Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Studi linguistici e letterari

CORSO DI DOTTORATO DI RICERCA IN: SCIENZE LINGUISTICHE, FILOLOGICHE E LETTERARIE
CURRICOLO: LINGUE E LETTERATURE ANGLO-GERMANICHE

CICLO: XXIX

Gesture and prosody: cognitive and communicative effort in L1 and L2

Coordinatore: Ch.ma Prof.ssa Annalisa Oboe

Supervisore: Ch.ma Prof.ssa Maria Grazia Busà

Dottoranda: Sara Brugnerotto

Contents

List of figures **7**

List of tables **9**

Abstract **11**

Riassunto **15**

Introduction **21**

Part I | Theoretical Background **25**

1/ Gestures **27**

1. Defining Gestures **27**

2. Recognizing Gestural Actions **28**

3. Anatomy and Structure of Gestures **30**

3.1 Gesture structure 30

3.2 Gesture space 33

4. Classifying Gestures **34**

4.1 A semiotic categorization, Wilhelm Wundt 34

4.2 Categorizations of gestures and their relation with speech 35

4.2.1 David Efron 35

4.2.2 Paul Ekman and Wallace Friesen: the repertoire of nonverbal behavior 37

4.2.3 David McNeill 43

4.2.4 Kendon's continuum 47

4.2.5 Speech interaction with gestures – Pragmatic meanings or rhythmic markers 48

2/ Gestures in relation with speech **51**

1. Introduction **51**

2. Speaking and Gesturing. When Words are Complemented by Gestures **53**

3. Gestures in L1 and L2 Acquisition **54**

3.1 L1 acquisition 54

3.2 L2 acquisition 56

4. Gestures and narration **59**

3/ Cognitive and communicative functions of gesture 61
1. Introduction 61
2. Gesture, cognition and thinking for speaking 62
3. Beyond cognition. The communicative function of gestures 66
3.1 The role of representational gestures in communication. 72
4/ About speech prosody and its relations with gestures 77
1. Prosody 77
2. On Intonation 78
3. About Pitch Range 79
4. Pitch in L2 80
4.1 Measuring global pitch range 82
5. Prosody and Gestures 82
6. Pitch and Gestures 84
Part II | Experimental analyses 87
5/ Does communicative effort influence pitch and gestures? 89
1. A note on terminology 89
2. Introduction 89
3. Subjects and Materials 92
4. Procedure 92
5. Methods and Analysis 94
6. Gesture data 96
6.1 The gesturing styles: some examples from the corpora 99
7. Results 105
7.1 About Pitch 105
7.2 Speech data 106
7.3 Occurrence of gesture types 110
7.4 Relation of pitch and gestures 112
8. Discussion 114
6/ Does repetition help to be more communicative? Experiment 2 119
1.Introduction 119
2. Subjects and Materials 120
3.Experiment procedure 121

4. Methods and Analysis **123**
5. Prosodic and Speech data **123**
6. Gesture data **125**
6.1 Gesturing styles and annotation criteria. Some examples from the corpora 127
7. Results **130**
7.1 About Pitch 130
7.2 Speech data 132
7.3 Gesture data 135
8. A little insight into the vagueness issue **138**
9. Discussion **139**
7 / Integrating the results **143**
Conclusions **147**
References **153**
Appendix 1 ..... **173**
Appendix 2 ..... **175**
Appendix 3 ..... **177**
Appendix 4 ..... **179**
Appendix 5 ..... **183**
Appendix 6 ..... **185**

List of figures

Fig. 1. Kendon's hierarchy 30
Fig. 2. Example taken and adapted from Kendon 2004, pp.120-121 32
Fig. 3. Gesture space, taken from McNeill, 1992; p. 89 33
Fig. 4 Relationships among types of 'usage' according to Ekman & Friesen, 1969 39
Fig. 5. Metaphoric gesture 44
Fig. 6. Deictic gesture 45
Fig. 7. The relation between iconic gestures generation and speech production 63
Fig. 8. Experimental set 93
Fig. 9. Annotation in Elan 96
Fig. 10. Example of iconic gesture 97
Fig. 11 Student C in her first (on the left) and second telling of the fable in English 101
Fig. 12 Student B while conveying the sentence "it picks up the cheese" 101
Fig. 13 Beat Gestures (annotated discursive) in the three different conditions 102
Fig. 14 Cyclic, vague and other pragmatic gestures 103
Fig. 15 Emblem occurrence 104
Fig. 16. Anova results for PVQ values 106
Fig. 17. Comparison of the frequency of co-occurrence of pauses in relation to spoken words in the three datasets 108
Fig. 18 Gesture occurrences (in raw numbers) for every macro category 110
Fig. 19. Comparison between the pitch variation and the Gesture Rate 113
Fig. 20. Experimental set 122
Fig. 21. Example of Elan annotation 125
Fig. 22 Students G and C performing iconic gestures 127
Fig. 23 Student A while producing metaphoric and abstract deictic gestures 128
Fig. 24 discursive gestures 129
Fig. 25 PVQ in the four conditions 131
Fig. 26. Gesture sequence in which the Speaker L is saying "the cheese falls" 135
Fig. 27. Occurrencies of the two macro-categories in the four conditions 135

Fig. 28 Total Gesture rate 136
Fig. 29 Gesture rates of the four conditions for each gesture type 137

List of tables

Tab. 1 The seven types of illustrators (Ekman and Friesen, 1969; Ekman, 2004) 41
Tab. 2. Gesture categorization comparison, Efron, Ekman & Friesen and McNeill 46
Tab. 3 L1 development (Masur 1990) 55
Tab. 4 Gesture use in relation to age (adapted from Gullberg, 1998) 56
Tab. 5 gesture macro-categorization 99
Tab. 6. Pitch values for the three experimental conditions 105
Tab. 7. Speech data in the three experimental conditions 106
Tab. 8 Gesture categorization 126
Tab. 9. Pitch values in the four experimental conditions 131
Tab. 10 Average duration of speech features 132
Tab. 11 Average number of occurrencies in speech and speech rate 133
Tab. 12. Percentage of gesture usage in each experimental condition 136
Tab. 13. Gesture rates 137
Tab. 14. Gesture rate for the vagueness category 138
Tab. 15. Speech rate in the four conditions 138
Tab. 16. Pitch Variation Quotient for the four conditions 139
Tab. 17. Comparison of the two investigations results 146
Tab. 18. Comparison of the PVQ results 148

Abstract

Gestures and speech are two interconnected features of human communication. Studies have shown that they develop together and they both covey semantic and pragmatic meanings (Kendon, 2004; McNeill, 1992). Furthermore, gestures seem to be also linked to the prosodic features of speech since they also share synchronicity aspects and are often temporally aligned (McClave, 1991; Esteve-Gibert & Prieto, 2013).

