hear but also see the conversation, and this is also why multimodal corpora usually consist in a kind of video databank. Only compiling video does not seem enough, however. In their work, based on other research and studies, Abuczki and Ghazaleh (2013: 88) state that annotation should assume an important place too. This can be done either by adding a transcription of the video speech in which information about the gestures are simultaneously given, or by transcribing and annotating the audio and the video files as two distinct elements. In the explanation of their study on gesture alignment, Hahn and Rieser (2010: 99) also give an important status to annotation as they specify that they have “completely annotated” the multimodal corpus they have created, but it is interesting to notice that they have only annotated the kind of gestures they want to analyse in their study instead of annotating every movement possible. One could wonder whether the annotation of the corpus is seen by the different researchers as a defining feature of multimodal corpora, implying that an unannotated collection of videos could not be considered as a multimodal corpus, or whether multimodal corpora can be unannotated, implying that the annotation comes after as a part of the corpus analysis. An interesting approach to this issue can be found in the work of Allwood (2014: 207), who explains that there is not one but several possible definitions of multimodal corpora. The narrowest definition he gives requires “that the audiovisual material should be accompanied by transcriptions and annotations or codings based on the material”, which coincides with the annotation necessity of Abuczki and Ghazaleh (2013) (Allwood 2014: 207). Yet, Allwood (2014: 207) explains that

this transcription and annotation necessity is not present in all the multimodal definitions and, according to him, a corpus can be considered as multimodal once it refers to a “digitized (...) collection of language and communication-related material drawing on more than one (...) modality”. An interesting approach could be to consider multimodal corpora in the same way as other types of corpora, only replacing the written data by audio-visual ones.

3.4.2. Annotation issues in multimodal corpora

Be it considered as inherent in the multimodal corpus definition itself or as one of the many steps in the research process, annotation is seen as a crucial part in order to study multimodal corpora. It makes indeed possible to extract specific information from the whole set of multimodal data and to transform it in a written format (be it a coding system, with numbers and/or letters, or be it a ‘verbalisation’ of the data, writing with words what is seen or heard). This allow researchers to analyse their multimodal data with a quantitative as well as a qualitative approach, in the same way as what is done with other corpora.

What is important to remember, however, is the subjectivity which is automatically present in any annotation. As Allwood (2014: 218) summarises, “any standard of transcription or schema of annotation/coding is dependent on a theoretical perspective and objective”. Even though some attempts are sometimes made to lessen this subjectivity, it is impossible to create an annotation type free from any previous theoretical orientation or intuition. Leech (1997: 6-8) provides some key resolutions that linguists should remember when they start their annotation in order to avoid orienting their observation too much. According to him, the unannotated corpus must always be possible to obtain next to the annotated one. Once the corpus has been annotated, it must be possible to extract the annotation from it and explanation about the annotation system used must always be provided. He finishes by adding that this annotation system must be presented as a practical tool and not as a standard to follow and it should ideally not be based on any specific theory. Also, two statistic elements can be used to calculate and verify the reliability of a result: *inter-rater* and *intra-rater reliability*. Inter-rater reliability refers to “the extent to which two independent parties (...) arrive at matching conclusions”², which means that it is possible to ask several researchers to conduct the same study in order to compare their results and refine the reliability of the final results. Intra-rater reliability refers to “the extent to which a single individual (...) constantly produces the same

² <https://medical-dictionary.thefreedictionary.com/interrater+reliability>

results while examining a single set of data”³, which means that the results of an experiment made by only one person can be reliable if this person follows a routine and, for instance, always annotates the same types of gestures in the same way. Annotating is thus necessary in order to correctly observe and study the data in a multimodal corpus, but it also potentially biases, or at least influences, the data analysis depending on the theory or theories behind the annotation system, which is why rigorously following the pre-set annotation system is essential in order to ensure reliability.

Furthermore, annotating multimodal corpora is not an easy task. Two main reasons for this are given by Abuczki and Ghazaleh (2013: 87). The first reason lies in the time it takes to annotate a multimodal corpus. Some annotation tools exist and it is possible to (partly) automatize speech and/or gesture recognition via specifically designed software, but manual verification is unavoidable (Allwood 2014: 217). Annotating data is always time-consuming but, once more, the nature of multimodal data makes it even more time-consuming. Indeed, if all data had to be fully annotated, it would mean that, for instance, each visual aspect, from the facial mimics to the corporal expressions, should be made clear through an annotation system in order to be analysed. As Brunner et al. (2017) say, it is impossible for linguists to include each and every possible aspect in their annotation of a multimodal corpus and this is why there is a need for selection. Researchers should clarify what they want to observe, and therefore what they need to annotate, before actually starting with their annotation. The second difficulty suggested by Abuczki and Ghazaleh (2013: 87) comes from the apparent absence of standard regarding the annotation procedure in multimodal corpora. It might be a consequence of the first limitation: because of the impossibility to annotate everything, linguists have to orientate their annotation towards what they want to investigate and will therefore not be subject to the same annotation necessities. Still, as Allwood (2014: 218) points out, even for studies which have been conducted on the same subjects, like gestures for instance, many different annotation systems have been proposed. Annotating multimodal data is thus time-taking, which forces linguists to refine their criteria and select which phenomena they annotate. Moreover, a standardised annotation system is yet to be created, which makes it harder to bring differently annotated data together.

³ <https://medical-dictionary.thefreedictionary.com/intrarater+reliability>

3.4.3. Annotating gestures in multimodal corpora

An element which comes back in various studies and annotation systems is the division of gestures into smaller parts. The idea of segmenting gestures in order to better analyse them has been suggested and developed by Kendon (2004: 108-126). According to him, each gesture, or *gesture unit*, can be subdivided into four main steps. There is first the *preparation*, when the hands leave a neutral position, called *position of rest or relaxation*. Then, the gesture is actually performed and the hands move during what is called the *stroke*. The stroke is sometimes followed by a *post-stroke hold*, namely a short pause, when the gesture has reached its final position and is kept unmoving. The meaning of a gesture is conveyed by the stroke and, if there is one, by the hold. Finally, the hands come back to their original (or rest) position: this is the *recovery*. Because it is sometimes possible to perform several gestures before coming back to the rest position, the concept of *gesture phrase* is also added in this gesture segmentation. This includes the preparation, the stroke and the hold, but not the recovery, which marks the end of the entire gesture unit. One gesture unit can thus sometimes contain more than one gesture phrase. This segmentation of gesture is often used, totally or partially, by researchers when they explain their annotation system and methodology. For instance, in the annotation system of Hu et al. (2016), the stroke is the signal for the beginning of the gesture and they start their annotation by writing its beginning time between square brackets. This is followed by the gesture name, by information about the hand use, and finally by the stroke duration, again between square brackets. An example of annotation they give is: “([1.90s]Cup, RH [0.46s])”, which refers to a gesture performed at 1.90 seconds of the video with the right hand (RH) forming a cup, the stroke lasting 0.46 seconds in total (Hu et al. 2016: 3451). Rosborough (2014: 249-250) also identifies the stroke in bold in his annotation. The particularity of his system is that the gestures are indicated in the transcription of the speech, and then developed in italics under the transcription. An example he gives is: “Okay. [So 3 **hands**]”, which means that the participant gestured while saying “So 3 hands” and that its stroke was performed when saying the word “hand” (Rosborough 2014: 240). He also adds a full description of the hand movement in italics under his transcription, in this case: “*Points and touches with RH index finger “hands” in the left column*” (ibid.). Tellier et al. (2012: 52) stress that annotating the gestures phrase is important – though complex – because it makes possible to easily synchronise gestures and speech, comparing for instance the gesture stroke and the word uttered at the same exact time.

A second element which is often found in annotation systems is the relationship between speech and gestures. There are different ways to express it. A first example has already been mentioned with the annotation system of Rosborough (2014), who includes the gesture annotation within the speech transcription, which suggests a close relationship between both. Another example is given by Allwood (2014: 211), who speaks about “synchronized alignment”. In his approach, the two elements are first analysed separately and then the different files are combined and synchronised to be compared so that the researchers can “hear and view the part of the interaction” which is analysed (*ibid.*). Because the annotation of gestures is first made independently of the speech transcription, he says, this allows the researchers to adapt their gesture annotation system to the aim of their study. According to him, the relationship between gestures and speech can be done at a later stage, which means that researchers can exclude the reference to speech from their gesture annotation scheme and still take this relationship into account in their observation, at a later stage. A third possibility which expresses the relationship between speech and gestures is to include it directly in the annotation system. This is the suggestion made by Holler and Bettie (2003: 85-86, 99-101), who use a code system to describe their gestures. The binary code (0-1) they use indicates whether the gestures they observe represent information which is present in the speech or not: 0 means that “the information that is contained in the accompanying speech or in the original scene is represented by the (...) gesture” and 1 means that “the information that is contained in the accompanying speech or in the original scene is not represented by the (...) gesture” (Holler and Beattie 2003: 100). They explain that in their previous scheme, which they used in another study, they included a 0.5 value for information which “was considered as ‘unclear’ or ‘ambiguous’” (Holler and Bettie 2003: 86). This value is not present in their new study because all the gestures they observed could clearly be scored with 0 or 1, but they specify that “theoretically it could be (...) less easy to decide” and in that case the 0.5 value could be kept (Holler and Beattie 2003: 99). The relationship between speech and gestures is present in various methodologies, either as a step after the gesture annotation or as an inherent part of the annotation system, including information about gestures in the speech transcription or adding information about the speech in the gesture annotation.

Some annotation schemes describe the observed movements in detail, specifying what body part is moved and how it is moved. This is the approach, for instance, of the CASE Taxonomy for non-verbal elements (2017: 1-4). This system of annotation has been created “to refrain, as far as possible, from interpreting during annotation” (CASE Taxonomy for non-verbal elements 2017: 1). They base the transcription of their gestures (and more generally of

what they call NVE, namely Non-Verbal Elements) on verbs which must be conjugated in the third person present tense. To that main verb, information can be added if it is found relevant, such as the body part. The annotation is written between curly brackets. A list of potential verbs which could be used to describe hand gestures is given and some examples are “{drums fingers}” or “{moves hands in circle}” (CASE Taxonomy for non-verbal elements 2017: 3-4). The annotation system used by Hahn and Rieser (2010: 102) also describes the positions and movements performed during the gesture, detailing for instance the “hand-shape, the wrist-movement or the palm-direction”. The italicised gesture descriptions which are added by Rosborough (2014: 237-241) to his speech transcription also deliver detailed information about how the gestures were performed. Some researchers thus describe gestures in their annotation scheme in such a way that the video support could be replaced by the written description.

Instead of describing the movements in detail, or in addition to the detailed description, many annotation systems are based on the classification of gestures. This might be less theory-neutral, but as annotating is usually done in the context of a study which is based on a certain theory background, attributing the observed gestures to pre-established gesture types can be very useful. This is an element which is included in most annotation systems, but the types of gestures which are considered vary, depending on the theoretical background which is promoted by the annotators. For example, Esteve-Gibert and Prieto (2013: 305) include the pointing gestures, reaching gestures, conventional gestures, emotive gestures and “other gestures”, whereas Tellier et al. (2012: 44) consider deictic, iconic and metaphoric gestures as well as beats, emblems, butterworths, interactive gestures (similar to regulators) and aborted gestures (started but stopped before being fully performed). There are many other classifications of gestures, some of which have already been discussed (see Section 1.2). This element is a crucial part in many annotation schemes and also highlights the fact that standardising annotation schemes is hard because it depends on the theoretical conception and categorisation of gestures.

While some annotation systems are mostly based on one of the given elements (segmentation of gestures, relation to speech, detailed movement description or gesture type classification), some take several of them, if not all, into account in their annotation scheme. This is the case in the study of Colletta et al. (2009), who divide their gesture analysis into five steps. To study gestures, they first use the gesture segmentation suggested by Kendon (2004: 108-126) in order to identify their gestures. Then, they place their gestures into one of the following categories: deictic, representational, performative, framing, discursive, interactive or

word searching. After categorising gestures, they define their relation to speech by suggesting six possibilities: according to them, gestures can reinforce, complement, supplement, integrate, contradict or substitute speech. In the fourth step of their methodology, they situate gestures and speech temporally, expressing whether the gesture is made before, during or after the corresponding speech. As a final step, they describe the gesture form, describing the body part which is used, the trajectory which is followed, the potential hand shape and the movement performed, for instance circular gesture. Colletta et al. (2009) decide to annotate gestures on various levels via the software ELAN (see Section 4.3) because it allows them to analyse the relationship between gestures and speech in detail.

The annotation system which is suggested by Kong et al. (2015) also takes most of the previously given elements into account. They base their annotation system on other studies, but they consider their own annotation system as “more refined for gestural analyses given its independent coding for gesture forms and functions” (Kong et al. 2015: 105). These researchers indeed base their annotation on a dual coding system, made possible by ELAN (see Section 4.3). On the one hand, they annotate gestures based on their categories. Like Colletta et al. (2009), they base their categorisation on different studies in order to fix six gesture types for their own study: iconic, metaphoric and deictic gestures as well as emblems and beats, plus a sixth open category for the gestures they consider as “non-identifiable” (Kong et al. 2015: 101). On the other hand, they annotate gestures according to the function (or relation to speech, in other words) which can be attributed to them. They establish eight possibilities, stressing the fact that most gestures can be attributed more than one function. Their options are: “providing additional information to the message conveyed”, “enhancing the speech content”, “providing alternative means of communication”, “guiding and controlling the flow of speech”, “reinforcing the intonation or prosody of speech”, “assisting lexical retrieval” or “assisting sentence re-construction” (Kong et al. 2015: 101) plus an eighth option when “no specific function” can be found (Kong et al. 2015: 102). They do not describe the gesture movement in detail but they specify that they only consider arm and hand movements. Nor do they refer to gesture stroke, but they explain that “a unit of gesture” in their study corresponds to “the duration from the start of a movement until the hand(s) returned to its resting position”, hence partly including the concept of gesture segmentation (Kong et al. 2015: 100).

4. Data and methodology

4.1. Research questions and hypotheses

The objective of this thesis is to gain better insight into how learners use gestures when they have to produce speech in their L2, depending on their proficiency level. Based on the literature and in the light of what has already been said in different fields about gestures, a pseudo-longitudinal corpus-based study will be carried out. This type of study gives access to the data of a panel of participants from various levels which have been collected on the same moment: this allows the researchers to gain insight into the learners' actual use of language and gestures across time without the need of waiting for the participants to improve their mastering of the language. Instead of observing one participant throughout his or her acquisition process, several participants are observed, each of them representing a stage in the process. In this study, eight interviews from learners between B1 and C2 level will be analysed. In order to conduct the analysis, the following questions will be addressed:

- Does the proficiency level of the learners influence their use of gestures during their speaking turn in terms of frequency?
- Does the proficiency level of the learners influence their use of gestures during their speaking turn in terms of gesture types?
- Is there a relation between the use of gestures and the speech production?

To answer the first question, it is hypothesised that there will be variation in the frequency of gestures used by learners depending on their proficiency level. It is expected that a higher proficiency level – and therefore a better mastering of L2 – will lead to a lower need for gestures because more information could be transmitted via speech. On the contrary, a lower proficiency level is expected to lead the learners to use more gestures, especially in order to palliate a certain lack of linguistic knowledge and/or fluency.

The expected answer to the second question is that the types of gestures used by learners also vary depending on their proficiency level. In that way, *deictic* and *representational* gestures, depicted in the literature as culturally more neutral but semantically relatively loaded, are expected to be very frequent in the beginning and to decrease along with a rising proficiency. It is indeed believed that these gestures might replace words regardless of the language which is spoken, and that learners who do not master a lot of words in the target language might tend to use them a lot, but that once the words are mastered, these gestures

would become less necessary. This hypothesis might not be verified in this study, however, on the one hand because there is no access to real beginners (A1 or A2) and on the other hand because the context of the data analysed in this study (interviewees sitting on a chair in an office) is expected to limit the number of *deictic* gestures occurrences. It is also thought that learners will tend to use more *butterworths*, which are used for word-retrieval, when their proficiency level is lower because they would need more time to retrieve some lexical items in the target language. The use of *self-adaptors* is also considered as important in this study, despite the fact that most previous studies do not take them into account. It is indeed believed that this type of gestures can be used by learners as a way to focus on their thoughts and to gain access to their knowledge of the L2 more easily. These gestures could also be used to help the speakers relax if they are nervous. This is why *self-adaptors* are expected to be mostly found in the data of learners with a lower proficiency level, both because they might need to focus more on their language knowledge in order to actually produce speech, and because they might be more nervous due to a lack of proficiency. *Beats* are believed to be very useful for all learners, because they seem to structure both the speaking and the thinking processes, but since advanced learners are expected to use fewer gestures in general, they are expected to use significantly fewer beats too. *Emblems*, *affect displays* and *regulators* are not expected to be the most frequent gestures, regardless of the proficiency level, because they are expected to be more related to the speech content than to the language knowledge.

The hypothesis for the third question, i.e. the relation between the use of gestures and the speech production, is that the observed gestures will influence the speech produced by the learners. It is expected that the relationship between gestures and speech will be different from gesture type to gesture type, but that all gestures will either support speech (adding or stressing information) or will support the thinking process which leads to speech production. In other words, it is believed that gestures will be used differently depending on the proficiency level of the speaker, but that gesturing while communicating in a L2 will help the learners in their second language acquisition process. Gestures are believed to have a dual function, namely to activate speakers' thinking and speaking processes on the one hand, enabling them to remember some words and structures they know from the target language, and to facilitate listeners' comprehension on the other hand, enabling them to understand some concepts without directly accessing the words in the target language. It is believed that intermediates will tend to use gestures to optimise their access to the language knowledge itself and therefore their speech production and to give access to what they mean despite a potential lack of linguistic

knowledge. Conversely, it is believed that advanced learners will tend to use gestures to structure the way they transmit speech in order to optimise the impact on the listeners' reception in the same way as native speakers could do, and that they will tend to use more "spoken" strategies instead of gestures when they need to organise their thoughts. This would explain the expected variation in the frequency and in the types of gestures which are used by the learners: the advanced learners having access to more vocabulary could use spoken strategies to fill their pauses and would need less gestures to activate their linguistic knowledge.