The co-production of gestures and speech, thus, seems to have a number of different functions both at the cognitive and at the communicative levels. On the one hand, in fact, gestures seem to help the process of speaking: they have a scaffolding function in the lexical and rhythmical organization of speech and they help information packaging (Butterworth & Beattie, 1978; Esteve-Gibert et al., 2014; Krauss et al, 1996; Kita, 2000, 2010). This is supported by the evidence that speakers gesticulate also when they don't see their addressees (for example when they are speaking over the phone (Cohen & Harrison, 1973; de Ruiter, 2003). On the other hand, gesturing is intended to communicate and is part of the speaker's communicative effort (Kendon, 2004).

Gestures also play an important role in L2 development and communicative strategies (Gullberg, 1998) and it is possible that L1 gestures influence gestures during L2 language development, and that gestural transfer co-occurs with linguistic transfer (Brown & Gullberg, 2008; Pika, Nicoladis, Marentette, 2006). It has also been suggested that, since gestures play an important role in facilitating language access in speech production, bilingual/L2 speakers may gesture more than monolingual speakers. This would be due to the cognitive load deriving from speaking a different language from the native one (Kita, 2000; Krauss & Hadar, 1999).

Until now, gestures and prosody have been studied mostly in their temporal interaction and coordination (McClave, 1991, Loehr, 2004) or in the possible similarities between their pragmatic functions (Tuite, 1993). Little attention has been paid to the relationship between speakers' global pitch range and use of gestures (and gesture categories) in conditions of high cognitive effort, or when they increase their communicative effort in their speeches. In fact, there

seem to be a lack of scientific evidence about the use of pitch and gestures variability as a communicative strategy of the speaker.

The aim of this thesis is to investigate aspects of the role and functions of gestures and speech in L1 and L2 and how the cognitive and communicative efforts may influence speakers' global pitch range and the use of gestures in a story telling task.

The investigation consists of two experiments. The first one (presented in chapter 5) aims to examine if speakers' increase in the communicative effort produces changes in pitch variation and in the use of different categories of gestures. The second one (presented in chapter 6) examines the effects of speakers' decrease in cognitive effort in the production of speech and gestures.

In the first experiment 8 Italian speakers, Italian learners of English (L2) and students of a Public Speaking class, were asked to read and tell a fable in English to their classmates (Italian was used only as a control and recorded only one time). One week later, the subjects were asked to repeat the task with the instruction to be as communicative as possible. The audiovisual material was analyzed with the software Praat (phonetic analysis) and Elan (gesture analysis). For speech, the aim was to examine possible changes in the speakers' fluency and pitch variation after they received the instruction to be communicative; for gestures, the aim was to examine the variations in both speakers' overall gesturing and representational gestures in the communicative task. The hypothesis tested was that the communicative task causes an increase in fluency in the L2 (shown through features like a higher speech rate and a decrease of disfluencies and pauses), and that the communicative task leads to an increased use of representational gestures, since they might be considered as helpful for the addressee to better understand what is said. The results of the first experiment led to the conclusion that if a person is asked to be communicative in the L2, they will probably increase both their pitch variation and the total number of gestures produced, with a significant increase in iconic and representational gestures. The results on pitch variation, though, do not exclude the possible effect of other variables, first of all of task repetition. In fact, it is possible that by repeating the task, the speakers became more confident and could be more focused on being more communicative in telling the story.

The second experiment tests the effect of task repetition on L2 speech and gesture. In this experiment (chapter 6) only the cognitive facilitation and the effect of the decrease in cognitive load was considered. This time, in fact, the subjects (10 Italian students of English L2) were

asked to watch a short cartoon and to tell the story, both in Italian and English, in front of a small audience while being video-recorded. The subjects were asked to repeat the task one week later in the two languages with the same modalities. The analyses followed the same procedure as the previous experiment (chapter 5). The hypotheses tested were that repetition itself could influence the communicativeness of the speakers in both the speech and the gesturing levels, with an increased fluency and a different use of gestures compared to the first narration attempt. This could be caused by the facilitation of the task in terms of cognitive load and better memorization. The results showed that with the decrease of the cognitive load, speakers reach a higher fluency in L2. In Italian, however, the speakers show no significant difference in fluency or liveliness (Hincks, 2005) and this might be due to the fact that repetition and memorization did not help their fluency as much as they did in the L2. As for gestures, the speakers employed a greater number overall gestures when speaking English, their L2. The use of gestures, though, did not change in spite of the repetition, with a greater use of discursive gestures even if representational gestures seem to be used more in the L2 than in Italian.

Overall, the investigation shows that the results of the two experiments can be integrated and offer a wider picture on the use of gesture and the employment of representationality with the conscious intent to be communicative. When people are asked to tell a story in a more communicative way, their response is to increase the representationality carried in their gestures and consequently use more representational gestures, this does not occur whenever the subjects only repeat the narration with a lighter cognitive load.

Riassunto

La comunicazione parlata è fortemente legata a quella gestuale. Il sistema gestuale e quello del parlato, infatti, sono considerati parti integranti di uno stesso sistema linguistico all'interno del quale si sviluppano assieme e si condizionano e influenzano l'un l'altro (Kendon, 2004; McNeill, 1992). I gesti, infatti, supportano l'organizzazione e "l'impacchettamento" delle informazioni espresse nel parlato (*Information packaging*, Kita 2000, Alibali & Kita, 2010). I gesti, inoltre, integrano e contribuiscono alla resa semantica e pragmatica del parlato (Kendon, 2004). Inoltre, la gestualità sembra ricoprire un ruolo anche per quanto riguarda alcuni aspetti ritmico prosodici del parlato, sia nella prima che nella seconda lingua (Esteve-Gibert & Prieto, 2014; McCafferty, 2004).