4.2. The corpus: NESSI

This study is based on the analysis of videos included in a corpus called NESSI, which stands for New Englishes Student Interviews. The project of this corpus, launched by the CECL (Centre for English Corpus Linguistics) in Louvain-la-Neuve, is to record and collect videos from "young speakers of New Englishes"⁴. A specificity of this corpus is that "detailed information [...] about the (socio)linguistic profile"⁵ is made available for each participant. Some general characteristics are given, such as the age or the nationality of the participant, and some language-related elements are also indicated, such as their own or their parents' L1, the language(s) spoken at home or at school, and the language(s) they master beyond their L1. Each interview is conducted following the same pattern: the university students have to perform three tasks, namely talk about a set topic, then hold a free discussion and finally describe a picture. In this corpus, some interviews have been carried out with Chinese university students from Hong Kong specialising in English, who speak English as a L2. These are the interviews from which the data have been selected for the current study. Because the data were not originally recorded for gesture analysis, some of the recording attention points mentioned by Allwood (2014: 216) have somewhat been overlooked. Allwood (2014: 216) indeed suggests routinising the recording sessions, opting for a fixed camera and a single viewpoint englobing as much of the scenes as possible, also indicating that some participant positions or video framings could limit the observation of some kinds of body movements. The data in the NESSI corpus have been recorded using a fixed camera, which is positive as it enables to focus on each participant without changing viewpoint. However, some videos have been closely framed on the participants' face, which limits the observation of some of their gestures. This detail had to be taken into account in order to select the relevant data to use in this study.

⁴ <https://uclouvain.be/en/research-institutes/ilc/cecl/nessi.html>

⁵ Ibid.

4.3. The software: ELAN

It has already been mentioned that annotating multimodal corpora is not an easy task. However, some tools exist which facilitate manual annotation. The software which has been used for the current study is called ELAN⁶. ELAN (or EUDICO Linguistic ANnotator) is a linguistic annotator based on the EUDICO (European Distributed Corpora) project. This project was conceived in 1997 at the Max Planck Institute for Psycholinguistics, The Language Archive, in Nijmegen, in the Netherlands (Russel et al. 2000: 1). The EUDICO project aimed at opening the access of multimodal resources to the scientific community, enlarging and easing the annotation systems and analyses of multimodal data. Therefore, tools were created to enable a theory-neutral approach in order to “allow users to compose their own annotation layers setup, including the definition of new annotation layer types” (Russel et al. 2000: 1). According to Wittenburg et al. (2002: 180), the annotating tool ELAN “has now reached a flexibility and functionality which makes it one of the most advanced tools for multimodal work”. This tool has been used and praised by many other researchers such as Colletta et al. (2009), Abuczki and Gazaleh (2013: 88), Esteve-Gibert and Prieto (2013: 305), Rosborough (2014: 236) or Kong et al. (2015). The main advantages which are attributed to ELAN are that researchers can apply their own annotation system to the software and that it allows a multi-layered annotation. It is indeed possible to create several *tiers* in order to code different elements (for instance speech and gestures) separately but simultaneously. Another interesting aspect of ELAN is that it is constantly enhanced in order to be as efficient as possible, remaining “a tool designed to work locally on personal computers” to stay available to the public (Wittenburg et al. 2006: 4). Moreover, several manuals and YouTube tutorials⁷ which explain how to properly use the software are available in order to ease the utilisation. Once the basic functions have been mastered, the software proves to be very efficient and useful.

4.4. The selection of data

To compare the use of gestures by L2-learners depending on their proficiency level, a pseudo-longitudinal analysis has been carried out. This means that instead of observing the evolution of learners during their learning process, different learners were observed at the same time, each representing a certain proficiency level. In order to avoid any bias, the videos of the Hong Kong students collected in the NESSI corpus were not watched beforehand. Instead, a first selection of the videos was made purely based on the proficiency level of the interviewees.

⁶ <https://archive.mpi.nl/tla/elan>

⁷ <https://www.youtube.com/watch?v=wB9vfp8evEI&list=PLQJhL9BD4W1Rm3Hby44zuVMtgT4lXY7wy&index=2>

The level of each interviewee had previously been rated by three native speakers of British English, based on the CEFR (or Common European Framework of Reference for Languages) descriptors. This “structured set of illustrative ‘can do’ descriptors”⁸ make it possible to define the proficiency level of people based on objective criteria (Council of Europe 2001). These descriptors “specify progressive mastery of each skill”⁹, and language users placed on a continuum, from A1, referring to beginners, to C2, referring to native(-like) speakers. There are 6 levels: A1, A2, B1, B2, C1, C2. Further variation can be added with plus and minus signs, placing for instance B1+ between B1 and B2. The proficiency level often varies depending on the skill which is taken into account, be it receptive (listening or reading) or productive (speaking or writing). However, a general proficiency level can always be attributed which gives an overall impression of the proficiency level which is reached by the participant. This is the score which was established by the three raters to determine the level of each interviewee.

Because the participants in the corpus had each been given three different scores (one per rater), the first step was thus to combine these three values in order to give a mean CEFR score for each interviewee. This mean of the three CEFR scores was calculated based on the methodology of Dumont (2018: 383-384) and Thewissen (2012). The proficiency levels of the recorded data varied between B1 and C2+: there was no available data for beginners, which means no A1 nor A2 scores. Following the previously mentioned methodologies, each level which was present in the data was translated into a number, from 1 for B1 to 11 for C2+. The three ratings per student were codified following this scheme in order to convert the more abstract proficiency levels into concrete numbers to calculate the mean: summing the three numbers and dividing the sum by three. The mean, once obtained, was then reconverted from number to proficiency level. The table used for this conversion was based on the table found in Dumont’s work (2018: 384).

Mean of the three CEFR scores	CEFR proficiency level
1 to 1.4	B1
1.5 to 2.4	B1+
2.5 to 3.4	B2-
3.5 to 4.4	B2
4.5 to 5.4	B2+

⁸ <https://www.coe.int/en/web/common-european-framework-reference-languages/the-cefr-descriptors>

⁹ <https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions>

5.5 to 6.4	C1-
6.5 to 7.4	C1
7.5 to 8.4	C1+
8.5 to 9.4	C2-
9.5 to 10.4	C2
10.5 to 11	C2+

Table 1: Conversion from mean score to CEFR proficiency level (based on Dumont 2018: 384).

After classifying the participants according to their mean score, three were selected per level, except for level C2 for which only two interviewees had been recorded. Priority was given to the candidates who had initially been rated similarly by the three raters because their proficiency level was considered to be the most reliable. If it was not possible to find three interviewees who had been assigned the same score by all raters, candidates whose mean score was a finite number, corresponding precisely to a certain proficiency level, were selected. For instance, a participant whose final score was 4 would be chosen over a participant whose final score was 3.6 or 4.3, even if the final score of these three potential participants resulted in a B2 level (see Table 1). After having given priority to these interviewees, other participants were haphazardly chosen to obtain a selection of three interviewees per level (and two for C2). The videos of the pre-selected interviewees were then watched and two videos were kept for each of the four levels, from B1 to C2. As the videos had not originally been recorded for gesture analysis, only videos in which gestures were fully observable were kept. The participants who were selected for the analysis, based on their proficiency level and on the video framing, were the following ones:

Participant	Rater 1	Rater 2	Rater 3	Mean score	CEFR level
HK026	B2-	B2-	B1	$(3.3.1):3=$ 2.33	B1
HK034	B1+	B2	B1	$(2+3+1):3=$ 2.33	
HK015	B2	B2	B2	$(4+4+4):3=$ 4	B2
HK021	B2	B2	B2	$(4+4+4):3=$ 4	

HK037	C1	B2	C1	$(8+4+8):3=$ 6.67	C1
HK042	C2	B1	C1	$(10+4+7):3=$ 7	
HK009	C2+	C2	C2	$(11+10+10):3=$ 10.33	C2
HK055	C2+	C1	C2	$(11+8+10):3=$ 9.67	

Table 2: Selected participants' mean score and CEFR level.

It is important to know that the score given by the raters to the interviewees was not based on the entire interviews but only on around 5 minutes for each. In order to be consistent in the current analysis and to maximise the validity of the results, the same video excerpts as those rated by the native speakers were annotated and analysed. The video recordings of the selected participants were aligned with the audio files which had been used for the rating and were cut accordingly in order to only keep the relevant part. Then, the videos excerpts were uploaded on ELAN in order to be annotated and analysed.

4.5. The annotation of data

It has been mentioned that some issues could arise with the annotation of multimodal corpora and this is why some elements had to be taken into account before starting the annotation of the data in this study. Three major points which could complicate the annotation were the subjectivity which can arise from annotations, the time-consuming aspect of multimodal data annotation and the lack of standardised annotation system for multimodal data (see Section 3.4.2). Completely erasing the subjective character of the study was impossible, as said in the literature, because the gestures had to be interpreted in some ways in order to be attributed a certain annotation. Still, the subjectivity could be limited by establishing a clear and consistent annotation scheme: all gestures were annotated on the same basis, which made possible to ensure the relevance of the results. The time-consuming aspect could also be reduced by precisely delimiting the types of gestures which would be observed and taken into account, and therefore annotated, to answer the research questions specific to this study. Finally, a solution was found against the lack of standardised annotation scheme by comparing the positive and negative aspects which could be found in the annotation systems and the

methodologies which had previously been used by other researchers (see Section 3.4.3). Based on their studies, a specific annotation scheme could be suggested for this study.

Once the objectives of the thesis had been established (see Section 4.1), a clear definition of what would be considered as a gesture in this thesis could be given, based on the definitions previously suggested by various researchers (see Section 1.1). The movements which were taken into account were thus hand and arm movements (from the fingers to the shoulders) which were produced by the learner during his or her speaking turn and which appeared to play a role in the communication process. This means that some body movements were excluded from the study. Movements performed by the learners when they were listening to the interviewee were not taken into account: learners had to be focussing on their speech production only (when speaking or thinking about their speech) and not on their listening comprehension, for instance, to allow the researcher to come to valid conclusions regarding the role of learners' gestures in their L2 production. Movements from another part of the body, such as the head for instance, were not considered as gestures either. Focussing on the hands was a way of restricting the observations and it was justified by various studies which highlight the strong connection between speech and hand movements (see Section 1.1). Not all the hand movements were considered as gestures, however: some self-touching movements which seemed to have no link with speech production – not used to influence the produced speech nor to influence the thinking process leading to the speech production – were not counted as gestures. This communicative necessity is often an argument for researchers in order to select the movements they include or not in their study (see Section 1.1). However, a distinction should be made here between the gestures which were considered as self-adaptors and the self-touching movements which were excluded from the study. When the learners moved their hands to clearly scratch their head or their nose, for instance, the movement was not counted as a gesture, whereas when the learners touched their hands or fingers while speaking or thinking during their speaking turn, the movement was considered as a self-adaptor (see Section 1.2).

Figure1	Participant HK009 puts her hair back	Self-touching movement	Not a gesture
Figure2	Participant HK026 twists her fingers	Self-adaptor	Gesture

Table 3: Difference between self-touching movement and self-adaptor.



Figure 1: Example of self-touching movement (from participant HK009).

The movement which is represented on Figure1 was performed by one of the C2 learners (participant HK009) at 0.20 minutes of her interview excerpt. It lasted about 2 seconds. This participant moved her hand to put some of her hair back behind her ear. She made this movement while saying “but now”, midden in a sentence and without any pause or hesitation. These are elements which led to the conclusion that this movement did not impact the communication. This movement could be linked to the hypothetical example of a learner scratching his or her head because it was itching (see Section 1.1) which would not be considered as a gesture in this study. Therefore, this movement performed by participant HK009 was not considered as a gesture.

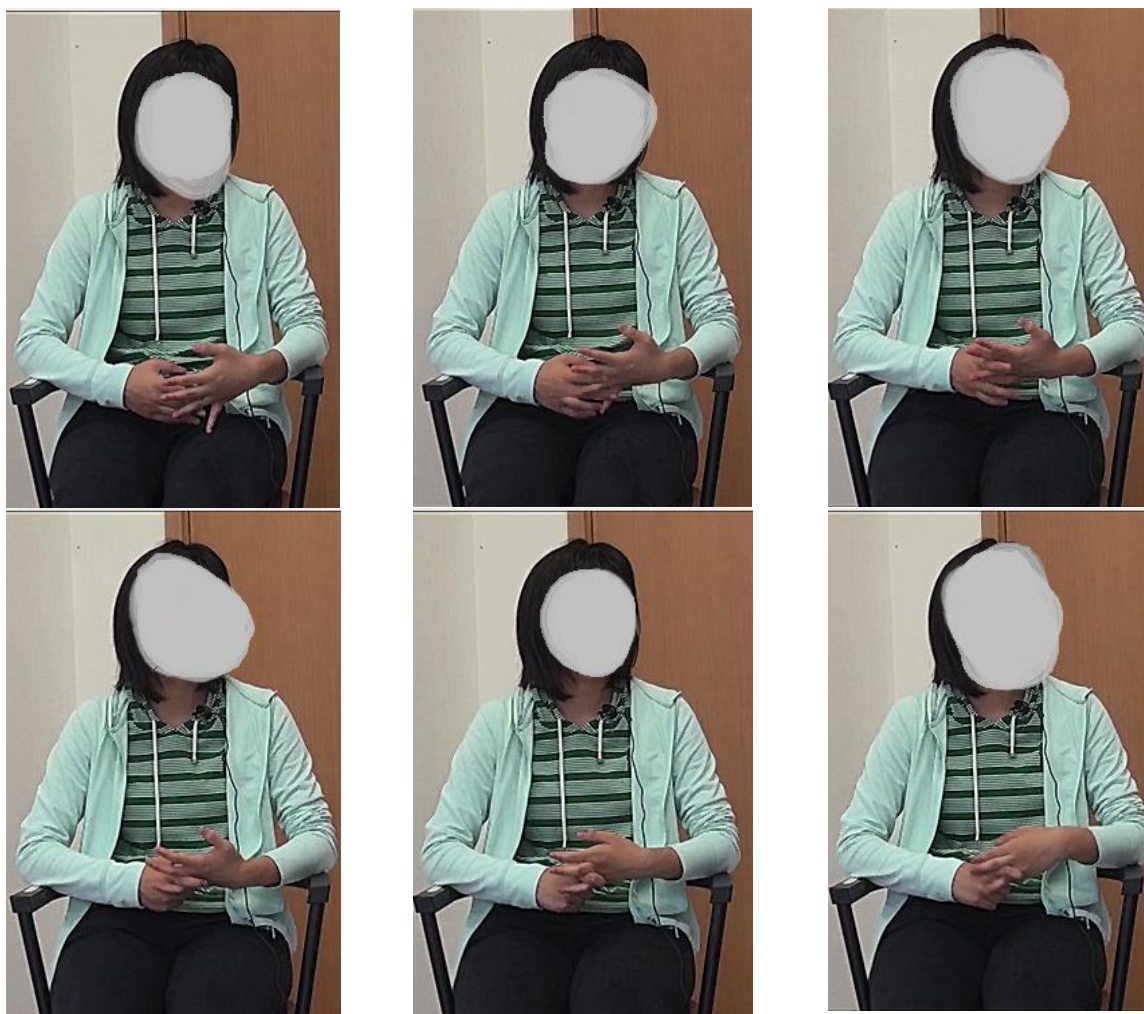


Figure 2: Example of self-adaptor (from participant HK026).

The movement which is represented on Figure2 was performed by one of the B1 learners (participant HK026) at 2.19 minutes of her interview excerpt. It also lasted about 2 seconds. This participant touched her hands and twisted her fingers during these two seconds while saying “aaaand some” (slowing down her speech rate and producing very long vowels). The situation clearly pointed out that the participant was thinking about the speech she had to produce and trying to find how to complete her sentence. This is consistent with the function which has been attributed to self-adaptors by Ekman and Friesen (1969:85), i.e. “facilitate [...] sound-making and speech” (see Section 1.2). It could thus be concluded that this movement possibly had an impact on the speech production and the speaker’s thinking process. Therefore, this movement, performed by participant HK026, was considered as a gesture and more specifically as a self-adaptor.

Because of the lack of annotation standard, various annotation systems and methodologies were analysed and compared (see Section 3.5). A few elements which appeared to be recurrent could be extracted from these different schemes. Kendon's (2004) gestures segmentation theory, which had been used by various researchers in their gesture analyses, was not directly used within the annotation (no direct reference to the stroke was made, for instance) but well implicitly at an earlier stage, for the gesture segmentation. Before annotating and analysing gestures, the researcher must indeed select the gestures which will be observed and delimit them on ELAN, using the segmentation mode. Once the gestures had been segmented, they could be annotated. Two elements were included in the annotation: the gesture type (or types) and its relation to speech.

Indicating the gesture type in the annotation is often a central element in the schemes which have been proposed (see Section 3.5) even though the categories of gestures which are included are not always the same. Eight types of gestures could be selected for this study, based on various existing classifications (see Section 1.2). It was anticipated that affect displays would not be very frequent because they are often performed with other body parts than the hands, but they were added because some topics in the analysed interviews which could be more personal would perhaps require this type of gestures. The beat gestures were included because of their anticipated impact both on the speaking and the thinking processes, punctuating words as well as ideas. The butterworths were added because their direct relation to the thinking process and word-retrieval was considered as relevant for the study. Because deictic gestures depend more on the direct environment than on linguistic capacities, they were also considered as relevant for this study and believed to be useful for learners despite their potentially lower language proficiency. The emblems were included because their culturally dependent and intrinsically meaningful characters were expected to potentially influence the gesturing of learners who had to use another language (and its associated culture). Regulators were added too, because the analysed data consisted in interviews, hence including both the observed interviewee and an interviewer, which could result in direct interactions and turn-takings. Representational gestures were considered as one gesture type, despite the distinction between iconic and metaphorical gestures which is often made in other analyses, because the difference between abstract and concrete representations was sometimes too thin to be made, or considered as irrelevant in practice for this study, though it seemed important to acknowledge this differentiation in theory. This gesture type was added in the study because it has often been defined as directly illustrating speech and appeared thus essential in this study. Finally, self-

adaptors were included in the classification, unlike most gestures studies which often consider them as non-communicational, because analysing their use and also their non-use across the proficiency levels would be valuable in order to gain insight into how and when they tended to be used. It has indeed been suggested that their use could be related to speech production.

After a first gestures selection and observation, it was noticed that the interviewees sometimes seemed to ‘freeze’ their gestures in positions unrelated to their rest position. This observation was supported by a study of Graziano and Gullberg (2018), who count these freezing gestures as holds (hence referring to Kendon’s (2004) segmentation) and who conclude that “holds were significantly more likely to occur with disfluent than fluent speech” and that “L2 speakers were significantly more likely than L1 speakers to produce holds” (Graziano and Gullberg 2018, 9). This justified adding the annotation of *holds* as a separate gesture (or precisely gesture interruption) category. An important distinction has to be made between the interruptions of gestures which were considered as holds and the resting position of the learners which was considered as a non-gesture, or an absence of gestures.

Figure 3	Participant HK015 does not move	Resting position	Not a gesture
Figure 4	Participant HK015 ‘freezes’	Hold	Gesture

Table 4: Difference between resting position and hold.



Figure 3: Example of resting position (from participant HK015).



Figure 4: Example of hold (from participant HK015).

Figure 3 is one of the many examples of a participant in a resting position. The resting position of the learners was often simply to cross their hands on their knees. Figure 3 represents the position of the B2 participant HK015 at 0.35 minutes of her interview excerpt. She kept her hands in that position while continuing her story about joining a choir: “[I joined a] choir before, but then, no I don’t really [...]”. The resting position itself was not considered as a gesture because there was no information which could be inferred from it regarding the communication nor the thinking process leading to speech production. Figure 4, on the contrary, was considered as a gesture (or gesture interruption, namely a hold) because the situation in which this ‘absence of movement’ took place clearly marked the thinking process of the interviewee. Indeed, Figure 4 represents the position of the same participant (HK015) at 1.33 minutes of her interview excerpt. She was gesturing while telling her story and she suddenly stopped her gestures and stayed in this ‘frozen’ position while saying “when I waaaaas” (slowing down her speech rate and producing very long vowels, like participant HK026 during the production of her self-adaptor, which is represented on Figure 2). The situation clearly pointed out that the participant was thinking about the speech she had to produce and trying to find how to complete her sentence. Therefore, this ‘absence of movement’, performed by participant HK026, was considered as a gesture and more specifically as a hold.

Many researchers have, directly or indirectly, included information about the gestures-speech relationship in their annotation system. The approach used by Colletta et al. (2009) and Kong et al. (2015) were inspiring as they also include both the gesture types and the relation to speech in their annotation systems. However, while they opt for a multilevel annotation, using different tiers in ELAN to annotate the two types of information, it was decided to combine everything on one tier in this study. By doing so, the use of gestures could be analysed by focussing on the two elements separately but also conjointly: the relation that every gesture type tended to have with speech could easily be observed. The coding system suggested by Holler and Beattie (2003) was inspiring in this respect. To each gesture type, a code was thus added. The 0-value was added when the gesture was performed without any speech produced. The 1-value was attributed when the gesture was used with speech and apparently matched what was being said. The 2-value was attributed to gestures which were made at the same time as speech but without any apparent or direct connection. The 2-value was considered as more relevant than the 0.5 value suggested by Holler and Beattie (2003) in this study because the meaning behind the values was slightly different. Indeed, in their coding system, 0.5 could be used for gestures situated between 0 (the gesture matches the speech) and 1 (the gesture does

not match the speech), whereas in the current coding system, it was not a question of continuum but well a presentation of three different situations. The studies of Colletta et al. (2009) or Kong et al. (2015) were interesting in order to justify the importance and relevance of adding the function of gestures related to speech, but their categories were considered too detailed for this study: most functions developed by these researchers could be covered by the 1-value (or 0, depending on the context). A coding system with three values seemed efficient to express the relationship between the interviewees' gesturing and speech.

Once it had been made clear what would be considered as gestures and the interviews excerpts had been uploaded on ELAN, the annotation could start. For each interview, a tier was created and named with the code of the participant and then the word *gestures* (for instance: HK009_gestures). The interviews were played and each movement which was considered as a relevant gesture, following the previously mentioned definition, was segmented on the tier, via the *segmentation mode*. The segmentation was first global to highlight the moments when a gesture was observed and then refined in order to exactly match the beginning and the end of the gesture. A gesture started when the hands left their rest position, and stopped either when they came back to it or when the trajectory changed to form a new gesture, as suggested by Kendon (2004). Once all the gestures had been segmented on the tier, each gesture could be watched separately and annotated. The different annotations were pre-set as *controlled vocabulary* so that a list was created and the annotations could be applied to the corresponding gestures. Each annotation was formed as follows:

1. The gesture annotation was delimited by square brackets [].
2. The gestures type was determined: affect displays [af.], beats [be.], butterworths [bu.], deictic [de.], emblem [em.], regulators [rg.], representational [rp.] and self-adaptors [ad.]. The category hold [hold.] was also included.

It has been shown in previous studies (see Section 1.2) that many gestures could be placed in different categories, which was confirmed in the observed data. Moreover, it was possible for some gestures to be ambiguous, so that there was hesitation between two potential types. When a gesture could be given two types, it was encoded as [X/Y], with X being one type and Y being the other, but without hierarchy between the two types.

3. A code number was used to highlight the relation of gestures to speech: 0 for gestures performed without vocalisation; 1 for gestures directly matching the speech; 2 for gestures with no or indirect connections to speech.

<i>code</i>	<i>Link between speech and gestures</i>
0	gestures without speech
1	gestures and speech match
2	gestures and speech without direct connection

A hypothetical example of annotation could be: [rp.0] if the interviewee wants to evoke something in the past and moves a hand to signify “before” without talking; [rp.1] if this gesture is performed by the interviewee while saying something like “when I was in Ireland last year”; [rp.2] if the interviewee says “I am in Ireland” while performing the gesture.

The three following examples are taken from the observed data which have been annotated. Similar examples frequently occurred in the data and it appears important to explain the distinction between the three tags to justify that all the similar gestures were coded with the same annotation.

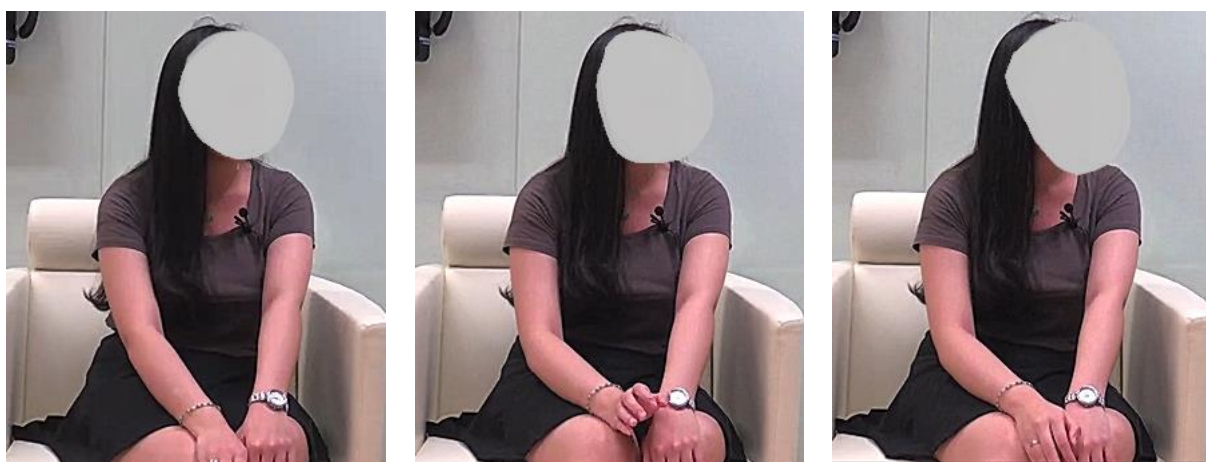


Figure 5a: Example of [be.1] (from participant HK021).

Figure 5a was the first gesture made by the B2 participant HK021 in the interview excerpt which was analysed. This gesture was performed when saying “I’m actually **teaching**”. The word which was being stressed, with the intonation as well as with the gesture, was the word *teaching*. This movement was considered as a beat gesture because it punctuated the speech of the participant. It was attributed the code 1 because the gesture gave the same information as the speech: they were both stressing the same word, and thus matching.

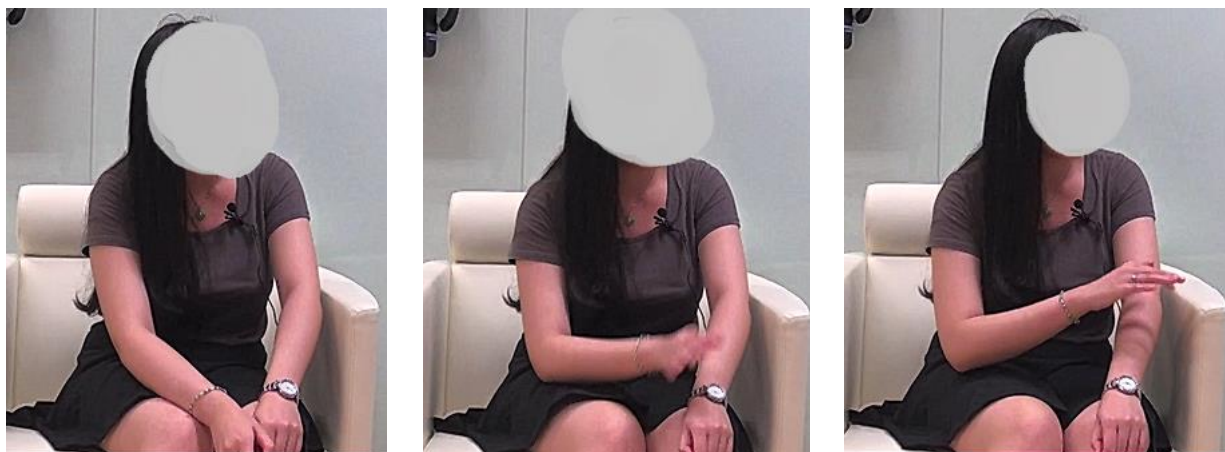


Figure 5b: Example of [be./rp.1] (from participant HK021).

Figure 5b represents a gesture which was also made by participant HK021. She performed this gesture at 0.37 minutes of her interview excerpt while saying “when I was a **kid**”. Again, she stressed a word, here *kid*, with her intonation as well as with her gesture, which is why the gesture was considered as [be.1]. Another dimension was added to this gesture as it did not only stress the word *kid*, but it also illustrated it, by placing a hand as if to show the height of the kid. The gesture was interpreted as representing what was being said and was thus also considered as a representational gesture, matching speech so with the code 1.

The situation represented on Figure 5c was performed by the same participant. She performed this gesture at 2.07 minutes of her interview excerpt, while saying “on the **lyrics**”. The gesture stressed a word, here *lyrics*, and was therefore also considered as [be.1]. Compared to Figure 5a, it appears clear that this gesture expresses more than punctuation but compared to Figure 5b, the gesture does not directly illustrate the word *lyrics*. Actually, the gesture seems to ‘materialise’ the word, as to ‘visually’ place it in front on the listener, without directly representing the ‘thing’. Therefore, this kind of gesture was considered as [be.1/rp.2].

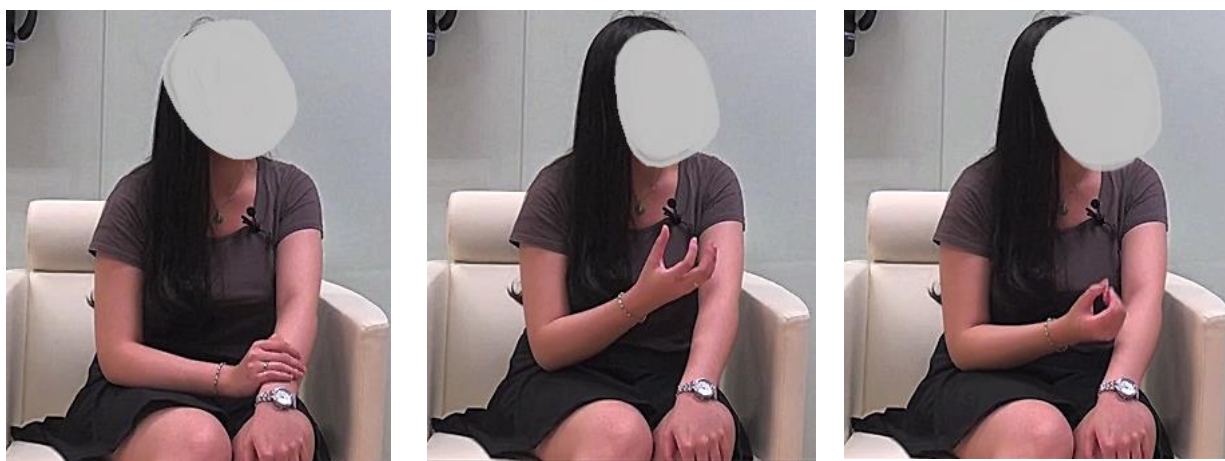


Figure 5c: Example of [be.1/rp.2] (from participant HK021).

5. Results

5.1. Frequency per proficiency level

In order to answer the first research question, namely investigating whether the proficiency level of the learners could influence the frequency of their gesturing or not, two elements were taken into account. The first element was the total number of gestures performed by each participant. The second element was the proportion of minutes in the learners' speaking turn during which they used gestures. All participants were indeed observed on the basis of a fragment of their interview which lasted for about 5 minutes for each of them: comparing the total number of gestures performed by learners seemed thus relevant in order to see if a significant difference or a clear trend could be observed between the levels, because the timing during which gestures were counted was similar. However, it seemed important to avoid drawing conclusions too fast as some participants could have been using short gestures repeatedly but at one specific time in their speech and not having gestured for the rest of their speech whereas others could have been constantly gesturing but using a few but longer gestures. Only comparing the number of gesture occurrences could have led to the conclusion that the first type of participants gestured more, overlooking the fact that they only gestured when talking about one topic and that the second type of participants actually gestured during the whole speech. This is the reason why it was decided to also take the duration of the gestures into account: for each participant, it was thus possible to count the time during which they used gestures and during which they kept their hands still in a resting position when it was their turn to speak. Also, the mean duration of each gesture type used by the participants was calculated. Including these two elements (total number and duration) in the analysis would give a better picture of the gesture use within and between the different proficiency levels.

The first indication, given by the total number of gestures performed, appeared relevant in order to compare the number of gestures which were used by learners when it was their turn to speak in the conversation. Each gesture was counted once, even when it had been attributed different types. For instance, a gesture which had been annotated [be./rp.1] was only counted as one gesture and not as one beat plus one representational gesture. It has indeed been decided to consider each movement as one occurrence instead of counting it twice: it had been given two functions but it remained one single gesture occurrence.

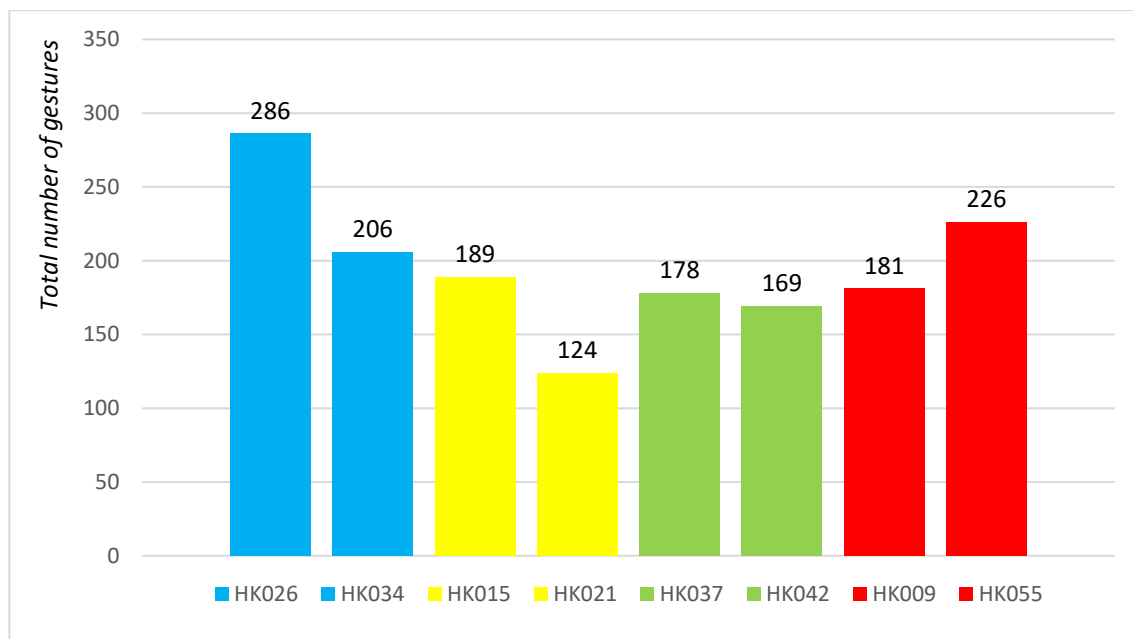


Figure 6a: Total number of gestures per participant.

When comparing the total number of occurrences in the different interviews, it was observed that the number of gestures fluctuated from interviewee to interviewee and that it was difficult to clearly establish any generalisation directly related to the proficiency level. For instance, Figure 6a shows a bigger difference between the two B1 participants (in blue) than between the second B1 participant (HK034) and nearly all the other participants. This result is consistent with many previous studies which stress that various factors can influence the use of gestures and not only the proficiency level (see Section 1.3). It has indeed been shown that the personality of the speaker or the content of the conversation could influence the use of gestures, hence justifying the individual variations which are observed on Figure 6a. However, that the learners' personality or other factors could influence their use of gestures did not mean that the proficiency level was not impacting. This is why, though acknowledging that no complete equivalence could be established between the participants of a same level in this study, the mean number of preformed gestures was calculated for each proficiency level. This made possible to observe if any tendencies could be revealed regarding the number of gestures between the levels.

From Figure 6b can indeed be observed that the mean number of gestures varies depending on the proficiency levels and forms a curve from B1 to C2. This indicates that the number of gestures performed by the participants in this study decreases from B1 to B2 and then increases from B2 to C2. The mean number of gestures performed by the C2 participants is still lower than the mean number of gestures performed by the participants from B1 level.

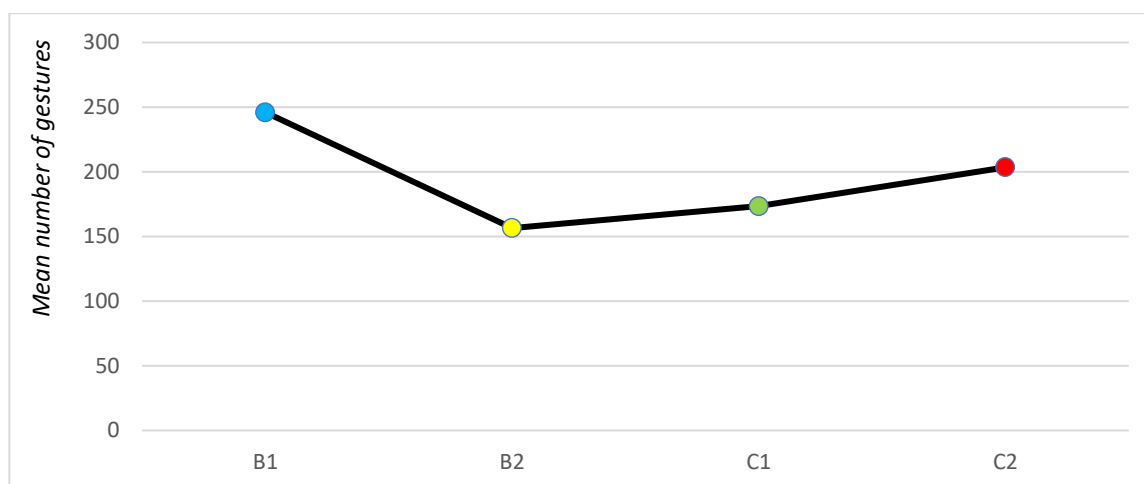


Figure 6b: Mean number of gestures per proficiency level.