Generalmente, gli studi che mettono in relazione prosodia e parlato si concentrano sulla sincronicità dei picchi intonativi con determinate categorie gestuali (generalmente i *beats* o *batonici,* caratterizzati da movimenti ritmici di dita, mani o braccia). Molta meno attenzione è stata data alla relazione esistente fra la variabilità intonativa e l'uso delle categorie gestuali. Mancano quindi studi che investighino il comportamento di gesti e prosodia in condizioni di comunicatività e sforzo cognitivo differente. Anche se diverse lingue hanno diverse caratteristiche prosodiche e intonative che le contraddistinguono, esistono comunque degli *universali prosodici* (Ladd, 2008) che caratterizzano il parlato e ne comunicano alcuni aspetti pragmatici (un universale prosodico, ad esempio, è l'alzarsi del contorno intonativo quando la frase viene interrotta da una pausa ma non è ancora considerata conclusa dal parlante). Per esempio, è stato dimostrato che la variabilità dell'intonazione influenza la percezione della vivacità del discorso (Hincks, 2004, 2005). Ne potrebbe essere una dimostrazione il fatto che, nei corsi di *public speaking*, gli studenti vengano incitati a modulare il tono di voce in modo tale che risulti variato. Un'altra istruzione che viene data agli studenti di *public speaking* riguarda la modulazione del proprio linguaggio del corpo e, quindi, della gestualità. Slemberebbe in fatti che anche la variabilità del linguaggio del corpo abbia come effetto la percezione di una maggiore vivacità comunicativa e aiuti quindi l'ascoltatore.

Tuttavia, nel panorama della ricerca scientifica, scarseggiano gli studi che dimostrano l'esistenza una relazione tra la variabilità dell'intonazione e dei gesti del parlante. Una parte consistente del lavoro di questa tesi è infatti rivolto all'approfondimento della funzione comunicativa di gesti e intonazione e alla verifica di possibili correlazioni fra questi due aspetti.

Un'altra questione che viene affrontata all'interno di questa tesi è l'effetto della ripetizione e la conseguente facilitazione del compito di narrazione in L2. Comunicare in una lingua straniera, infatti, comporta un complesso sforzo cognitivo: oltre a dover organizzare il discorso e recuperare il lessico adeguato, il parlante deve conferire al proprio discorso anche la pronuncia, la prosodia e il ritmo della lingua che sta parlando. Questo sforzo cognitivo può portare a una diminuzione della fluenza, ad un aumento di esitazioni nel parlato ed a una minore variabilità intonativa (Zimmerer et al. 2014). Un aspetto che verrà investigato all'interno di questo lavoro riguarda, appunto, il rapporto fra la facilitazione del compito narrativo in L2, attraverso la ripetizione del task narrativo, e l'uso di prosodia e gesti.

Questa tesi si inserisce negli studi relativi a gesti e prosodia e si focalizza sulle differenze gestuali e prosodico-intonative in parlanti L2 di inglese che raccontano una favola e la ripetono poi a distanza di una settimana. L'ipotesi che ho voluto testare è che quando cambia il contesto comunicativo cambia anche la gestualità e alcuni aspetti ritmici e linguistici del parlato. In particolare, quando domina la funzione cognitiva del parlato, e quindi il soggetto si concentra sulla struttura della storia, prevalgono gesti di tipo ritmico-discorsivo anche associati alle esitazioni del parlante; quando invece domina la funzione comunicativa, aumentano i gesti di tipo rappresentativo e diminuiscono le esitazioni nel parlato.

Il lavoro è strutturato in due parti: la prima parte, *Theoretical Background,* è intesa come la presentazione dell'argomento e della letteratura riguardante i temi d'interesse. La seconda parte riguarda i due esperimenti condotti per testare le ipotesi di ricerca.

Più nello specifico questo lavoro è strutturato in 7 capitoli così suddivisi:

Dopo l'introduzione al lavoro generale, il primo capitolo riguarda una presentazione generale sugli studi sulla gestualità, iniziando dalla definizione di gesto e dai criteri che rendono possibile la distinzione di un gesto da un non-gesto. In questo capitolo, inoltre, viene spiegata l'anatomia dei gesti e l'uso dello spazio da parte dei parlanti. Infine, vengono presentate le principali categorizzazioni che sono state date dai principali studiosi. Lo studio dei gesti, infatti, è una

disciplina relativamente giovane, che si è sviluppata soprattutto a partire dalla seconda metà del novecento e per la quale non sono ancora state definite categorie univoche.

Il seguente capitolo (capitolo 2), tratta la relazione dei gesti con il parlato. Il rapporto esistente fra parlato e gesti non è del tutto chiaro. Se da una parte, infatti, i gesti e il parlato sono considerati parte di uno stesso sistema linguistico che si sviluppa e interagisce nell'atto comunicativo (McNeill, 1992; Kendon, 2004), dall'altra essi vengono considerati come il risultato di due processi indipendenti (Hostetter, Alibali & Kita, 2007). In generale, comunque, i gesti possono in alcuni casi sostituirsi al parlato, in altri, arricchire il significato semantico e pragmatico di ciò che viene detto, o possono marcarne il ritmo (McNeill, 1992). All'interno di questo capitolo, viene inoltre discusso il rapporto che gesti e parlato hanno sullo sviluppo della prima lingua e sull'apprendimento della seconda lingua, nonché le funzioni che i gesti hanno nella narrazione.

Il capitolo 3 si focalizza sul dibattito riguardante le funzioni cognitive e comunicative dei gesti partendo da considerazioni e studi da considerarsi fondativi in questo ambito. Nella prima parte del capitolo vengono presentati i più importanti studi a supporto dell'idea che i gesti siano principalmente legati alle funzioni cognitive e quindi organizzative e lessicali del parlato (v. Beattie & Coughlan, 1999; Hadar & Butterworth, 1997; Rauscher et al., 1996; Kita, 2000, 2010; Krauss et al., 1996). La seconda parte del capitolo illustra gli studi che si concentrano sulla funzione comunicativa e pragmatica che i gesti ricoprono nel parlato (Bavelas et al. 2008; Kendon, 1985, 2004) e, nello specifico, su quale sia il ruolo che i gesti iconici hanno nella comunicazione (v. Cohen & Harrison, 1973; Melinger & Levelt, 2004),

Il quarto capitolo è dedicato alla prosodia e alla sua relazione con il sistema gestuale nel parlato. In questo capitolo vengono chiariti i concetti di intonazione e variabilità intonativa sia in L1 che in L2 (v. Hincks, 2004; Mennen et al. 2007; Patterson, 2000).

La seconda parte, *Experimental analyses*, comprende i capitoli 5, 6 e 7 dedicati al lavoro sperimentale.