A potential explanation could be that gestures have different functions and that they are not used in the same way by the learners, depending on their proficiency level. Gestures have indeed been attributed various functions, two of them directly related to language production: on the one hand, gestures seem to have an impact on the communication process, improving the understanding of the listeners for instance and structuring the speakers' speech production, and on the other hand, they seem to have pedagogical implications, helping the learners in their language learning and supporting the thinking process, for instance for word retrieval (see Section 1.4). This hypothesis could explain the curve which is observable on the graph. Participants from B1 level would have access to less lexical knowledge than the participants with a higher proficiency level and would therefore tend to use more gestures to help them produce speech and transmit ideas correctly (pedagogical function). Participants from B2 level would potentially have access to more language knowledge and would therefore tend to use less gestures to find their words than B1 learners. Moreover, they would perhaps still need to think of their L2 production and would therefore tend to use fewer gestures than advanced learners to focus on the structure of their speech (structuring function). Gestures used to optimise speech transmission would be acquired with language proficiency, which would explain why the use of gestures increases again from B2 to C2 level.

To support this hypothesis, that the gestures used by the participants were not the same depending on their proficiency levels, a comparison with the other results was: the proportion of gesture duration of the participants, the type of gestures they used and the relation their gestures tended to have with the produced speech. These comparisons help to test whether the curve which is observed regarding the total number of gestures is indeed possibly related to the fact that gestures had not the same function when performed by B1 and by C2 participants or whether this curve must be interpreted differently.

The second indication which was observed to determine if the proficiency level of the participants had an influence on the frequency of gestures was given by the gesture duration. If the individual variations tended to be marked between the interviewees of a same level when observing their total number of performed gestures, the period of time during which they used them during their speaking turn was relatively stable within the proficiency levels. For each participant, the exact time of their speaking turn was calculated. This means that because the current study aims at analysing the use of gestures when learners are speaking (or focus on their speech production and not on their listening comprehension for instance) and because the observed videos were interviews, the interviewer's interventions were excluded from the analysis. In the remaining time, corresponding to the participants' speaking turn, the proportion of minutes during which they used gestures was calculated.

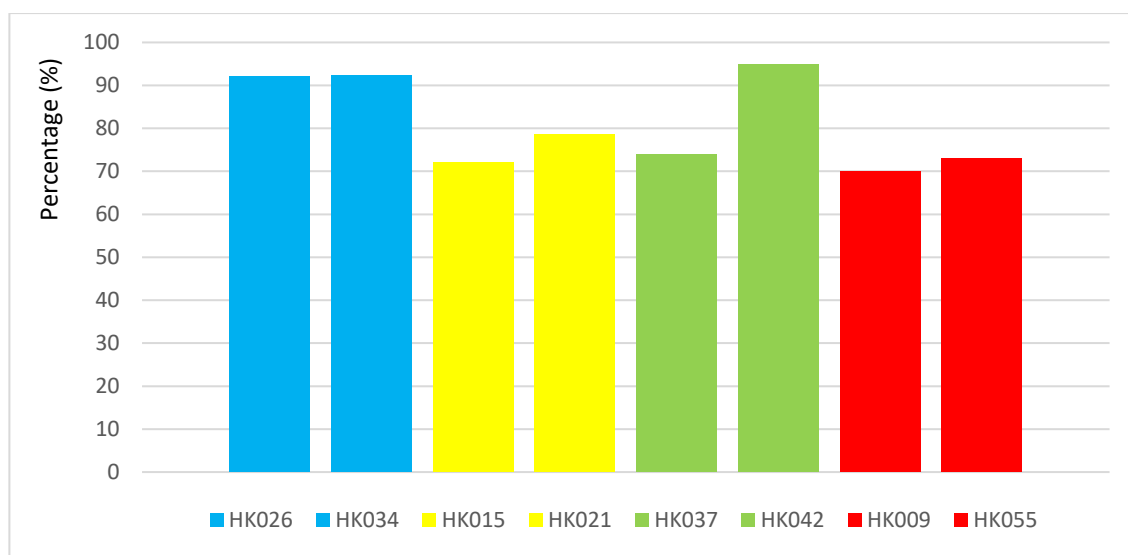


Figure 7a: Percentage of gesturing time per participant.

Contrary to the indications of the total number of gestures, Figure 7a shows a certain stability between the participants of a same level concerning the percentage of their gesturing time. An exception could however be noticed for the participant HK042 (second participant of C1 level, in green). His high gesturing frequency could be justified by the fact that his resting position had most of the time to be considered as self-adaptor, because he constantly moved his fingers. The fact that HK042 was the only male participant could perhaps have influenced his use of gestures, but this hypothesis could not be tested in the present thesis. What can be said, however, is that the results of HK042 often diverge from the general tendencies which the other participants display, as will be further repeated.

Figure 7a also interestingly shows that all participants used gestures during 70% to 95% of their speaking time, which means that gestures are important in all the levels when the learner needs to perform speech. If all participants tend to gesture a lot, the ones in B1 level gesture more than the participants of the other levels as they use gestures during more than 90% of their speaking turn, hence keeping their hands still in a rest position for less than 10% of the time. Again, the mean proportion of gesturing time was calculated in order to draw conclusions on the proportion of gesturing time between the proficiency levels and not only between the participants.

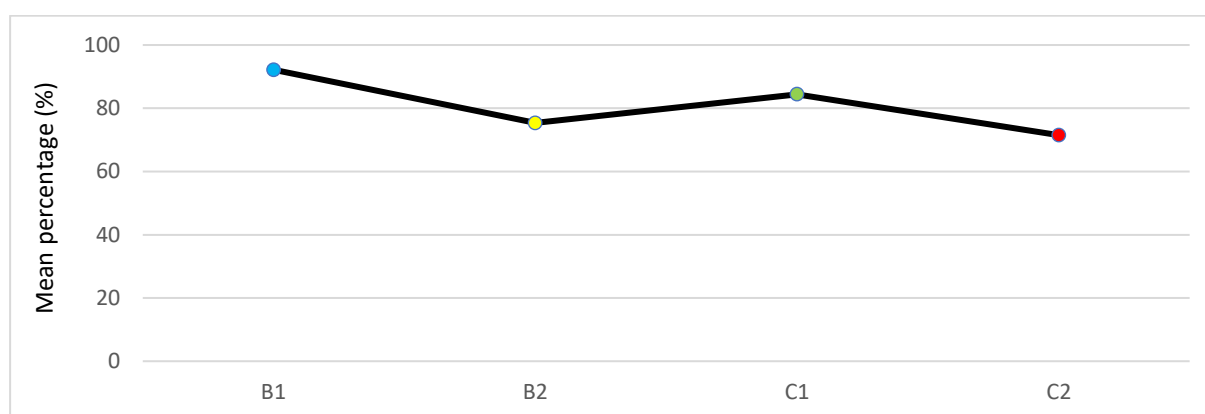


Figure 7b: Mean percentage of gesturing time per proficiency level.

Figure 7b confirms that the proportion of gesturing time of the participants decreases from B1 to C2, although this mean percentage is constantly higher than 70%. The increase between B2 and C1 on the graph could be due to the surprisingly high gesturing time of the second participant, as already mentioned. Based on Figure 7b, it can also be said that there is a split between the percentage of gesturing time for the participants of B1 level and the others: the B1 participants proportionally gesture more than 90% of their speaking time whereas B2 participants proportionally gesture for around 75% of their speaking time, which signifies a

decrease of about 15% whereas the difference between the mean gesturing time of the participants in B2 and of the participants in C2 is less than 5%. Observing the gesturing time of learners with a lower level (A1 and A2) would thus be particularly interesting in order to confirm whether beginners indeed gesture more than advanced learners and whether there is a significant turning point in the gesturing time between B1 and B2 levels. What can be said from the present analysis is that the gesturing time of the participants of a same proficiency level tends to be similar, which was less evident when only considering the total number of gestures made. Indeed, participants keep their hand in a resting position more often when their proficiency level is higher.

The elements of information which can be inferred from the analysis of the number of gestures (Figure 6 a and b) and of the percentage of gesturing time (Figure 7 a and b) are different but complementary. By combining the curved tendency observable on Figure 6b and the decreasing tendency observable on Figure 7b, one could conclude that the participants in this study used gestures differently. It could be concluded that the participants in B1 and C2 levels apparently tended to use more gestures than the participants in B2 and C1 levels, but that the gestures performed by the two B1 participants lasted proportionally longer than the gestures performed by the C2 participants. This hypothesis could be confirmed by calculating the ratio for each participant between their total gesturing time and their total number of gestures:

$$ratio = \frac{\text{total gesturing time in seconds}}{\text{total number of gestures}}$$

Participant	Ratio	Proficiency level
HK026	0,88	B1
HK034	1,17	
HK015	1,00	B2
HK021	1,34	
HK037	1,08	C1
HK042	1,44	
HK009	1,08	C2
HK055	0,77	

Table 5: Ratio between the total gesturing time and the total number of gestures per participant during their speaking turn.

Table 5 shows that the ratio for most participants was around 1 second. Again, the result of the participant HK042 stands out as it was higher than all the other ratios. The mean ratio was calculated in order to compare the ratios between the four proficiency levels. Figure 8 clearly shows that the mean ratio decreases from B1 to C2, once more with an increase from B2 to C1 which could be related to the high ratio calculated for HK042.

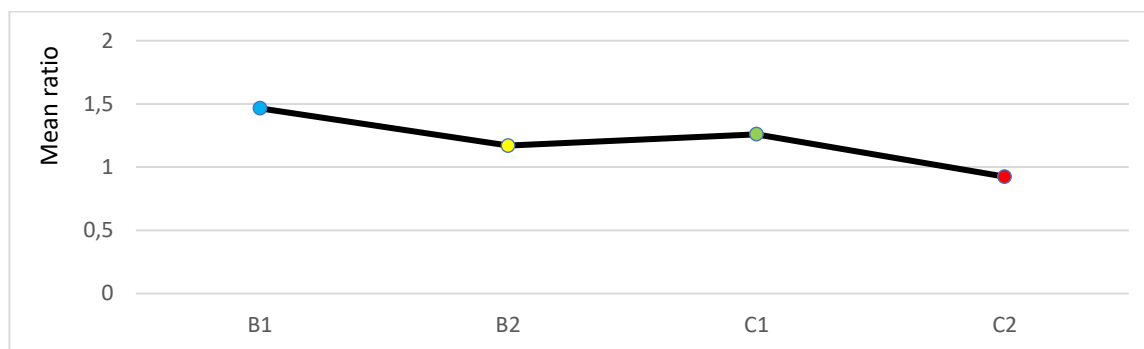


Figure 8: Mean ratio per proficiency level.

The last aspect which could be measured regarding the frequency of gestures and which verifies the hypothesis that the types of gestures which are used could influence the gesturing time was the mean duration of each type of gesture. Some statistical pieces of information were automatically measured on ELAN, the software which was used to annotate the data. It was thus possible to access the total number of occurrences of each annotation token (for instance [be.0], or [be.1], or [be.2]), as well as to the total duration of all the segments which had been annotated with the same token and their mean duration. Other elements were available, but these were the most relevant regarding the frequency of gesturing. By combining the information which was measured by the software for each participant individually, it was possible to calculate the mean duration of each gesture type, regardless of the proficiency level.

It appears clearly from Figure 9 that the mean duration of each type of gesture varies. Most gesture types last around 1 second. This is the case for the gestures which can be considered as beats and representational gestures ([be./rp.]) as well as for the regulators ([rg.]), the holds ([hold.]), the representational gestures alone ([rp.]) and the affect displays ([af.]). Some gesture types do not last as long, namely the deictic gestures ([de.]), the butterworths ([bu.]) and the beats ([be.]), which last on average around 0.7 and 0.8 seconds. The self-adaptors ([ad.]) tend to last significantly longer as their mean duration nearly reaches 2 seconds.

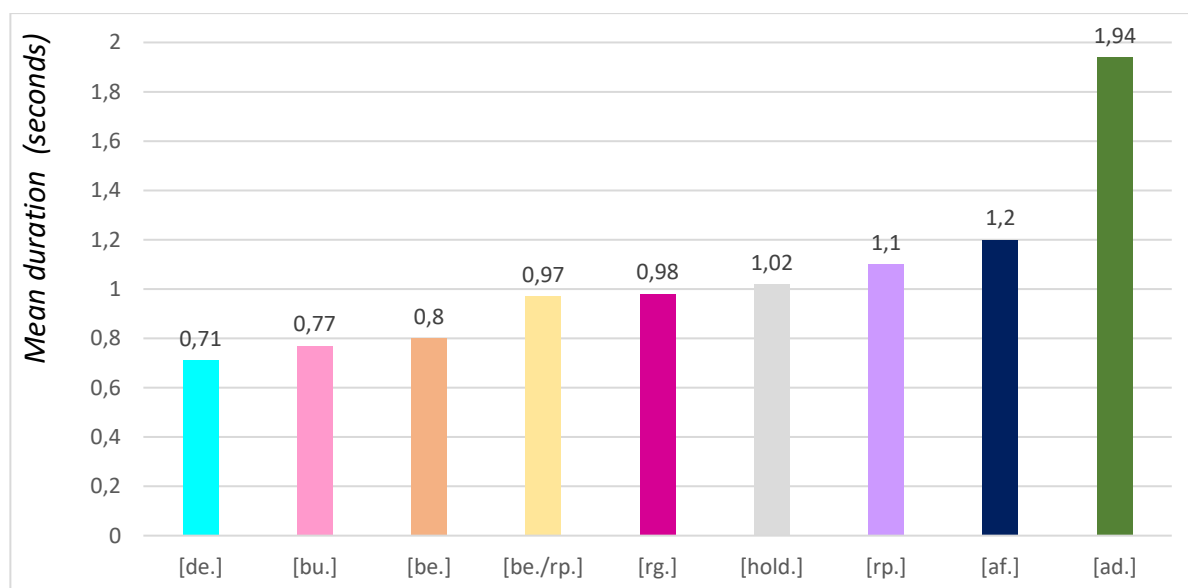


Figure 9: Mean duration per gesture type.

This confirms the hypothesis that the gestures which are used by the participants representing B1 level must be different than the gestures which are used by the participants representing C2 level. Although Figure 6b shows that the total number of gestures used by the participants increases from B2 to C2 after a sharp decrease from B1 to B2, the proportion of gesturing time of the C2 participants, represented on Figure 7b, is smaller than the one of all the other participants, which indicates that the gestures used by the advanced learners must proportionally last shorter. This is confirmed by Figure 8, which shows that the gestures used by the B1 participants last proportionally longer than the gestures in the other levels, and by Figure 9, which confirms that the type of gesture has an influence on its duration.

5.2. Gesture types per proficiency level

Comparing which types of gestures tended to be used by the participants when it was their turn to produce speech made possible to answer the second research question, namely analyse whether the proficiency level of the participants had an influence on the types of gestures they would use or not. Also, it was important to verify whether the results obtained here coincided or not with the results previously obtained regarding the frequency of gestures. From what has been suggested in the literature (see Section 2.3.2) as well as from the results about gesture frequency (see Section 5.1), a variation was expected to be found in the types of gestures which were used between the proficiency levels.

In order to compare which gesture types were used by the participants, the types of gestures performed by each interviewee individually were ranked in decreasing order and then the results of each participant were brought together to be compared. Ranking the types of gestures in decreasing order brought to light which types of gestures tended to be used by the learners. It allowed the researcher to highlight a few gesture types which are prevalent among the proficiency levels. Also, it made possible to gain insight into the diversity of gesture types which were used by the learners. Indeed, some participants could constantly use the same type or types of gestures whereas some others could perhaps tend to use a lot of different types and to vary a lot. Afterwards, bringing the results of all the participants together offered a general overview of the use of the different gesture types within and between the proficiency levels. On the basis of the information which had been extracted, it was possible to identify general tendencies regarding the use of the various gesture types in the proficiency levels. Figure 10 represents the proportion of the gesture types which were used by each participant individually and Figure 11 represents this proportion throughout the levels. A detailed comparison of the mean proportion per proficiency level is then given for some of the gesture types, which were considered as relevant to analyse because they were used by all the participants and/or in a large proportion.

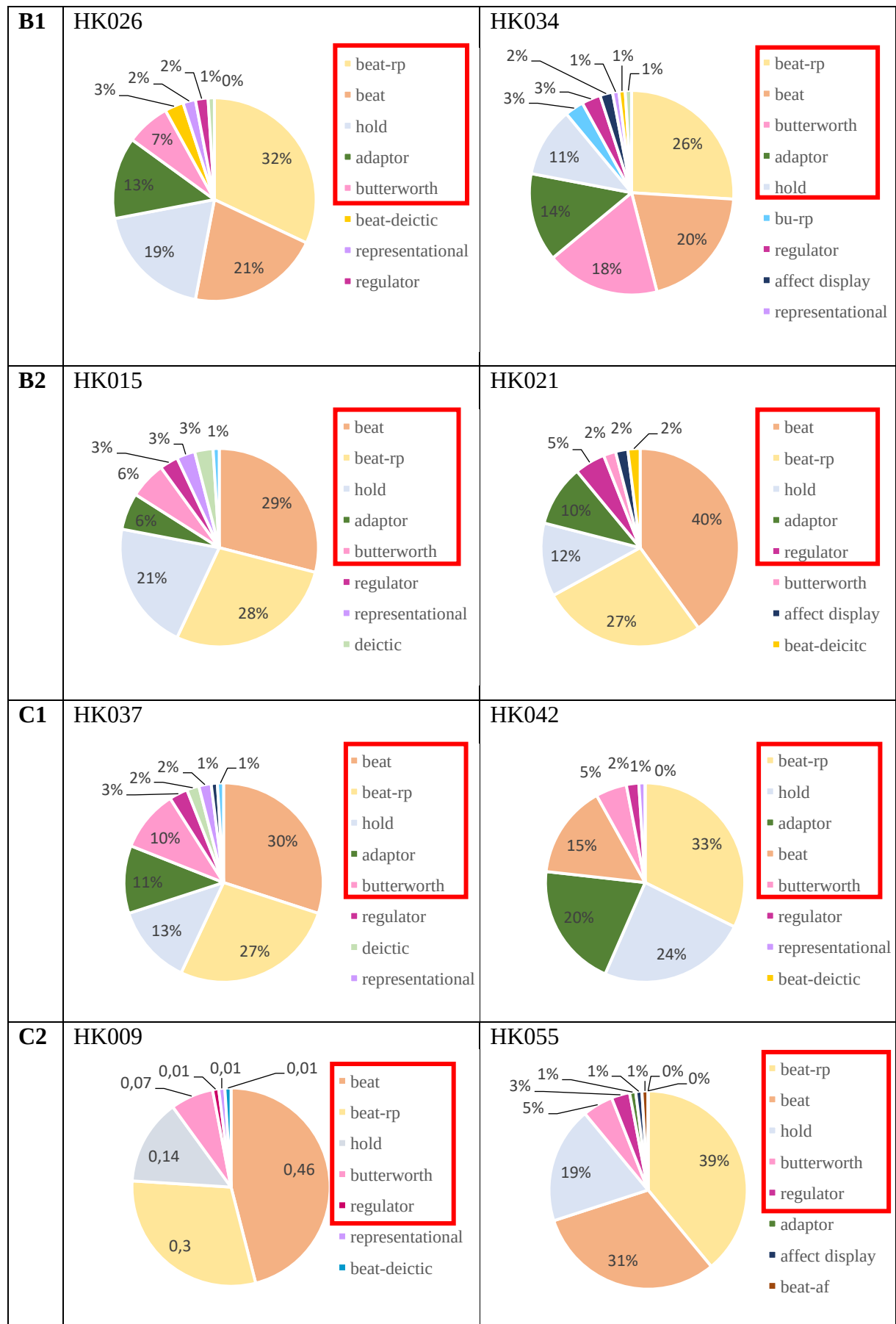


Figure 10: Proportion of gesture types per participant.

Basing one's observations on Figure 10, it can be said that the distribution of the types of gestures used by the participants varies but that some elements are stable. If it is true that all participants did not use the various gesture types in the same proportion, it is nevertheless noticeable that the gestures which they make the most are similar. The first five types used by each participant are indeed always five of the six following gestures: self-adaptors, beats, gestures combining the beat and representational types (beat-rp on Figure 10), butterworths, holds and regulators. The type which is the most used by all participants is always the beat type or the combination beat and representative. No participant used any gesture of the emblem type—neither alone nor in a combination with another type—in the interview excerpts which were analysed. It is interesting to notice that the self-adaptor type is one of the five most frequent types for all the participants except for both participants representing C2 level.

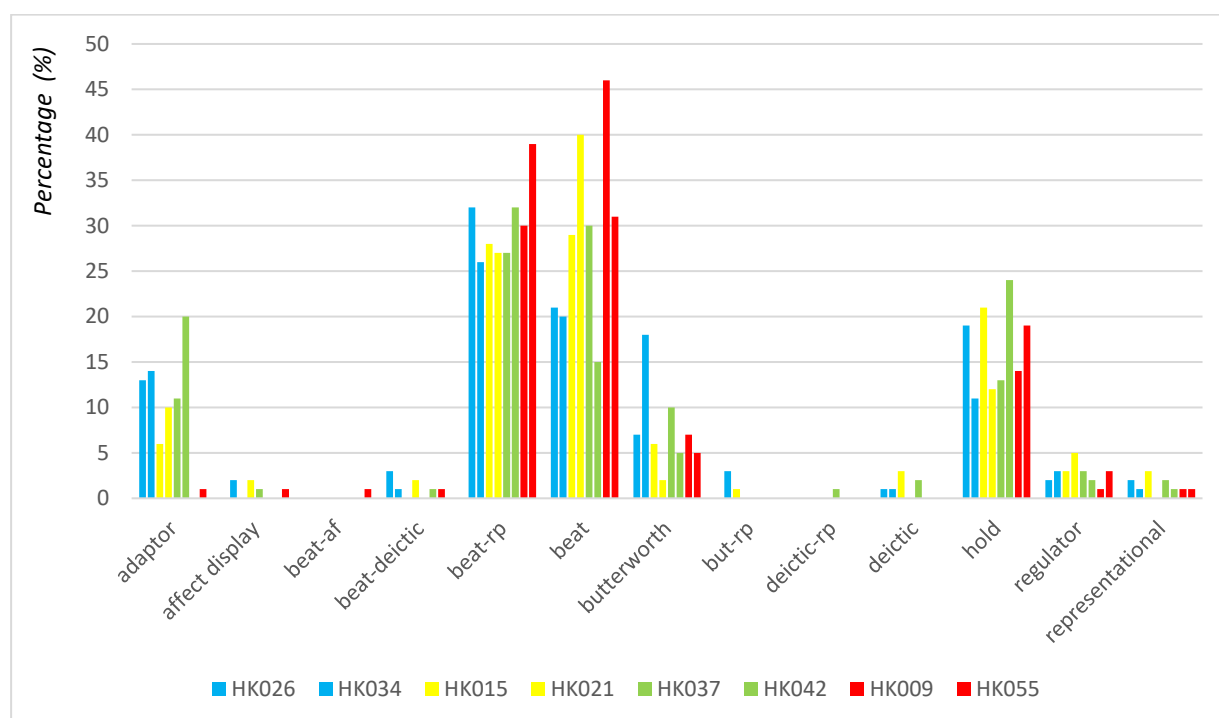


Figure 11: Percentage of gesture types per participant.

Figure 11 shows that the use of some gesture types is similar between the participants of a same level – this is the case for instance of the regulators and of the representational gestures – but that the two participants of each proficiency level can have contrasting results – this is the case for instance of the butterworths and of the holds. Drawing generalisation about the use of gesture types among the participants is thus difficult because there seems to be a lot of individual variations, depending on the gesture type which is observed. However, it can be said that the use of different gesture types is diversified as none of the types alone reaches the

50%, which means that all the participants tend to use various gesture types. Yet, some gesture types are indeed used in a greater proportion and this is the case for the beats and the beat-representational gestures, which are the most performed gestures for all the participants. It is interesting to notice that if gestures from the beat type were quite frequent, representational gestures were nearly never made alone. This be explained by the fact that representational gestures, which illustrate speech, are often made simultaneously with the performed speech, hence stressing the words which are used: they are thus frequently combined with beat gestures. On the contrary, beat gestures can be performed to punctuate the speaking or thinking process without per se illustrating the content and they can thus often be used alone.

Furthermore, the information which can be inferred from Figure 11 confirms the former observations based on Figure 10. Some types of gestures are indeed prominent for all the participants and they coincide with the six types which were formerly mentioned, namely self-adaptor, beat-representational, beat, butterworth, hold and also regulator even though this latter type is less present than the other types. On the contrary, some gesture types are used by very few participants and/or in a low percentage. It seems thus justified and relevant to focus on the use of some gesture types more specifically: comparing the production of gestures of the type beat-affect display (beat-af on Figure 11) would be irrelevant as only one gesture in the whole set of data had been placed in that category. This is why only the use of the gesture types which were performed by all the participants and/or which were used in a large proportion will be analysed in greater detail between the proficiency levels.

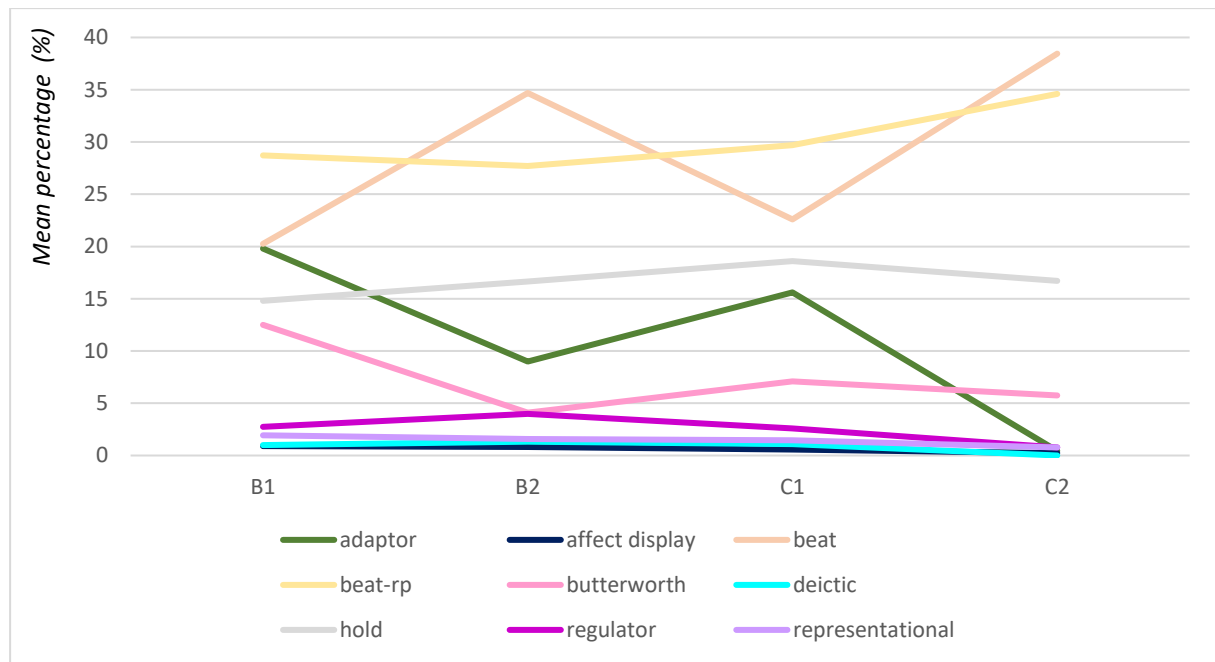


Figure 12: Mean percentage of frequent gesture types per proficiency level.

By observing the use of the various gesture types which are used by most of the participants, some tendencies can be displayed. From Figure 12 can be said that the use of representational gestures, of affect displays and of deictic gestures are relatively stable and low across the proficiency levels. The use of regulators is also stable in spite of a small increase from B1 to B2 and then a small decrease from B2 to C2. The use of butterworths and self-adaptors follows the same evolution between B1 and C1 since it first decreases from B1 to B2 and then increases from B2 to C1. The use of butterworths is relatively stable from C1 to C2 whereas the use of self-adaptors literally drops from C1 to C2 and becomes almost nil. Interestingly, the use of self-adaptors and of the use of beats seem inversely proportional: when the proportion of self-adaptors increases the beats decrease and when the self-adaptors decrease, the proportion of beats increases. Figure 12 thus shows that the mean proportion of self-adaptors decreases from B1 to B2, increases from B2 to C1 and then drastically decreases from C1 to C2. On the contrary, the mean proportion of beats increases from B1 to B2, decreases from B2 to C1 and then considerably increases from C1 to C2. Once again, it is possible that the high proportion of self-adaptors which has been used by the C1 participant HK042—and which is clearly observable on Figure 11—might have biased the mean proportion of the C1 level. This exception can once more be explained because the resting position of this participant had most of the time to be counted as self-adaptors because he tended to touch and move his fingers most of the time when he was talking or thinking of his speech, trying to find his words for instance.

Generally, it could thus be said that, in spite of a high proportion of self-adaptors in the gesturing of the participant HK042, the use of self-adaptors tends to decrease from B1 to C2, with few to no occurrence for the participants with the highest proficiency level: self-adaptors are indeed used more by the B1 participants than by all the other participants. Conversely, it can be said that the use of beats tends to increase from B1 to C2, despite some fluctuation between the levels. B1 participants are indeed the ones who perform beat gestures the less and C2 participants are the ones who perform them the most: C2 participants proportionally produce twice as many beats as B1 participants. The use of beat-representational gestures also rises with the proficiency level, though the difference between the proportion of beat-representational gestures made in the B1 level and in the C2 level is smaller than the difference which is marked for the beat gestures. Concerning the beat and beat-representational types, they are clearly the two most prominent types, regardless of the proficiency levels. Finally, the use of holds is relatively stable between the proficiency levels, although a slight increase can be observed from B1 to C1, followed by a slight decrease from C1 to C2.

5.3. Relation between gestures and speech

The third question which needs to be answered in this thesis is if the current analysis can bring to light any information about the relation between the gestures which are performed by the learners when it is their turn to speak and the actual speech they perform or try to perform. In order to study the potential relation between gestures and speech, two approaches were combined.

The first approach was to regroup all the gestures of the participants and compare the percentage of gestures which had been made without speech (with the code value 0), with speech and in direct relation to it (with the code value 1) or with speech but without direct or apparent connection to it (with the code value 2). The holds were analysed separately because they had been considered as gesture types but they were different in the way that they were precisely a non-moving gesture or in other words, a gesture interruption. Therefore, they had been coded with a binary code, 0 signifying that the speech had stopped with the gesture and 1 signifying that the gesture had stopped but the participant had kept on talking (usually with onomatopes like “erm” or with chunks like “how could I say that”).

The second approach was to compare the relation to speech for each gesture type among the participants. It was indeed possible that some gesture types would automatically or generally have a certain relation to speech. For instance, it was expected that beat gestures would per definition match speech if they were made while talking, because it would be difficult to punctuate speech without coinciding with it. Conversely, it seemed hard to think of any example of self-adaptor which would directly represent speech. Because it had already been shown that the use of the various gesture types was different from participant to participant, avoiding to draw generalisations about the relation between all gestures and the production of speech in general seemed relevant. It was thus established that verifying the relation to speech, and therefore their potential function, for all the gesture types separately could be valuable.

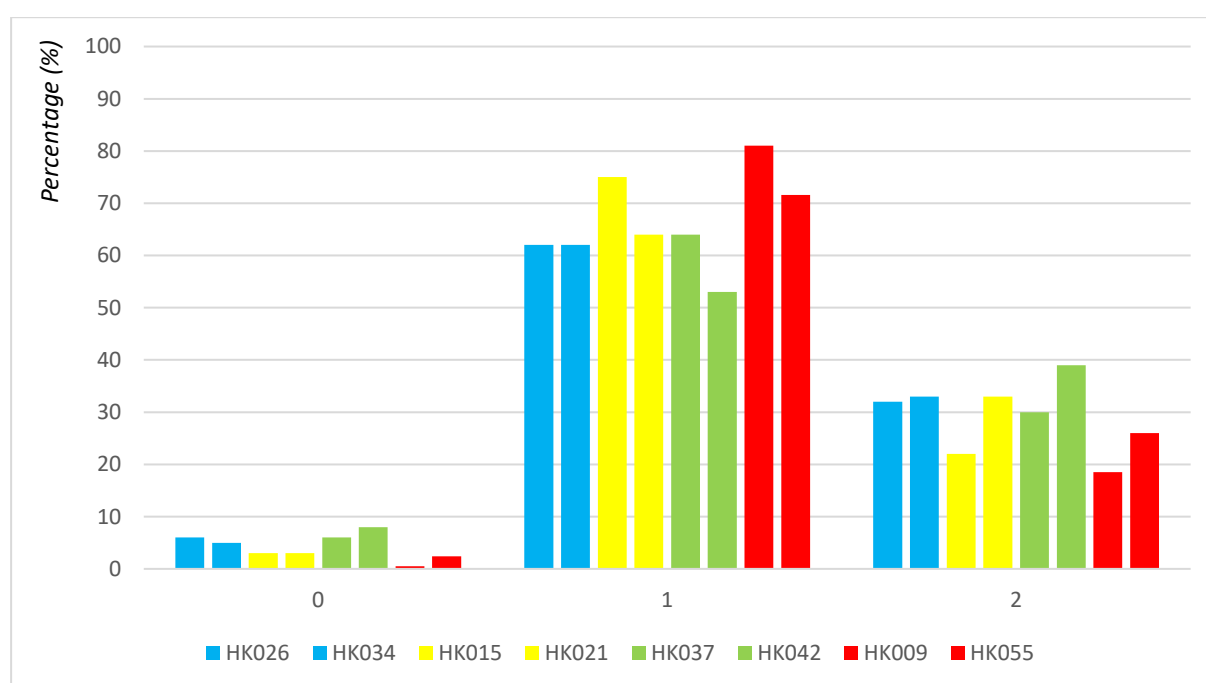


Figure 13a: Percentage of gestures-speech matching per participant.

Figure 13a shows a consistency between the participants of a same level as well as between the different proficiency levels. Indeed, it can be observed that for all the participants, more than half of their gestures matched what they were saying (code 1). Moreover, all participants used less gestures which did not match what they said (code 2) than gesture matching their speech (code 1) when they talked and gestured simultaneously, but they used even less gestures alone, without speech (code 0). Still, the participant HK042 was again an exception as the difference between the gesture matching and not matching speech (codes 1 and 2) is only of about 10% whereas this difference for all the other participants lies between nearly 30% and slightly more than 60%.

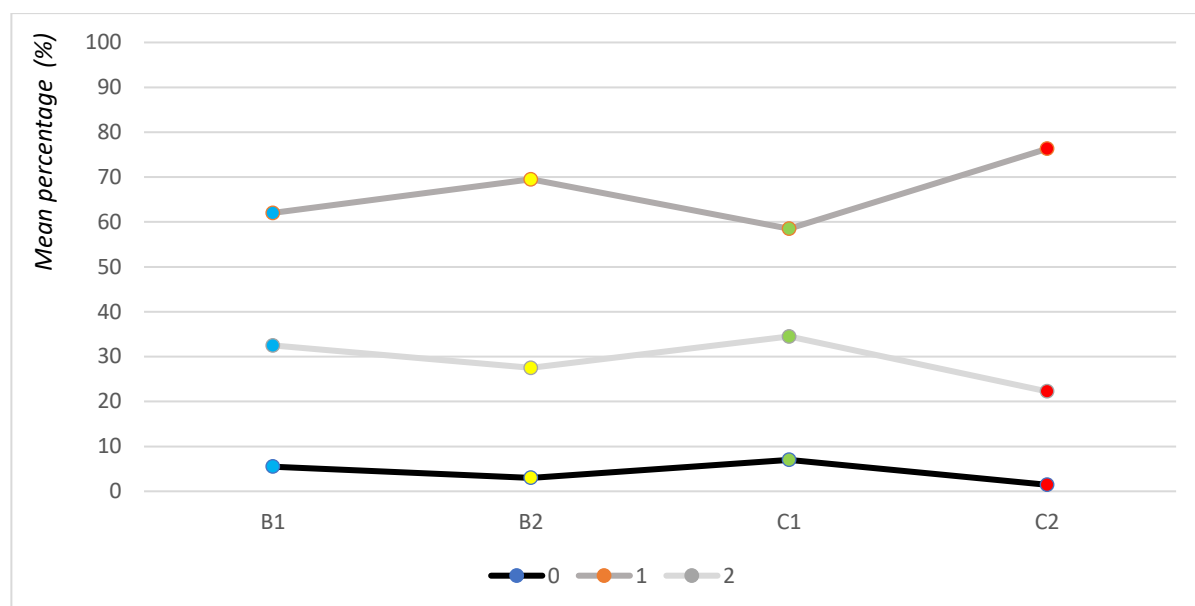


Figure 13b: Mean percentage of gestures-speech matching per proficiency level.