Nel quinto capitolo viene esposto il primo dei due lavori sperimentali. L'ipotesi di partenza è quella che un maggiore sforzo comunicativo da parte del parlante L2 abbia come effetti, da un lato, una maggiore variabilità intonativa rispetto a quella della L1; dall'altra un uso distinto dei gesti rispetto alla condizione L2 nella quale tale sforzo comunicativo è meno consapevole. Il primo dei due effetti, la maggiore variabilità intonativa, sarebbe dovuto al fatto che i soggetti si

concentrano maggiormente nella resa comunicativa del parlato e quindi riescono a conferire una variabilità intonativa. In particolare, i parlanti riescono ad avere un'intonazione più simile a quella della L2 (che in questo caso è l'inglese, generalmente percepito come una lingua con maggiore variabilità rispetto all'italiano). Il secondo effetto sarebbe invece dovuto al fatto che i soggetti, per risultare più comunicativi, si concentrerebbero su quei gesti che considerano più importanti a livello comunicativo e percettivo.

Il primo esperimento ha coinvolto otto studentesse, frequentanti un corso di Public Speaking all'interno dell'Università di Padova. Alle studentesse è stata fatta leggere la favola di Esopo "Il corvo e la volpe" ed è stato chiesto loro di raccontare la storia due volte: la prima senza ricevere alcuna istruzione specifica e la seconda con la richiesta di cercare di essere il più comunicative possibile. L'analisi del corpus di materiale audiovisivo è stata effettuata:

- Tramite un'analisi fonetico-prosodica degli audio (con l'utilizzo del software Praat http://www.fon.hum.uva.nl/praat/) con il quale sono stati misurati: disfluenze, ripetizioni, correzioni, pause piene, pause silenti e pause respiratorie, altezza tonale. Sono state inoltre conteggiate le parole.
- Tramite l'analisi gestuale del corpus di video (con l'utilizzo del software Elan, https://tla.mpi.nl/tools/tla-tools/elan/download/) con cui sono stati analizzati i gesti concomitanti al parlato suddividendoli in macro-categorie (*rappresentazionali,* che descrivono caratteristiche fisico o metaforiche dell'oggetto del discorso; *discorsivi,* che hanno una relazione ritmica con il parlato e che ne marcano continuità e coesione; *emblemi,* gesti tipici di una cultura e che sono capiti e condivisi solo da chi fa parte di un determinato gruppo linguistico e/o sociale).

I risultati di questo primo esperimento suggeriscono che un maggiore sforzo comunicativo porta ad un aumento della quantità di gesti totali e della percentuale di gesti di tipo iconico-rappresentazionali, oltre che ad una maggiore fluenza, e ad un uso più variato dell'altezza tonale da parte dei soggetti sperimentali. Tale risultato porta alla conclusione che, quando viene richiesto loro di essere comunicativi, i parlanti migliorano la fluenza del parlato, modificano la variazione dell'altezza tonale e implementano un numero maggiore di gesti iconici.

Il sesto capitolo è dedicato alla seconda ricerca sperimentale, nella quale le parlanti sono state registrate fuori dal contesto di insegnamento e, nella seconda ripetizione, non è stato chiesto loro di concentrarsi sulla resa comunicativa del proprio racconto. L'esperimento mira ad integrare i

risultati raggiunti dall'esperimento con focus sulla resa comunicativa dei soggetti e vuole verificare quali siano le differenze prosodico-gestuali in un contesto in cui alle parlanti (in questo caso sia di L1 che di L2) viene richiesto solamente di ricordare e ripetere il racconto senza dover migliorare la propria performance comunicativa.

Nel secondo esperimento sono state coinvolte 10 studentesse di lingue e linguistica di età compresa fra i 22 e i 25 anni, con livello B2 di inglese (CEFR).

Alle studentesse è stato fatto vedere un video con un breve cartone animato che illustrava la favola "Il Corvo e la Volpe" di Esopo. Subito dopo aver visto il cartone, le ragazze sono state videoregistrate mentre raccontavano la storia sia in italiano che in inglese. Le stesse studentesse hanno ripetuto l'esperimento dopo una settimana. Per evitare l'effetto di *common ground*, per il quale si omettono informazioni verbali quando l'interlocutore condivide o è a conoscenza di ciò di cui si sta parlando, il pubblico al quale le studentesse raccontavano la storia è stato cambiato ad ogni ripetizione, sia in italiano che in inglese.

Le seguenti analisi sono state effettuate sul materiale audiovisivo:

- Analisi fonetico-prosodica del materiale audio con il software di analisi acustica Praat (disfluenze, ripetizioni, correzioni, pause piene, pause silenti e pause respiratorie, altezza tonale). Sono state inoltre conteggiate le parole.

- Analisi gestuale del materiale visivo con il software di analisi multimodale Elan. Per poter confrontare i risultati con quelli ottenuti nel primo esperimento, l'annotazione dei gesti co-occorrenti al parlato ha mantenuto la stessa categorizzazione dei gesti in macro-categorie: *representational*, che comprende tutti i gesti di tipo rappresentazionale, che indicano o rappresentano fisicamente o metaforicamente l'oggetto o l'azione di cui si parla; *discursive* (ritmico-discorsivi), cioè tutti quei gesti che aiutano il parlante ad organizzare il proprio discorso e/o a marcarne il ritmo.

I risultati di questo esperimento hanno mostrato che, nonostante ci sia un effetto di ripetizione, che comporta un miglioramento nella fluenza del parlato (soprattutto in L2, condizione nella quale la differenza fra le due ripetizioni è risultata statisticamente significativa), le parlanti non modificano significativamente i gesti.

Il settimo capitolo è dedicato all'integrazione dei risultati e delle conclusioni delle due ricerche e quindi alle considerazioni finali. I risultati dei due esperimenti conducono alla conclusione che all'aumentare dello sforzo comunicativo da parte dei parlanti corrisponde anche una maggiore gestualità, visibile in particolare nell'aumentare della quantità di gesti appartenenti alla macro-categoria rappresentativa (*representational*). Allo stesso tempo, la riduzione dello sforzo cognitivo nella seconda ripetizione comporta effetti sia sul parlato che sui gesti. Nel parlato, il fatto di essere di fronte ad un compito relativamente più facile (per effetto della ripetizione della storia) porta le studentesse a migliorare la loro fluenza in L2 (si riducono le disfluenze come esitazioni, pause piene, correzioni e ripetizioni). La maggiore facilità nel raccontare la storia, inoltre, permette di aumentare significativamente la variazione dell'altezza tonale nella seconda ripetizione in L2, consentendo alle studentesse di imitare le caratteristiche prosodiche dell'inglese. Dal punto di vista gestuale, invece, non si presentano differenze statisticamente rilevanti fra le ripetizioni, né in italiano, né in inglese. L'unica differenza significativa rilevata è quella dell'uso della categoria rappresentazionale fra la prima ripetizione in Italiano e la prima in inglese. Quando raccontano la storia in inglese, infatti, le studentesse usano un maggior numero di gesti di tipo rappresentazionale, suggerendo così che la rappresentazionalità dei gesti non è solo effetto del maggior intento comunicativo (come si era visto nel primo esperimento) ma è anche effetto del minor sforzo cognitivo.