Figure 13b confirms that the gesture-speech matching is similar across the proficiency levels. It can be observed for all proficiency levels that the proportion of gestures made without speech (code 0) is lower than the proportion of gestures made with speech and that within the gestures performed with speech, the proportion of gestures matching speech (code 1) is always higher than the proportion of gestures which do not match speech. If the general tendencies are overall similar between the proficiency levels, Figure 13b shows nevertheless that the mean proportion of gestures matching speech, though being the highest in all the proficiency levels, is even higher for the learners of the C2 level. Indeed, they tend to proportionally make more than 70% of their gestures simultaneously with speech and matching what they say (gestures with the 1-value).

The variations in the use of matching speech gestures is inversely proportional to the variations in the use of non-matching speech gestures. The use of matching gestures increases from B1 to B2, decreases from B2 to C1 and increases again from C1 to C2, indicating a general increasing tendency: the proportion of matching gestures in C2 is around 20% higher than the proportion in B1, whereas on the contrary, the use of non-matching gestures decreases from B1 to B2, increases from B2 to C1 and decreases again from C1 to C2. The fact that the proportion of matching and non-matching speech varies so much between B2 and C1, either re-increasing or re-decreasing, can, again, possibly be related to the high proportion of self-adaptors which are made by the C1 participant HK042. Indeed, as will be further shown, the matching or non-matching character of some types of gestures can possibly be directly linked to their nature.

This is the case for self-adaptors which, per definition, do not match speech when they are used simultaneously with it.

Finally, Figure 13b also shows that the proportion of gestures performed without speech are rather stable between the proficiency levels and relatively low: it never exceeds 10% of the whole set of gestures. When comparing the proportion of gestures performed without speech to the proportion of gestures performed with speech (matching it or not), it can thus be established that most participants combined gestures and speech. It does not mean that the gestures performed without speech did not have meaning or were not semantically loaded, but it means that they were either performed instead of speech or that they were used while thinking about what could be said without producing any sound.

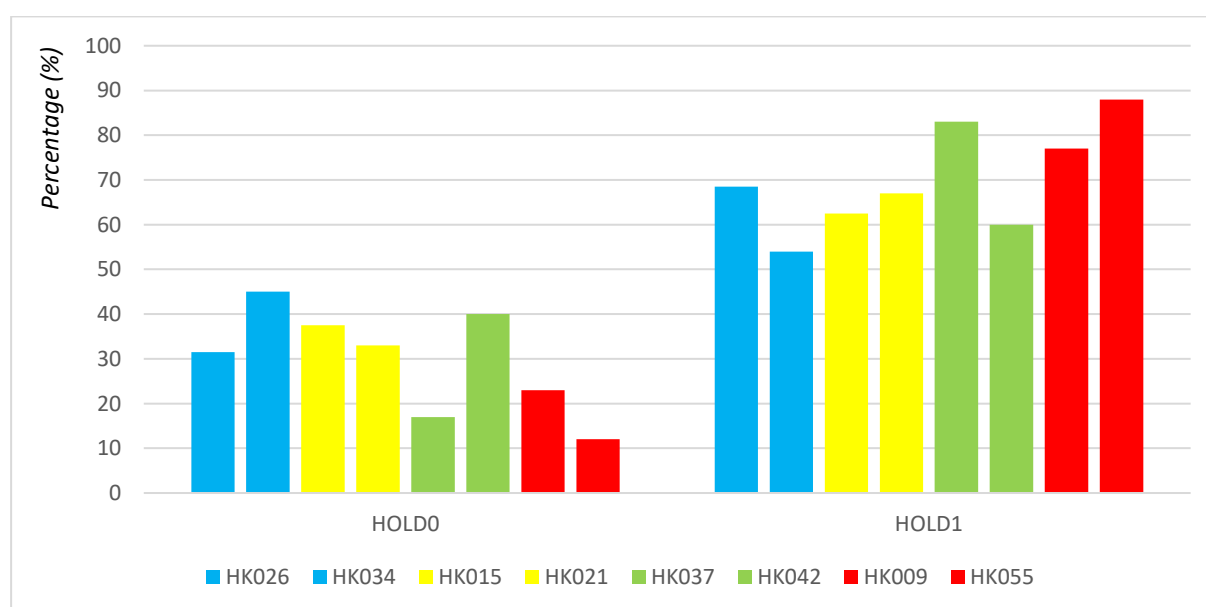


Figure 14a: Percentage of hold without or with speech per participant.

The observation that the participants usually gestured and spoke at the same time, matching what they said or not, is also to be found for the case of holds, or 'gesture interruptions', as illustrated on Figure 14a. The proportion of holds which have been made by the participants without interrupting their speech (code 1) is always superior to the proportion of holds they used when they also stopped talking (code 0). This is the case for all the participants.

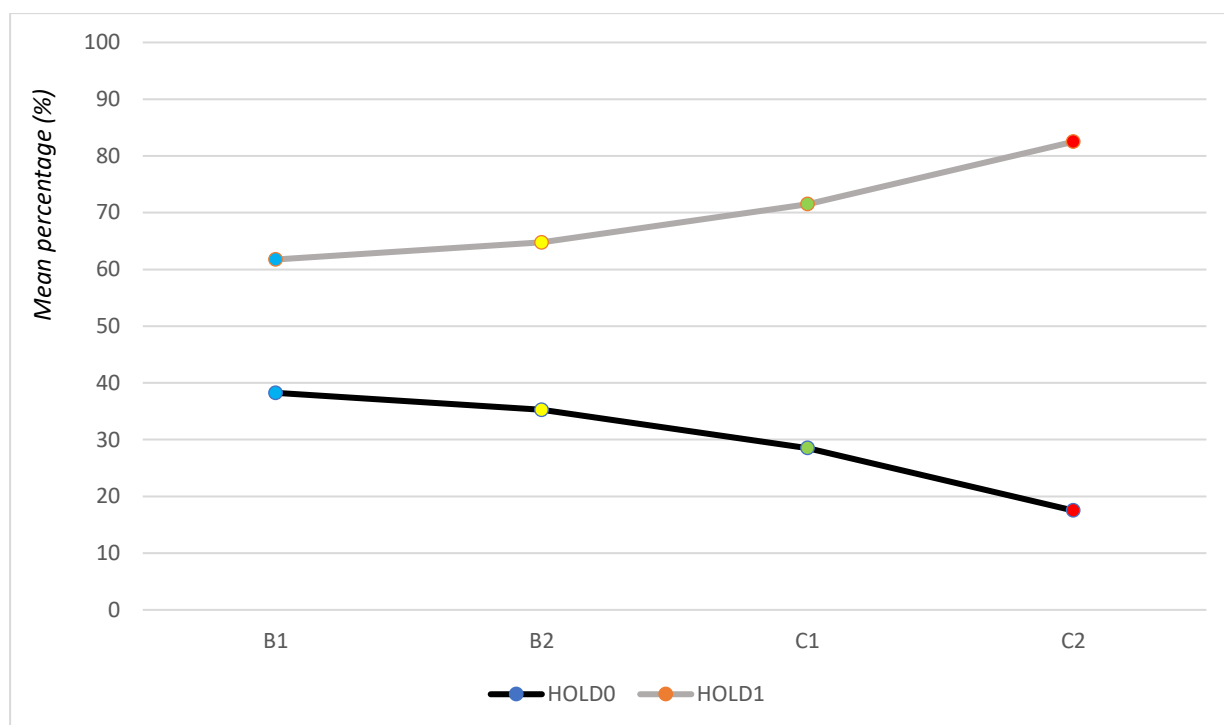


Figure 14b: Mean percentage of hold with or without speech per participant.

This is higher use of holds with speech than without speech is verified when the use of holds is compared between the proficiency levels, on Figure 14b. Yet, an increasement in the percentage of holds performed with speech is noticeable with the rise of proficiency. Indeed, a steady increase can be observed from B1 to C2. It could be explained by the fact that advanced learners tended to fill their thinking pauses with elements such as “I would say that” or “what is the word again”, whereas intermediate learners tended to focus efforts on their thoughts in order to find the words they lacked. This hypothesis could be further developed and investigated on with a larger panel of participants, including beginners from A1 and A2 levels.

Figure 15a confirms what has been observed on the previous graphs, namely that the participants tend to proportionally use more gestures when they talk than when they are silent. An exception can be found with the category of regulators in which the participants appear to gesture in a similar way, with or without speech. Figure 15a also confirms what has previously been induced regarding the gesture types as the same categories stand out as being the most used, i.e. beats, beat-representational gestures, butterworths, holds, regulators and self-adaptors. These categories have been regrouped in Figure 15b to give a better overview and ease their comparison.

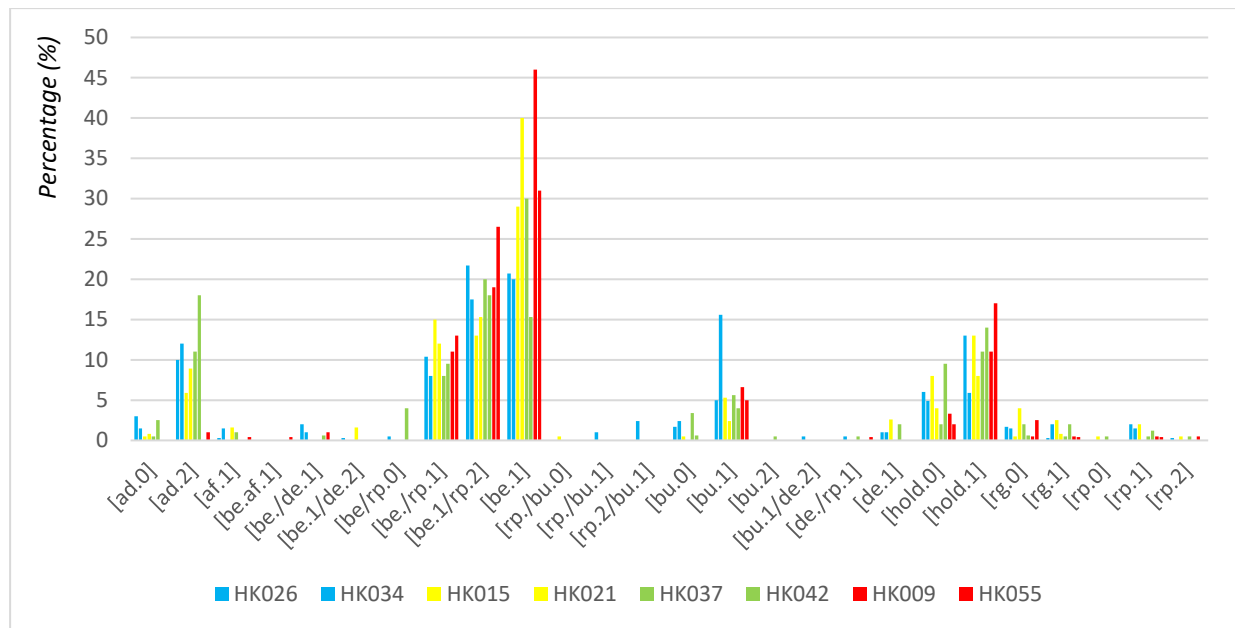


Figure 15a: Percentage of gesture types in relation to speech per participant.

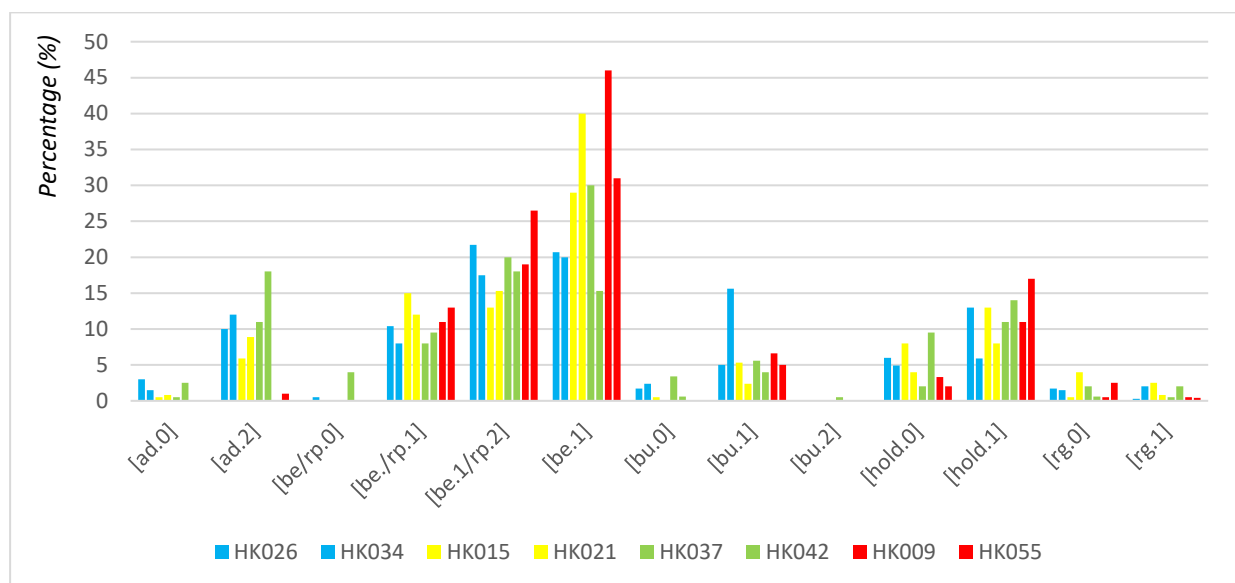


Figure 15b: Focus on the most used gesture types in relation to speech.

The added value of this graph is that it establishes the intrinsic relation that some gesture types seem to have with speech. Indeed, most gesture types can easily be performed with or without speech, even though it has already been mentioned that gestures mostly occurred with speech. This relation to speech, matching it or not, seems however to vary from type to type, as expected. In that way, it can be observed that no example of self-adaptor was found directly matching speech ([ad.1]). On the contrary, only beat gestures which were performed simultaneously with speech and matching it ([be.1]) were made by the participants. This type of gesture (beats) was often combined with the other types and especially with representational

gestures. It is interesting to notice that in this combination, the representational part could either be considered as matching the speech or not, whereas the beat part was always considered as matching the speech. The butterworths used by the participants in this study were most of the time performed simultaneously to speech and matching it (with utterances like “erm” or “I wonder if”, which express the same information as the gesture type, namely the thinking process), to the point that no butterworth gesture was performed without speech by the two participants who represent the C2 level. There is no occurrence of hold with the code value 2 because, as already explained, this “type” was coded separately, with a binary code expressing the interruption or non-interruption of speech instead of a matching or non-matching characteristic.

It should be mentioned that the matching or non-matching character with speech was directly linked to the context in which the gesture occurred and the type of gesture which was used. Indeed, the various types of gestures do not give the same kinds of information. Therefore, the fact that no self-adaptor was found with a 1-value (matching speech) is not surprising. Indeed, self-adaptors tend to be excluded from most studies precisely because they are believed to have no communicative meaning (see Section 1.2). In the study, this type of gesture was actually mostly used by learners when they had to produce speech. When they occurred simultaneously with speech, it was usually either because the learners were filling their pauses but without actually expressing any new idea nor content in their speech, or because they were talking about something while gesturing to help them constructing their next ideas. Therefore, it would hardly be possible to associate self-adaptors directly with speech. Another type of gestures which has been attributed a similar function is the butterworth gesture. Still, this gesture can explicitly be interpreted by anybody as signifying “I am thinking” or “I am searching for a word”. This is why when this type of gesture was used by learners when they said similar words, verbally directly referring to their word-searching, these gestures were attributed a 1-value: they were considered to directly match speech. With or without the words, the interpretation would have been similar, which is not the case with adaptors. The fact that no beat was found with a 2-value (non-matching speech) is not surprising either as their function is precisely to punctuate the speech (and/or perhaps the thoughts when no speech is produced). Therefore, it would be hard to imagine a gesture ‘mispunctuating’ speech. So, per definition, when a beat gesture was performed with speech, it automatically matched what was being said. Through this study, the function of the different gesture types could be confirmed and/or refined.

5.4. Functions of gestures and frequent patterns

Annotating the data which were observed made possible to extract quantitative information and draw statistically relevant conclusions regarding the frequency of the gestures which were performed, the types of gestures which were made and their relation to speech, this within and between the proficiency levels. Next to these quantitative results, some qualitative tendencies could also be observed regarding the use of gestures and more precisely regarding the function they might have for the participants and the context in which they tended to be used. The following examples aim at showing that gestures tended to be used in two major contexts: sometimes they seemed to be used in order to communicate and transmit information to the listener, in addition to speech or instead of it, and sometimes they seemed to be used in situations in which speech was more difficult to reach, depending on the proficiency level and consequently on the mastering of the L2, possibly to help the learner produce speech. Of course, a lot of gestures which were made could be placed in both functions. It is indeed difficult to assert whether the performed gestures were intended to have a certain impact or not, and whether the performed gestures indeed had an impact or not, on the interviewee, the interviewer or both, especially without directly asking these people. Still, based on the observations, some patterns stood out and it was possible to interpret them. Some patterns indeed appeared to be frequent in the speech production of the participants, especially regarding the gestures which were made when the learners did not know what to say and/or when they corrected themselves. Further study is needed to support the following elements on a quantitative basis too. Still, the qualitative observations which will be presented here, based on specific examples, are relevant to point out and appear to coincide with the results which were previously obtained.

A first observation was made regarding the use of gestures when participants corrected a mistake they had just made or decided to rephrase something they had just said. In these situations, participants tended to insist on the second option they proposed, sometimes via speech, by modifying their intonation to stress the word they corrected, and almost always via gestures, by using a beat to stress their correction. Sometimes this beat was preceded by a short hold, as if the participants realised the mistake and froze before starting their gesture again. Sometimes the hold was absent and the beat was only stressed in the continuity of the conversation. Examples of this pattern were found in all the analysed interview excerpts. Figure 16 illustrates this pattern.

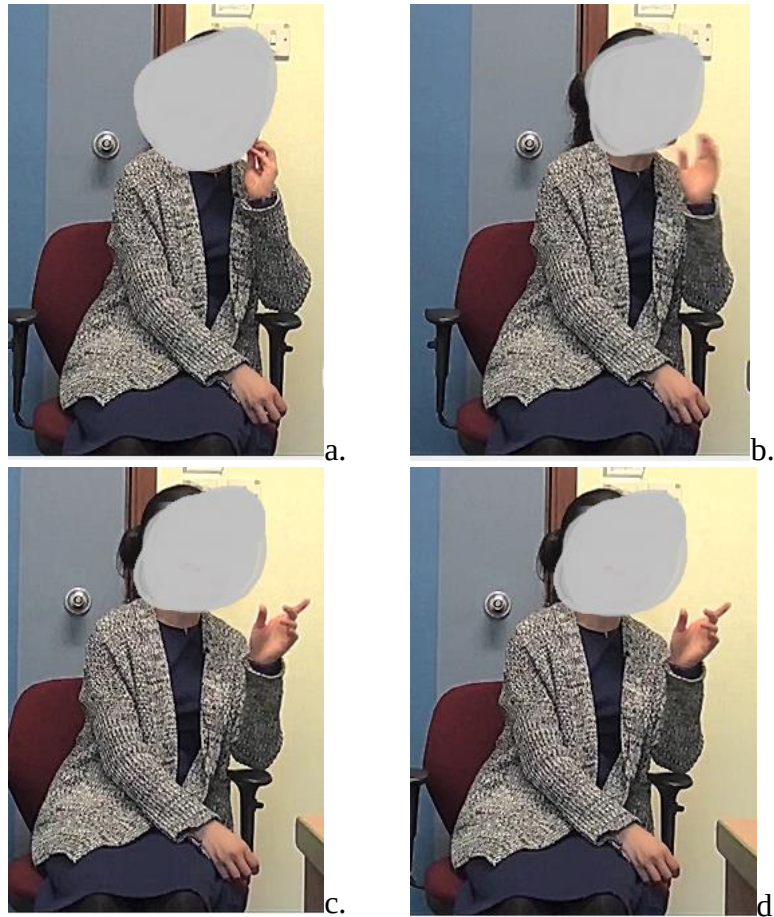


Figure 16: Example of beat to punctuate a self-correction (from participant HK037).

Figure 16 represents the gestures which were made by the C1 participant HK037 from 0.52 to 0.54 minutes of her interview. During this period of time she said “I usually spends* erm I usually **spend**”. She realised her mistake (I spends*) and corrected herself (I **spend**), stressing her correction with her intonation and with her hands, as can be seen on Figure 16 and especially on picture c. and picture d.

A similar observation was made when the participants repeated something. Indeed, most of the time, when they repeated words they had just said, they also repeated the gestures they had just performed. This observation corresponds to what has previously been said in the literature (see Section 1.5), namely that gestures and speech should be considered as two related channels which tend to be both repeated when an idea is repeated. It was also the case when some words were repeated to introduce a correction. In this case, the repeated gesture was usually combined with the above-mentioned use of beats to stress the correction: the initial movement was repeated and ended with a beat which was accentuated. An illustration of this pattern is found on Figure 17.



Figure 17: Example of gestures and speech repetition ended with a beat (from participant HK034).

Figure 17 represents the succession of gestures which were made by the B1 learner HK034 from 3.36 to 3.42 minutes of her interview excerpt while saying “a lots of*, of, a lot of, erm shop erm, lots of **supermarkets**”. Figure 17 shows that the succession of pictures a., b., c. is very similar to the succession of pictures d., e., f., which reflects the repetition in the speech of “a lot of”. The word supermarket ends this repetition and was stressed by the intonation as well as by the gestures, as can be seen from pictures h. and i.

Another observation was made regarding the use of gestures of the participants when they were thinking of the speech they had to produce, either because they were unsure of what to say or because they did not know how to say it in their L2. Indeed, three gesture types were observed to be mostly used by the learners in the speech and thoughts formation more than in the speech transmission. They were thus interpreted as helping the learners think of their speech production more than helping them communicate per se: self-adaptors, butterworths and holds. It was thus observed that self-adaptors and butterworths were frequently used together with holds and that, once more, these gestures were often followed by a beat gesture when the correct speech was finally produced. More precisely, it could be observed that the adaptors were most of the time either directly followed by a beat gesture or followed by a hold and then a beat. The same observation was made for the butterworths. Also, self-adaptors and butterworths often followed each other. A frequent succession was a self-adaptor followed by a butterworth and then either another adaptor or a hold, finally followed by a beat. More than once, the three gesture types—butterworths, self-adaptors and holds—were successively made without specific order, performing one and then another until finally the correct word was found and the gesture succession was ended with a beat gesture. The holds tended to be used together with self-adaptors and/or butterworths, but not only. They were also used alone, especially in the already mentioned context of self-correction: the learners interrupted their gestures when they realised their mistake and started it all over again. This is why holds were not only—or not immediately—followed by beats, but they were also often followed by representational gestures, beat-representational gestures and sometimes also deictic gestures, for instance. Still, the trio self-adaptors, beats and holds was generally observed in situations where the learners had to think of their speech production.

From the three previous observations, related to the self-correction, repetition and thinking process (especially word-retrieval), it can be said that beat gestures seem to punctuate speech by stressing important words. They tend to be used when the learners correct themselves or find the word they were looking for, but also simply throughout the conversation to punctuate and give rhythm to their speech, like they can do (and sometimes simultaneously do) with their intonation. This gesture type has been shown to be the most used (alongside beat-representational gestures) by the participants in all the proficiency levels (see Section 5.2), but regarding the contexts in which these gestures were made by the participants, it appears that they were not used for the same reasons. In the case of the intermediate learners, beats tended to be performed in the second situation, i.e. after having been searching for their words via

adaptors, *butterworths* and/or *holds*. The advanced learners, and especially in the C2 level, tended to use a lot of beats to punctuate their speech as if to give life to their conversation, so that the communication could be optimised. The beats which were used when the learners corrected or rephrased some words were found in all the levels when this situation occurred. Interestingly, the intensity of the gestures from the beat type was not always the same. Usually, when the learners were in their resting position (often with their hands joined on their knee), they punctuated a word or a syllable only by slightly moving their fingers, still staying close to their resting position. Sometimes, usually when they had already left their resting position, to explain something for instance, they made larger beat gestures, moving their hand and their arm to stress the words or word syllables.

Many gestures were considered as a combination of beat gesture plus another gesture type, often representational. This seems logical because, generally, when a word was illustrated by a representational gesture, the gesture simultaneously punctuated the word, hence belonging to the beat category. The beat part of the gesture was per definition always directly related to speech, but the other aspect of the gesture, namely the representational part, could either be directly related to what was being said or not. Many gesture occurrences were indeed annotated with the gesture type [be.1/rp.2]. These gestures were often made by the learners in situations in which they seemed to ‘materialise’ what they were talking about. They indeed placed their hands as if they were trying to physically place the words which they were saying in front of their interlocutor. The gesture did not directly match speech as it would have been impossible to interpret the gestures as more than ‘something’, unlike representational gestures which directly illustrated the words or concepts which were said. Still, they tended to be used a lot by the learners, especially when they were explaining something. This could be related to the pedagogical function which has been attributed to gestures in the literature (see Section 1.4), suggesting that some gestures tend to be made by teachers to enhance their pupils’ understanding. This could also be related to the fact, again largely agreed on in the literature (see Section 1.4) that gestures seem to be influent both for the speaker and the listener in order to optimise the communication. Representational gestures (iconic and/or metaphorical) tend indeed to be a type which is often referred to in the literature when talking about the communicative impact of gestures.

A gesture type which is also often related to communication is the deictic type. This type of pointing-gesture directly depends on the environment in which the conversation is held (see Section 1.2). In the observed data, the context limited the possibility of objects to point at, as expected. This is why there were relatively few occurrences of deictic gestures whereas this type of gesture has been suggested to be largely used by learners (see Section 2.3.2). The fact that the use of this gesture type was limited by the direct environment in which the interviews took place could be pointed out in the interview of the participant HK034, from B1 level, as illustrated on Figure 18.



Figure 18: Example of the limitation for deictic gestures (from participant HK034).

Figure 18 represents the movements made by the B1 participant HK034 from 1.37 to 1.39 minutes of her interview excerpt. She performed these gestures while saying “the, like, some, you know, the... the sign”. It can be seen that the learner tried to find a “sign” to point to (picture b.) but could not find any (picture c.) and tried to find another way of showing what she wanted to say by finally ‘pointing’ to her own hand as if her hand represented the sign (picture f.).

Gestures of the regulator type were used by the learners in this study, as expected, but not in a large proportion. Indeed, because the observed data were interviews excerpts, there were some interactions between the interviewees and the interviewer. The regulators were used to regulate these interactions, and especially the turn-takings, as suggested in the literature (see Section 1.2). However, the number of turn-takings was quite limited as the aim of the interviews were to make the interviewees talk to allow for an analysis of their speech production. This is why regulators were used by all the participants in all the proficiency levels. Usually, the regulators were performed in the same way by all the participants. Learners were performing other types of gestures and they put their hands down on their knee but facing the interviewer, as to indicate that they invited the interviewer to speak. There were two situations. Most of the time, the learners were explaining or telling something and they arrived at the end of their story, hence indicating via the regulators that their story was finished and they had nothing to add. The gestures which preceded the regulators in this situation were generally beats or (beat-)representational gestures. Sometimes, the learners were struggling with their words and/or ideas and did not know how to pursue, hence indicating via the regulators that the interviewer could help them by interrupting their story. This second situation was less frequent but occurred more than once and in various interviews. The gestures which were made before the regulators in this situation were adaptors. It is also important to stress that the movements considered as gestures in the study were only hand and arm movements, from the fingers to the shoulders, but that other elements such as head moves or mimics also influenced the communication and were sometimes useful to interpret the gestures. This was often the case with regulators. Indeed, when the learners performed a gesture which was considered as a regulator, they also moved their head, as if they were nodding, and they seemed to open the position of their whole body towards the interviewer, not only their hands.

Other parts of the body were also essential in the interpretation of gestures of the type affect display. As expected, this type of gesture was not frequently used by the participants because affect displays are often used in the literature to refer to mimics and less to hand movements (see Section 1.2). Still, some gestures were considered as affect displays. Learners generally performed affect displays by bringing one hand or both hands towards their heart. Here again, the context in which the gesture was performed was important to determine the type of gesture. Indeed, this type of movement could sometimes be considered as deictic, when the learners made them while talking about themselves: they said “I” or “me” while pointing at themselves, which was considered as an example of deictic gesture. On the other hand, if this

type of movement was made when talking about something emotional and/or personal, the fact that the learners brought their hands towards their heart was interpreted as an expression of their emotion, and thus an example of affect display. A good example of affect display, illustrated on Figure 19, was found at 4.57 minutes of participant HK021's interview excerpt. She made this gesture while saying "I may feel sad to [...]", hence directly referring to an expression. This is why the fact that she raised her hand towards her heart while talking about sadness was considered as an example of affect display.

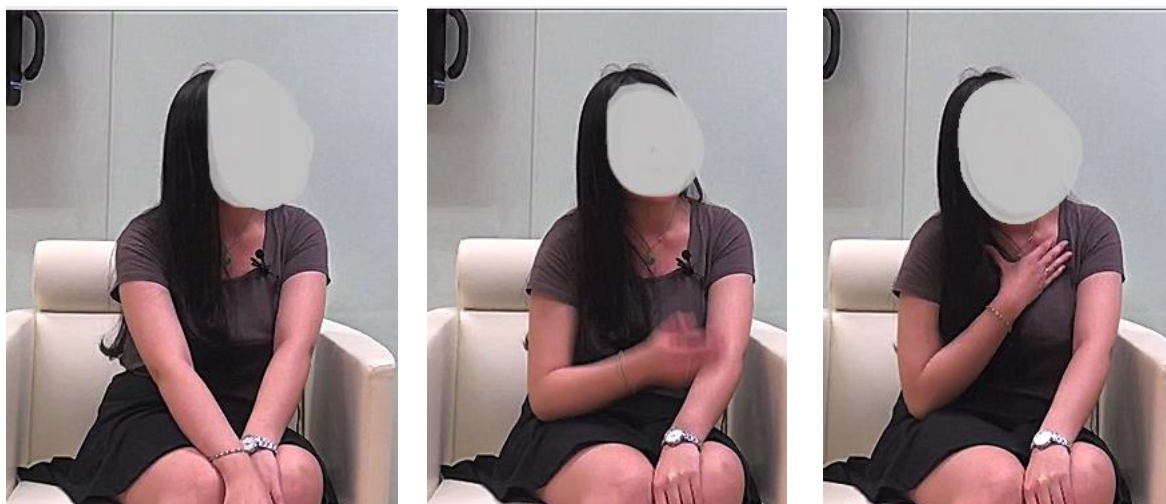


Figure 19: Example of affect display (from participant HK021).

The qualitative observations of the interviews highlighted two main functions which could be attributed to gestures. On the one hand, gestures seem to play a role in the 'external' communication, which means they are used similarly to speech, to express the same kind of ideas, either simultaneously with speech, in addition to it or instead of it. This function also seems to modulate the speech transmission, potentially like the intonation could do, for instance, to optimise the transmission and the communication. On the other hand, gestures seem to play a role in the 'internal' thinking process, which means they can be used by the learners to gain access either to ideas in general—when they search what they could say—or to linguistic knowledge—when they search how to say what they want in the target language. These two functions are not in a binary system which would imply that if a gesture has one function, it automatically excludes the other function. On the contrary, it is possible for one gesture to be attributed both functions. Moreover, as already said in the literature (see Section 1.2) and in the examples above, each type of gesture seems to have a specific function which can specify the two major functions previously mentioned: affect display tend to express emotions; beats give rhythm; butterworths seem to express the thinking process and word-retrieval; deictic gestures usually point to physical entities in the real and immediate environment; holds possibly help

the learners focus on their thoughts and often precede stressed elements which are punctuated by beats; regulators control the interactions and turn-takings in the conversations; representational gestures directly illustrate what is being said or seem to be used to ‘materialise’ words and ideas in 3D; self-adaptors seem to help the learners focus on their thoughts and express ideas correctly in words. Table 6 summarises which of the two major functions was fulfilled by the gesture types which were observed in this study. When the function had been observed in the analysis, a v was written in the corresponding case. When the function was not observed in the data, a x was written. If the function seemed to be fulfilled but was not the function which was the most observed in the study, a v was written in the corresponding case, but the case was greyed out.

Gesture types	Function related to the ‘external’ communication and speech transmission	Function related to the ‘internal’ thinking process and speech production
Affect displays	v	x
Beats	v	v
(Beat-)representational gestures	v	v
Butterworths	v	v
Deictic gestures	v	v
Holds	x	v
Regulators	v	x
Self-adaptors	x	v

Table 6: Functions of the gesture types observed in this study.

Some gesture types seem to play a role in the ‘external’ communication only (be it for the speakers, the listeners or both), like affect displays and regulators. Some others tend to be used for the ‘internal’ speech production and the thinking process only, like self-adaptors and holds. The beat type is the only gesture type which clearly plays a role in both situations: this confirms the previous observations and this justifies why it was used a lot by all participants. Gestures which have been attributed the dual function beat and representational as well as deictic gestures influence the speech transmission, by supporting, adding or replacing information which either is already conveyed or which could also be conveyed by the speech. These types might possibly play a role in the speech production too, because they might enable the learners to ‘visualise’ the words they need to find. Finally, butterworths clearly play a role in the thinking process, mostly for word-retrieval. They might also play a role in the external communication as they could sometimes explicitly call for the listeners’ help. Further study is needed to develop the role of gestures, but the present observations appear to justify this dual function of L2 learners’ gestures.

Conclusion

Based on the pseudo-longitudinal corpus-based analysis which has been carried out in this thesis, it can be concluded that gestures play a role in the second language acquisition process. The results obtained in this study indeed lead to the conclusion that the participants used gestures differently depending on their proficiency level: variations appeared between the proficiency levels regarding the frequency and the type of the gestures which tended to be used and also the context and the function which were attributed to them. Though the proficiency level does not seem to be the only factor which has an influence on the use of gestures—individual variations within the proficiency levels testify it – the results clearly showed that learners did not use gestures similarly depending on their L2 proficiency.

It was thus observed that the number of gesture occurrences tends to decrease from B1 to B2 and then to increase from B1 to C2. An explanation could be that the function of the gestures they use is different. This explanation is supported by the other results. A first validation comes from the proportion of gesturing time of the participants, which tends to decrease from B1 to C2. There is a slight increase from B2 to C1 which is explainable by the individual variations and the gesturing tendencies of the participant HK042 which often differs from the global tendencies. By combining these two results (the number of gestures performed and their duration), it was possible to calculate a ratio for each participant, indicating the mean duration of the gestures used in each level. This shows that the duration of the gestures performed by the C2 level participants is proportionally 0.5 seconds shorter than the duration of the gestures used by the B1 level participants. Moreover, it was shown that the mean duration of each gesture type was different, with self-adaptors lasting proportionally more than twice as long as beats, for instance. This confirms that the frequency of gestures used by the learners during their speaking turn varies depending on their proficiency level. Intermediates seem to constantly gestures whereas advanced learners tend to use many gestures, but much shorter gestures, which leads them to spend more time in a resting (unmoving) position when it is their turn to speak than intermediates. Comparing these results with data from beginners would be relevant to sketch the complete picture between all the proficiency levels and to see whether the same tendencies are observed or not. It could not be concluded from the results that advanced learners use less gestures than intermediates, but well that they gesture for less time.

Another way of confirming the hypothesis that the gestures used by the learners between the levels tend to serve different functions is found when analysing the gesture types which are used. Indeed, while the same six gesture types – beats, beat-representational gestures, butterworths, holds, regulators and self-adaptors – are always found in the ‘top-5’ of all the participants, one can notice that the use of self-adaptors tends to decline from B1 to C2 whereas the use of beats and beat-representational gestures tends to grow. Again, the rising from B2 to C1 in the use of self-adaptors (mirrored by a symmetrical decrease in the use of beats) could be attributed to the abnormally high use of self-adaptors of the participants HK042 from the C1 level, in comparison with the use of the other participants. These results coincide with the decreasing proportion of gesturing time previously calculated because beats are usually much shorter than self-adaptors in terms of duration. It could be related to the function of these gestures: beat gestures stress one word or word syllable whereas self-adaptors are most of the time used instead of the resting position to help the speakers access their thinking process, which justifies the short duration of beats and the longer duration of self-adaptors. The learners’ proficiency level thus seems to influence their use of the various gesture types too.

From this study can also be concluded that gestures are related to speech production. It has indeed been shown that all participants proportionally performed more gestures with speech than without speech. Furthermore, the proportion of gestures matching speech was always higher than the proportion of gestures which did not match speech, with an increase in the proportion of matching speech gestures across the different levels. Again, there was an exception which can be explained by the participant HK042’s high use of self-adaptors (non-matching speech): this could justify the decrease in the proportion of matching speech gestures from B2 to C1. This being said, the general rise in the proportion of matching speech gestures points the tendency among the C2 participants to use gestures and speech conjointly, as suggested in the literature. This also confirms that learners do not use gestures in the same way, depending on their proficiency level. Moreover, it has been established that the relation to speech varies depending on the type of gesture which is considered. Some gesture types can be placed in the three categories, namely being performed without speech, matching speech or with speech but without (directly) matching what is said. On the contrary, some gesture types, when performed with speech, will per definition always either match or not what is being said. This is for instance the case of self-adaptors, which are always performed either without speech or without matching it, and of beats, which generally match speech. The relation gestures have with speech seems thus to vary depending on the gesture type and on its function.