La tesi termina con le considerazioni finali e conclusive sul lavoro svolto e sui risultati ottenuti.

Introduction

Gestures and speech are an interrelated system and develop together. For example, it seems that gestures help reducing the cognitive load of the speakers (Goldin-Meadow et al., 2001; Mol et al., 2009), they seem to help fastening the learning processes (Goldin-Meadow, 2000; Roth, 2001). Furthermore, gestures seem to help speakers organize visual information with speech (Kita, 2000, 2010) but they are also an important feature that supports the delivery of semantic and pragmatic aspects of speech (Kendon, 2004; McNeill, 1992). The functions of gestures in speech are still objects of debate, and this will be illustrated in the present work.

Another important aspect of communication is represented by speech characteristics and prosody. Besides being different for every language, prosody seems to have an effect on the perception of liveliness in speech and thus, when a speaker intends to deliver a lively speech, they will probably use a more varied pitch range (Hincks, 2004, 2005). In fact, in public speaking classes students are urged to employ a varied intonation together with salient gestures in order to catch the attention of their audience. Even if these instructions are commonly given in many public speaking classes, there is a lack of scientific evidence that proves the actual effectiveness of these techniques. One of the aims of this work is to provide this scientific evidence. Another important instruction that is given in public speaking classes is to repeat and rehearse before delivering a speech, so that speakers can gain confidence and memorize their speeches, and this allows them to reduce hesitations and to improve the fluency of their speech. But does repetition also have an effect on speakers' pitch and gestures? Another aim of this work is to understand if repetition can elicit a different use of gestures and pitch.

Another important aspect that is investigated in this work is the role that communicative intent has in L2 gestures, speech and prosody. In fact, if a lively and communicative speech is characterized by an increase in pitch variation in L1 (Hincks, 2004, 2005), it is possible that the same effect will be shown in the L2. The communicative intent might also affect the use of gestural categories and this work will also investigate whether an increased liveliness in pitch might be accompanied by an increased liveliness in gestures, and, if so, how this relates gesture use.

Communicating in a second language is a complex cognitive process that includes word retrieval and speech structure organization. This cognitive load can cause a radical decrease in speech fluency and prosodic variation (see Jenkins, 2002; Zimmerer et al., 2014). In the present work, the role of cognitive load decrease will be investigated both in speech and gestures. In fact, since gestures seem to be involved in the lexical and rhythmical organization of speech (Butterworth & Beattie, 1978; Esteve-Gilbert et al., 2014; Krauss et al, 1996), their use could also be affected by the level of cognitive load speakers are facing in their utterances and this might also cause a different use of gestures in the L2 when the task is facilitated (through repetition).

The present work aims to test a number of hypotheses relating to the relationship between prosodic features of speech (pauses, disfluencies, speech rate, pitch variation) and gestures in situations in which communicative and cognitive efforts change their loads. In particular, the main hypothesis is that an increased communicative effort of the speakers will cause intonation to be more varied and gestures to be used differently. When the cognitive effort is decreased with task repetition, the hypotheses is that speakers will perform the task with a better fluency, increased pitch variation and different use of gesture categories.

The present work is divided into two parts: the first is an overview of the literature about gestures and speech; the second presents the two experiments carried out to test the experimental hypotheses.

In particular, chapter 1 aims to give an introduction on gesture studies, starting from gesture recognition and the discrimination of what can and what cannot be considered gesture. The chapter is also intended to explain the anatomy and the structure of gestures: what body parts can be used to produce a gesture and how speakers use space to perform gestures. This part mainly follows the definitions by Kendon (2004) and McNeill (1992), two of the main contributors in the gesture studies field. Chapter 1 also presents the main classifications of gestures based on the importance given to features such as type of movement or semantic and pragmatic functions.

Chapter 2 presents an overview of the studies on gestures in relation with speech. First of all, the five linkages between gestures and speech (McNeill, 1992) are presented. McNeill, in fact, explains five reasons why gesture and speech should be considered within a unified conceptual framework, and these represent one of the bases of modern studies of gestures and speech. Gestures convey pragmatic meaning but are also markers of speech rhythm. These functions are explained by Kendon (2004). Chapter 2 also aims to give an overview of gestures in L1

acquisition, and their facilitating function in L2 acquisition and communication. Finally, the chapter illustrates the importance of gestures in narration and of narrative levels that are linked to certain gestures and certain functions of gestures.

Chapter 3 is dedicated to the debate on the communicative versus the cognitive approach of gesture functions. This debate is illustrated in two sections. First, the chapter explains the evolution of the idea that gestures' main function is lexical facilitation (see Beattie & Coughlan, 1999; Butterworth & Hadar, 1997; Rauscher et al., 1996; Kita, 2000, 2010; Krauss et al., 1996). The second part of the chapter is devoted to the literature that claims that the main function of gestures is communication (Kendon, 2004; McNeill, 1992, 2005). The studies about the communicative role of gestures were carried on the observation that listeners' comprehension seems to be greater when they can see the speakers' gestures (Berger & Popelka, 1971; Graham & Argyle, 1975; Rogers, 1978; Riseborough, 1981) but also on speakers' observation in natural situations (Kendon, 1985) and on the analysis of conversations in natural settings (Goodwin, 1981; Goodwin & Goodwin, 1986, 1992; Moerman, 1990; de Fornel 1992; Streeck & Hartge, 1992; Streeck, 1993). The last part of the chapter is devoted in particular to the role that representational gestures have in communication (Bavelas et al. 2002, 2008; Cohen & Harrison, 1973; Melinger & Levelt, 2004).

Chapter 4 presents an introduction and a summary of what is prosody and what are its relations with gestures. In this chapter, the reader will be able to understand what is pitch and how this can vary cross-linguistically (Mennen et al. 2007, 2012; Patterson, 2000). The measure of pitch range used by Hincks (2004, 2005) represents one of the bases of the analysis run in the experimental part of this work.

Chapter 5 and Chapter 6 present the experimental investigations.

In Chapter 5, the experiment focuses on the effect that the communicative instruction given by the experimenter can have on pitch and gestures in the L2. The aim is to investigate global pitch range and use of gestures as produced by Italian speakers of English L2 before and after they are asked to be communicative in their speech. The main hypothesis is that the explicit instruction to be communicative may cause speakers to use a different pitch range together with a different way of gesturing, showing a more varied intonation and a different use of gestures. In particular, the aim of this experiment is to verify the possible relations between prosodic features and gestures in L2. The results will show that the instruction to be more communicative will bring

out a change in the way subjects use representationality in gestures and that these changes may be also linked to an increased prosodic variation and fluency.