By observing the situations in which gestures tended to be made, two main functions could be suggested. Some gestures seem to be more oriented towards the ‘external’ communication, supporting, adding or replacing information which could be or is conveyed by the speech, and others are more oriented towards the ‘internal’ speech production and the thinking process. These two functions are often intertwined, and it would be hard to establish whether the use of gestures impacts the listener, the speaker or both. Still, on the basis of the observations made during the analysis, some patterns and frequent situations in which specific gesture types were used could be pointed out. Self-adaptors, butterworths and holds seem to be mostly performed by the learners when they think of the speech they have to produce, either because they do not know what to say or because they do not know how to say it. It is interesting to notice that these three gesture types often occur together and that their use varies across the levels: for instance, learners in the C2 level tend to use holds more frequently with speech than the other participants and they tend to use few to no self-adaptors. The beats seem to punctuate speech and stress important words. Beats tend to be used in two situations: they are used by learners when they correct themselves and/or find a word they were looking for to stress the correct word, and they are used by learners to give rhythm to the conversation, like intonation can do. The first situation is more often found in the data of the intermediate learners, whereas most of the beats used by the advanced learners are made in the second situation. This dual function could explain why this gesture type is the most used by all the participants and it seems to confirm that gestures do not fulfil the same function depending on the proficiency level and/or the context in which they are made. The other gesture types, namely the affect displays, the deictic and representational gestures and the regulators appear to be mostly linked to the speech transmission and communication in general because they have a function which is similar to speech. These functions coincide with what has already been said in the literature, namely that gestures and speech seem to be connected and that the specific role that gestures play surely depends on their type. The gestures which were observed in this study make possible to suggest two major functions related to speech production (and second language acquisition): learners probably gesture on the one hand to optimise their communication with their interlocutor(s) and on the other hand to mentally structure and support their linguistic abilities.

General tendencies could thus be highlighted regarding the use of gestures in terms of frequency, types and relation to the produced speech and speech production. It should however be borne in mind that the study relied on a limited amount of data. Due to time and practical limitations, only two participants were observed per proficiency level and no access to beginners (A1 and A2 levels) was gained. However, the tendencies which are revealed in this thesis are encouraging. Therefore, extending this analysis to include beginners and observe more participants for each proficiency level could be valuable in order to confirm and develop them. Another point, related to the previous one, is that some other aspects than the proficiency level can influence the use of gestures. It has indeed been established in the literature that the culture as well as individual characteristics, for instance, could impact the use of gestures. The cultural variable was excluded from this study as all participants shared a similar cultural background: they were all university students from Hong Kong. Further studies could nevertheless include participants from other cultural backgrounds in order to check if the variation in their gesturing follows the same patterns as suggested here. The personal variable, on the other hand, was inevitably taken into account in this study. Hence, the results obtained for participants HK042, the second participant representing the C1 level, often diverged from the general tendencies observed among the other participants. It is worth noticing that this participant was the only male interviewee among the other female participants: this could potentially affect the way of gesturing, although there are not enough data here to draw any conclusions based on the participants' gender. Also, this participant is the one for which the proficiency level of the three raters was the most contrasted (see Table 2): his mean CEFR score was C1, but he had been attributed proficiency levels from B1 to C2, depending on the rater. This could also explain why his gesturing appears different from the general tendencies. The results of this study brought some interesting observations to light, but these are tendencies and not general truths. Yet, these tendencies are encouraging and they should be worth verifying in further studies as the proficiency level indeed seems to influence the way learners use gestures when they have to produce speech in a L2.

Bibliography

- Abuczki, A., & Ghazaleh, E. (2013) An overview of multimodal corpora, annotation tools and schemes. *Argumentum* 9: 86-98.
- Allwood, J. (2014) Multimodal Corpora. In Lüdeling, A., Merja, K. (eds), *Corpus Linguistics. An international Handbook*. Berlin: Mouton de Gruyter, 207: 225.
- Andor, J. (2004) The master and his performance: An interview with Noam Chomsky. In *Intercultural Pragmatics* 1(1): 93–111.
- Beattie, G. (2003) *Gesture Types*. Retrieved April 19, 2019, from http://changingminds.org/explanations/behaviors/body_language/gesture_type.htm.
- Bolton, K. (2002) The sociolinguistics of Hong Kong and the space for Hong Kong English. In Bolton, K. (ed.), *Hong Kong English: autonomy and creativity*. Hong Kong: Hong Kong University Press, 29-55.
- Brunner, M. L., Diemer, S., & Schmidt, S. (2017) "... okay so good luck with that ((laughing))?" - Managing rich data in a corpus of Skype conversations. *Studies in Variation, Contacts and Change in English* 19 [Big and Rich Data in English Corpus Linguistics: Methods and explorations, ed. by Turo Hiltunen; Joe McVeigh; and Tanja Säily]. Helsinki: Varieng.
- Butterworth, B., & Beattie, G. (1978) Gesture and silence as indicators of planning in speech. In Campbell, R.N., & Smith P.T. (eds) *Recent Advances in the Psychology of Language Formal and Experimental Approaches*. Plenum: New York, 347-360.
- Capone, N., & McGregor, K. (2005) The Effect of Semantic Representation on Toddlers' Word Retrieval. *Journal of Speech, Language and Hearing Research* 48: 1468-1480.
- CASE Taxonomy for non-verbal elements (2017) Birkenfeld: Trier University of Applied Sciences.
- Colletta, J.-M., Venouil, A., Kunene, R., Simon, J.-P., & Kaufmann, V. (2009) Multitrack Annotation of Child Language and Gestures. In Kipp, M., Martin, J.-C., Paggio, P., & Heylen, D. (eds) *Multimodal corpora: from models of natural interaction to systems and applications*. Berlin, Heidelberg: Springer-Verlag, 54–72.
- Corballis, M. (2010) The gestural origins of language. *WIREs Cognitive Science* 1: 2-7.
- Council of Europe (2018) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*.
- Darwin, C. (1872) *The expression of the emotions in man and animals*. London: John Murray.
- Davidse, K., Vandelanotte, L., Gentens, C., & Ghèsquière, L. (2014) Interrogating corpora to describe grammatical patterns. In Davidse, K., Gentens, C., Ghesquière, L., & Vandelanotte, L. (eds) *Corpus Interrogation and Grammatical Patterns*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1-11.

- Dortier, J.-F. (2003) Langage et évolution : nouvelles hypothèses. *Sciences Humaines* 114. Retrieved May 15, 2019, from https://www.scienceshumaines.com/langage-et-evolution-nouvelles-hypotheses_fr_3646.html.
- Driskell, J. (2003) The Effect of Gesture on Speech Production and Comprehension. *Human Factors The Journal of the Human Factors and Ergonomics Society* 45(3): 445-454.
- Dumont, A. (2018) *Fluency and disfluency: a corpus study of non-native and native speaker (dis)fluency profiles*. Unpublished PhD thesis, Université catholique de Louvain.
- Ekman, P., & Friesen, W.V. (1969) The repertoire of nonverbal behaviour: categories, origins, usage and coding. *Semiotica* 1(1): 49-98.
- Fast, J. (1970) *Body Language*. New-York: Simon & Schuster, Inc.
- Fillmore, C. (1991) "Corpus linguistics" vs. "computer-aided armchair linguistics". *Directions in Corpus Linguistics: Proceedings from a 1991 Nobel Symposium on Corpus Linguistics*. Stockholm: Mouton de Gruyter, 35-66.
- Esteve-Gibert, N., & Prieto, P. (2013) Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication* 57: 301-316.
- Gilquin, G., Granger, S., & Paquot, M. (2007) Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes* 6(4): 319-335.
- Goldin-Meadow, S., & Alibali, M.W. (2013) Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*. 64: 257-283.
- Goring, S. A. (2016) *The impact of co-verbal gestures on speech fluency for object descriptions*. Unpublished Master's thesis. University of Missouri, Kansas City.
- Granger, S. (2002) A bird's-eye view of learner corpus research. In Granger, S., Hung, J., & Petch-Tyson, S. (eds), *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, 3-33.
- Graziano, M., Gullberg, M. (2018) When Speech Stops, Gesture Stops: Evidence From Developmental and Crosslinguistic Comparisons. In *Frontiers in Psychology* 9: 1-17.
- Gregersen, T. (2007) Language learning beyond words: Incorporating language into classroom activities. *Reflections on English Language Teaching* 6(1): 51-64.
- Gregersen, T., MacIntyre, P. (2017) *Optimizing Language Learner's Nonverbal Behavior*. Bristol: Multilingual Matters.
- Gullberg, M. (1998) *Gesture as a communication strategy in second language discourse. A study of learners of French and Swedish*. Sweden: Lund University Press.
- Gullberg, M. (2006) Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon). *International Review of Applied Linguistics* 44(3): 103-124.

- Gullberg, M. (2014) Gestures and second language acquisition. In Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D., & Bressemer, J. (eds), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction Vol. 2*. Berlin/Boston: De Gruyter Mouton, 1868-1875.
- Gullberg, M., de Bot, K., & Volterra, V. (2008) Gesture and some key issues in the study of language development. *Gesture* 8(2): 149-179.
- Graziano, M., Gullberg, M. (2018) When Speech Stops, Gesture Stops: Evidence From Developmental and Crosslinguistic Comparisons. In *Frontiers in Psychology* 9: 1-17.
- Hahn, F., & Rieser, H. (2010) Explaining Speech Gesture Alignment in MM Dialogue Using Gesture Typology. *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue*, 99-109.
- Hişmanoglu, M. & S. Hişmanoglu (2008) The use of body language in foreign language learning and teaching. *Journal of Social Sciences* 19: 165-178.
- Hoetjes, M., Emiel, K. & Swerts, M. (2014) Does our speech change when we cannot gesture? *Speech Communication* 57(2014): 257-267.
- Holler, J. & G. Bettie (2003). How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process? *Semiotica* 146 (1-4): 81-116.
- Hu, Z., Dick, M, Chang, C., Bowden, K., Neff, M., Tree, J. E., & Walker, M. A. (2016) A Corpus of Gesture-Annotated Dialogues for Monologue-to-Dialogue Generation from Personal Narratives. *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*. Portorož, Slovenia, 3447-3454.
- Inaky, A., & Okita, T. (2006) A Small-Corpus-Based Approach to Alice's Roles. *Literary and Linguistic Computing* 21(3): 283-294.
- Ipek, H. (2009) Comparing and Contrasting First and Second Language Acquisition: Implication for Language Teachers. *English Language Teaching* 2(2): 155-163.
- Kachru, B. (1985) Standards, codification and sociolinguistic realism: The English language in the outer circle. In Quirk, R., & Widdowson, H.G. (eds), *English in the world: teaching and learning the language and literatures*. Cambridge: Cambridge University Press, 11-30.
- Kachru, B (1992) Models for Non-Native Englishes. In Kachru, B. (ed), *The Other Tongue; English Across Cultures (2nd edition)*. Urbana and Chicago: University of Illinois Press, 48-70.
- Kendon A. (1983) The Study of Gesture: Some Remarks on its History. In Deely J.N., & Lenhart M.D. (eds) *Semiotics 1981*. New York: Plenum Press.
- Kendon, A. (2004) *Gesture. Visible Action as Utterance*. Cambridge: Cambridge University Press.

- Kirschner, P. (2016) Cognitieve Belasting Theorie: Leuker kunnen we het maken, en ook makkelijker. Retrieved October 27, 2019, from <https://onderzoekonderwijs.net/2016/01/03/cognitieve-belasting-theorie-leuker-kunnen-we-hetmaken-en-ook-makkelijker/>.
- Kong, A.P.H., Law, S.-P., Kwan, C.C.Y., Lai, C., & Lam, V. (2015) A coding system with independent annotations of gesture forms and functions during verbal communication: Development of a Database of Speech and GEsture (DoSaGE). In *Journal of Nonverbal Behavior*, 39(1): 93-111.
- Koutsombogera, M., & Vogel, C. (2017) Ethical responsibilities of researchers and participants in the development of multimodal interaction corpora. *8th IEEE Conference on Cognitive Infocommunications*. Debrecen, Hungary, 277-282.
- Lee, D., & McGarrell, H. (2010) Corpus-Based/Corpus-Informed English Language Learner Grammar Textbooks: An Example of how Research Informs Pedagogy. *Teachers of English as a second language of Ontario* 37(2): 78-100.
- Leech, G. (1997) Introducing corpus annotation. In Garside, R., Leech, G. & McEnery, A. (eds) *Corpus Annotation. Linguistic Information from Computer Text Corpora*. London: Longman, 1-18.
- Littlewood, W. (1984) *Foreign and Second Language Learning: Language Acquisition Research and its implications for the classroom*. United Kingdom: Cambridge University Press.
- Macedonia, M. & K. von Kriegstein (2012) Gestures Enhance Foreign Language Learning. *Biolinguistics* 6(3-4): 393-416.
- McEnery, T. & Gabrielatos, C. (2006) English corpus linguistics. In Aarts, B., & McMahon, A. (eds.), *The Handbook of English Linguistics*. Oxford: Blackwell, 33-71.
- McEnery, T., Xiao, R., & Tono, Y. (2006) *Corpus-based Language Studies: An advanced Resource Book*. New York: Routledge.
- McNeill, D. (1992) *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- Mettewie, L. (2017-2018) *LNRLB301: Langue et linguistique néerlandaise. Questions approfondies. De verwerving van het Nederlands als NT2 of NVT [Syllabus]*. Namur: Faculté de Philosophie et Lettres, Département Langues Germaniques, Université de Namur.
- Morett, L., Gibbs, R., & MacWhinney, B. (2012) The Role of Gesture in Second Language Learning: Communication, Acquisition, & Retention. *Proceedings of the Annual Meeting of the Cognitive Science Society* 34: 773-778.
- Nicoladis, E., Nagpal, J., Marentette, P., Hauer, B. (2018) Gesture frequency is linked to story-telling style: evidence from bilinguals. *Language and Cognition* 10: 641-664.

- Nespoulous, J. L., & Lecours, A. R. (1986) Gestures: Nature and Function. In Nespoulous, J. L., Perron, P., & Lecours, A. R. (eds), *The Biological Foundations of Gesture: Motor and Semiotic Aspects*. Hillsdale/ New Jersey: Lawrence Erlbaum Associates, 49-62.
- Olivier, N. (2014) La communication : Quelle est la place du non verbal dans notre communication ? (Graduation Work.) Institut Saint-Michel, Neufchâteau.
- Privsec Report (2017) *Data masking: anonymisation or pseudonymisation?* Retrieved March 22, 2020, from <https://gdpr.report/news/2017/09/28/data-masking-anonymization-pseudonymization/>.
- Punchihetti, S. (2013) First, second and foreign language learning: how distinctive are they from one another? *Proceedings of The European Conference on Language Learning 2013*. University of Sri Jayawardenapura, Sri Lanka.
- Reffay, C., & Teutsch, P. (2007) Anonymisation de corpus réutilisables : Masquer l'identité sans altérer l'analyse des interactions. *Proceedings of the EIAH'2007 Conference: Environnements Informatiques pour l'Apprentissage Humain*. Lausanne, France.
- Rosborough, A. (2014) Gesture, meaning-making, and embodiment: Second language learning in an elementary classroom. *Journal of Pedagogy* 5(2): 227-250.
- Russel, A., Brugman, H., Broeder, D., & Wittenburg, P. (2000) The EUDICO project, multi media annotation over the internet. In *Proceedings of the Second International Conference on Language Resources and Evaluation*. Athens, Greece.
- Schneider, E. (2007) *Postcolonial English: Varieties around the world*. Cambridge, UK: Cambridge University Press.
- Schneider, E. (2009) Towards Endonormativity? African English and the Dynamic Model of the Evolution of Postcolonial Englishes. In Harrow, K. & Mpoche, K. (eds) *Language, Literatures and Education in Multicultural Societies: Collaborative Research on Africa*. Newcastle, UK: Cambridge Scholars Publishing, 283-305.
- Stam, G. (2001) Lexical failure and gesture in second language development. In Cavé, C., Guaïtella, I., & Santi, S. (Eds) *Oralité et gestualité : interactions et comportements multimodaux dans la communication*. Paris: L'Harmattan, 271-275.
- Stam, G. (2012) Gestes et Recherche Lexicale en Langue Seconde. In Vion, R., Giacomi, A. & Vargas, A. (eds.), *La corporalité du langage : Multimodalité, discours et écriture*. Aix-en-Provence, France: Presses Universitaires de Provence, 55-71.
- Stubbs, M. (2004) Language corpora. In Davies, A, & Elder, C. (eds), *Handbook of Applied Linguistics*. Oxford: Blackwell, 106-132.
- Tellier, M., Stam, G. (2010) Découvrir le pouvoir de ses mains : La gestuelle des futurs enseignants de langue. In *Spécificités et diversité des interactions didactiques : disciplines, finalités, contextes*. Lyon, France: 1-14.

- Tellier, M., Azaoui, B., Saubesty, J. (2012) Segmentation et annotation du geste : Méthodologie pour travailler en équipe. In *JEP-TALN-RECITAL 2012 : Atelier DEGELS 2012 (Défi GEste Langue des Signes)*. Grenoble, France, 41-55.
- Thewissen, J. (2012) *Accuracy across proficiency levels : insights from an error-tagged EFL learner corpus*. Unpublished PhD thesis, Université catholique de Louvain.
- Thompson, P. (2006) Assessing the contribution of corpora to EAP practice. In Kantaridou, Z., Papadopoulou, I., & Mahili, I (eds), *Motivation in Learning Language for Specific and Academic Purposes*. Macedonia: University of Macedonia.
- Torres, N. (2019) When you pitch an idea, gestures matter more than words. *Harvard Business Review* May-June 2019: 33-37. Retrieved December 11, 2019, from <https://hbr.org/2019/05/when-you-pitch-an-idea-gestures-matter-more-than-words>.
- Vandelanotte, L., & Dancygier, B. (2017) Multimodal artefacts and the texture of viewpoint. *Journal of Pragmatics* 122: 1-9.
- Wittenburg, P., Levinson, St., Kita, S., & Brugman, H. (2002) Multimodal Annotations in Gesture and Sign Language Studies. In *Third international conference on language resources and evaluation, European Language Resources Association*.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté de philosophie, arts et lettres

Place Blaise Pascal, 1 bte L3.03.11, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/fial