In chapter 6, the experiment focuses on the effect of repetition on gestures and speech in a story retelling task in L1 and L2. The aim was to test whether by making the speakers' task easier with the repetition, speakers used a more varied pitch range and increased gesturing as it happened when they were asked to be more communicative (in chapter 5). To do so, subjects were asked to watch a cartoon, tell the story to a small audience and repeat the experiment one week later, this time with no special instruction being given before the task. The results will show that repetition has an impact in fluency and in L2 prosody, while the way speakers gesture does not change significantly in the second repetition, either in L1 or in L2.

Chapter 7 presents the integration of the results of the two investigations and a general discussion of the research, which is aimed to contribute to the study of the interrelations between communicativeness and gestures, supporting the idea that when speakers have the intention to deliver a communicative speech, gesture representationality will increase relevantly. The results also show that repetition helps elicit a better fluency in L2, although it seems to have no impact on speakers' gesturing.

Part I | Theoretical Background

1/ Gestures

1. Defining Gestures

Humans interact with each other through a complex communication system that involves both the use of verbal codes such as speech, whether or not the language employed is understood, and nonverbal expressions, or paralinguistic features, that can add and specify lots of contents and pragmatic meanings. Besides the whole core of vocal production, human communication also implements facial expressions, proxemics, body motions and spontaneous hand movements called gestures (Birdwhistell, 1955; Ekman & Friesen, 1969; Hall et al., 1968). Traditionally, though, research on communication has focused on verbal features and speech production. Lately, however, the interest in communication as an ensemble of various aspects of human cognition, and the idea that language and action are two aspects of the same communicative process has grown.

One of the body actions that often occurs when a speaker is talking is gesture. The Oxford English Dictionary defines 'gesture' as "a movement of the body or of any part of it that is expressive of thought or feeling" (Kendon, 1997). The meaning of the word, though, evolved to the current notation that refers more to a person's manner of action and expressive meaning than to its physical appearance. However, it is always difficult to specify what a gesture is and what it is not: for example, it is sometimes uncertain to understand whether a postural adjustment, an object manipulation, or even a hair pat are gestures or not. As an object of his study, Kendon (2000) defines gestures as "that range of visible body actions that are, more or less, generally regarded as a part of a person's willing expression". According to Kendon (1997), gestures are intended as actions that are treated as communicative by co-participants in interaction, including conventional gestures, gesticulations and signing, but excluding postural shifts, self-touching and incidental objects manipulations. According to Kendon (2004), 'gesture' is a name for a visible action when it is used as an utterance or as a part of it. The term 'utterance' refers to any combination of actions that are viewed by others as an attempt by the speaker to convey some kind of information. This concept arises from the formulation made by Goffman (1963, 1981; reviewed also by Craib, 1978) in which he states that whenever people are co-present to each other, they cannot avoid providing

information to one another about their purposes and involvements, about their social status and about their own individual personality. So, the notion of 'utterance' used by Kendon refers to a unity that includes any action or group of actions made by the interlocutors as a communicative act. These units may be formed by speech or by visible body actions or by combinations of these two features (Goffman, 1981; Kendon, 2004).

Whenever we observe someone speaking, we will notice what appears to be a set of hands and arms movements together with their speech. Nonetheless, even if it is a feature that everyone uses in almost all circumstances, gesturing is a phenomenon that often passes unnoticed at the conscious level (McNeill, 2000). Gestures, in fact, are able to express meanings and images that cannot always be expressed through speech, and so, combining gestures and speech allows the speaker to connote the meaning of what they intend to say.

Not all body actions, though, can be defined as gestures. Usually, in fact, those body expressions that can be considered as innate or physiological are excluded from the gesture category. So, if actions like waving goodbye, pointing or pantomimes are included in the definition of gesture, laughter, smiling or crying cannot be considered gestures unless they are used on purpose as, for example, to imitate those genuine expressions of emotion (Ekman, 1993; Ekman & Friesen, 1974). In an interactional context, moreover, there are many other actions that can reveal the personal attitudes or feelings of the speakers towards each other. These are not usually referred to as gestures just because they are practical to the interaction itself. Thus, even if interpersonal distance and posture can establish the attitude or nature of a relationship between the interlocutors, movements and actions that take place to determine which is the right personal space and/or the most comfortable posture to interact with each other are not considered gestures. Interpersonal distance and posture, in fact, usually pass unnoticed unless they violate the common expectations about space and special orientation within the specific culture in which they are taking place (Mehrabian, 1972; Hall, 1968; Kendon, 1973).

2. *Recognizing Gestural Actions*

Action movements seem to be easily categorized even by the non-expert observer. In order to understand how body movements are perceived, in a study by Kendon (1978), a mute video of a man giving a speech was shown to 20 people who were asked to describe the movements made by the man. Without any exception, every subject first said that they saw deliberate, conscious

movements that were part of the man's speech and only later described the presence of some other movements, described as 'natural' that did not carry any meaning. Basing his analysis on what the subjects reported, Kendon classified these movements into four types:

Deliberate expressive movements, which are intentional movements in order to consciously communicate something.

Position movements, that are all those actions made by the speakers to change their body orientation. These kinds of movements might involve the whole body or even only head and eyes, whenever they intend to change the direction of attention.

Object manipulations, meaning all those movements that involve an object, such as changing the position of an object or clothing manipulation, and that are not intended to be communicative.

Incidental or nervous movements, that, during the experiment ran by Kendon, were always mentioned last, only after the experimenter had asked them to recall every movement they had seen. These kinds of actions were never considered to bring any specific meaning and were never seen as intentional.

Even if the subjects recalled movements that are distributed in these four categories, according to Kendon (2004), only the actions that are included in the first category can be considered proper gestures. In fact, he intends 'gesture' as a label for actions showing deliberate expressiveness and that are perceived as controlled voluntarily by the speaker. How these actions may be considered as deliberate expressive or not varies from one situation to another: what can be interpreted as an expressive action (or gesture) may be considered an incidental movement if it occurs in a different context or even if, for example, the recipient does not share or does not recognize the meaning of the actions made by the speaker.

However, it is possible to claim that there are different cores of visible body actions that are usually referred to as gestures. During the last decades, various classifications of all these different cores of gestures have been attempted.

3. Anatomy and Structure of Gestures

3.1 Gesture structure

In his analyses on the relationships between body motion and speech, Kendon (1972, 1980) established a hierarchy of gesture movements in order to provide a diagram of the relations between the units of discourse.

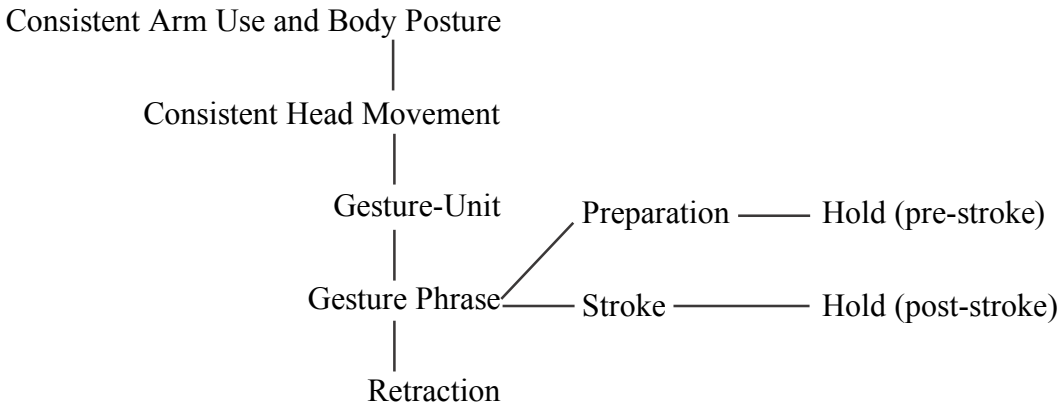


Fig. 1. Kendon's hierarchy

In Kendon's hierarchy above (Fig.1, taken and adapted from the summarization found in McNeill, Pedelty & Levy, 1990; McNeill, 1992) the first level is *Arm use and body posture* in which the speaker adopts different body postures and arm usage patterns. Within stretches of an arm and body movement, shorter actions occur in which *head movements* take place. So, when speakers are engaged in gestural actions, the body parts that they use to produce these actions undertake movement excursions or successions of such excursion starting from a rest position (also called Home position by Sacks & Schegloff, 2002) and, right after the excursion is completed, these body parts move back to their original rest position. The whole excursion, from the moment the articulators begin to move up to the moment they eventually return to the rest position is called *gesture unit* (also called G-Unit in McNeill, 1992).

Within the time of a G-Unit, it is possible to distinguish different *gesture phrases* and each of these finally consists of different *phases:*

1. *Preparation:* This is an optional phase in which the body parts move away from their rest position to a position in gesture space where the stroke begins. Usually the preparation phase tends to anticipate the co-occurring linguistic expression.

    - *Pre-stroke hold:* is the position and hand posture reached by the hand at the end of the preparation phase which can be held shortly just before the stroke begins. This phase is optional and appears only when, for some reason, the stroke onset is delayed (Kita, 1990).

2. *Stroke:* is "the peak of effort in the gesture" (McNeill, 1992) that is obligatory in the gesture hierarchy. In this phase the meaning of a gesture is expressed. Usually, the stroke is temporally synchronized with the co-expressive speech.

    - *Post-stroke hold:* is the position that comes at the end of the stroke. Occasionally this position can be held briefly before the beginning of the retraction phase. Usually, post-stroke holds appear in those cases in which the co-expressive linguistic segments are delayed (Kita, 1990).

3. *Retraction:* occurs when the arm/hand returns to a rest position, which can be different from the rest position occupied before the G-phrase.

While a G-phrase cannot exist without a stroke, all the other phases are optional. While almost all gestures contain a preparation component, retraction phases do not appear often because they are omitted when a gesture ends with the beginning of the following gesture. Furthermore, McNeill affirms that the preparation and the two-stroke phases can compensate for mismatches of speech-gesture synchrony.

In order to illustrate a typical G-Unit, Kendon (2004) described the following example (fig.2) in which is possible to recognize different gesture phrases inside the same G-unit.

(B) (C) (D) (E)

When you knock something down / you don't throw it away/ you use it again/ because it will last forever

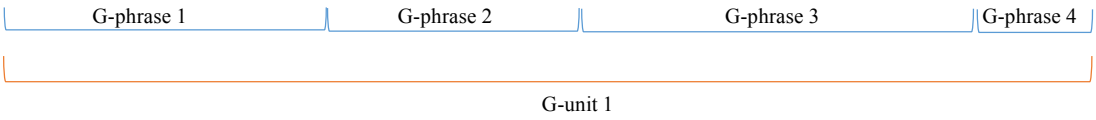


Fig. 2. Example taken and adapted from Kendon 2004, pp.120-121

Only the *stroke* phase is obligatory while all the other phases don't necessarily appear. For example, it is possible to see that the speaker, while ending G-phrase 3 and passing to G-phrase 4 did not engage in any preparation phase before performing the subsequent gestural movement, but instead passed directly from a *hold* position to the next *stroke* phase. Kendon (2004) analyzed also the phrasal organization of the gesturing in relation to speech, noticing that the stroke is often produced in close temporal proximity with the linguistic expression that has a similar semantic function. By coordinating the nucleus of the gesture phrase (which is the stroke and/or any post-stroke hold) with the semantic nucleus of the spoken expression, the speaker achieves the conjunction of two different expressive modalities with a coherent meaning (Kendon, 2004). In fact, when speaking, the speech is organized into a series of packages, identified as tone units, marked in terms of variations in voice, pitch level, loudness and pacing. The tone units tend to correspond to units of meaning, called by Kendon (1980) 'idea units'. Gestures seem to be organized in packages as well as speech, the gesture phrases. However, gestural expression usually corresponds with just one part of the entire idea unit expressed in the tone unit. Thus,

gestures may bring out an aspect of meaning associated with a verb in the phrase; or they may add an imagistic dimension related to a noun.

3.2 Gesture space

When gesturing, the speaker performs gestures in a specific space (fig. 3). McNeill (1992) affirms that this space can be "visualized as a shallow disk in front of the speaker, the bottom half flattened when the speaker is seated".

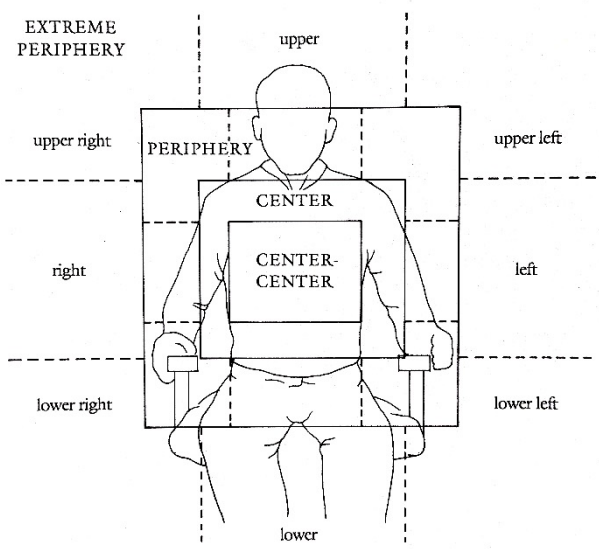


Fig. 3. Gesture space, taken from McNeill, 1992; p. 89

In order to better analyze gestural utterances, McNeill (1992) divided the space in front of the speaker as an imaginary concentric space, with four basic sectors: "center center", "center", "periphery" and "extreme periphery" which are further differentiated according to the features "upper", "lower", and "right" and "left". For example, a speaker that, while talking, performs a movement above his left shoulder, performs a gesture that would be described as produced in a "peripheral upper left" position.

Gesture space seems to be important for cohesion. In fact, gestures contribute to cohesion by repeating the same form or the same location in order to indicate discourse continuity (McNeill, 1992; Perniss & Özyürek, 2015). In narratives, for example, when introducing a new subject,

adult speakers usually assign it to a specific area by some gesture, either deictic or iconic (see §4.2.3) and, when the same subject is mentioned later, the speaker generally indicates the same location. These kinds of gestures seem to have a very decisive role in the avoidance of ambiguous interpretations of what is being said (Goodrich & Hudson Kam, 2012).

4. Classifying Gestures

Even if the importance of gestures for human expression has been recognized since the Greek and Roman era, when it was developed mainly in relation to the art of rhetoric and scenic gestures[1], and there are many descriptions of gestures even before the XX century (for example De Jorio in 1832 or even Bonifacio, in 1616), it is only in this century that the attention on these aspects of communication has grown. Thus, with the rise of the scientific interest in gestures, various attempts to categorize the gestural phenomena for scientific purposes have been made.

4.1 A semiotic categorization, Wilhelm Wundt

The first important categorization is the one made by Wilhelm Wundt (1921, edited in 1973). His classification is mainly semiotic and does not analyze gestures in relation to speech, but it refers to how the different forms of gestural actions are related to their meanings. He first divided gestures into two main categories: *demonstrative* and *descriptive.* In the first one, *demonstrative*, he included all those gestures that are pointing to something. For example, he refers to gestures that help to draw attention to present objects, to indicate spatial relations or to refer to body parts. The second category, *descriptive*, it is divided into more subcategories:

- *Mimic gestures:* which are those that directly imitate some action or object.
- *Connotative gestures:* in which some aspect of something is taken as standing as the whole.
- *Symbolic gestures:* which are those gestures that have a more complex relation with their meanings. Even if the shape of the gesture may refer to the object it is portraying or to the

[1] Before the XX century, gestures were analyzed as part of the oratorical technique and scenic art. The first discussion on gestures reported by Kendon (2004) is the one by Quintilian titled *Institutio Oratoria* ("Education of the Orator") written in the first century AD. In the XI book of the *Institutio* Quintilian divides *actio* or *pronunciatio* (Action or Delivery) into two components: *vox* (voice) and *motum/gestus* (movement and gesture).

action that can be made with it, sometimes it may also be standing for some more abstract concept. For example, a gesture that shapes the hand as a glass, not only delivers the image of glass but might also convey the general meaning of drinking.

4.2 Categorizations of gestures and their relation with speech

4.2.1 David Efron

Even though he never presented a real systematic categorization of gestures, one of the first detailed analysis of gestures and their relationship with speech was offered by David Efron in 1941. With his considerations (summarized and categorized by Paul Ekman[2]), Efron wanted to prove that the Nazi theory about a racial origin of gesture was wrong. In order to do so, Efron found, when comparing two different groups of immigrants in the U.S., Italians and Jews, that they used completely different repertoires of movements during conversation. Since he believed that these differences were due to the different environments in which they grew up while in Europe, he demonstrated that the amount of differences between the way of gesturing of the two groups of immigrants decreased in the first generation offspring: those who were assimilated into the American culture failed to maintain the gestural style belonging to their original cultures. To be able to analyze the differences between the two groups, Efron elaborated very detailed distinctions, focusing mainly upon hand movements, and to a reduced extent on head movements, with only occasional considerations of trunk positions. He did not consider facial expressions, posture, stride, or eye movements.

In his analysis, Efron distinguished and measured three aspects of hand-head movements:

1. *Spatio-temporal,* in which gestures are considered simply as movements, independently from the reason why they are produced.
   The features that are included in this distinction are:
   - *radius*, meaning the size of radius and axis of movement (for example, wrist, elbow etc.);
   - *form,* which can be distinguished between sinuous, elliptical, angular or straight;

[2] The classification was actually organized by Paul Ekman in his Introduction to the 1972 edition of the book "Gesture, Race and Culture" by Efron & van Veen.

- *plane,* that can be sideways-transversal, towards auditor-frontal, up-down-vertical, away from speaker and auditor-dorsoventral centrifugal;
- *body parts,* involved in gesticulation and ways in which they are employed: head gestures, whenever they are used as a substitute for hands, digital gestures, meaning the variety of positions and shapes of hands, ambulatory gestures, in which the motion is transferred from one hand to the other;
- *tempo,* which can be abrupt, dischronic or described as a flowing transition from one movement to another.

2. *Inter-locutional,* meaning all those aspects that consider the communicative or interactive elements of gestures, separately from the referential aspects:
   - *Familiarity with the physical person of the interlocutor* (touching the interlocutors to interrupt them or to capture their attention)
   - *Simultaneous gesturing of all the interactants*
   - *Conversational grouping* (use of space and distance between speakers and recipients)
   - *Gesturing with objects*

3. *Linguistic,* or the referential meanings of gestures. Among the linguistic aspects, Efron (1941) differentiated two wide classes of gestures based on whether the meaning of the gesture is independent or is conveyable only together with speech:
   - *Logical-discursive.* Gestures that refer to the course of the ideational process and do not refer to any object or thought so that they relate to how they enact the concept more than what they represent. These kinds of gestures are used to give emphasis to the content of the verbal behavior, so that they are physical re-enactments of the speech process, of logical pauses, inflections etc.

     The logical-discursive gestures are further subdivided into two categories:
     a) *Baton,* that are used to beat the tempo of the mental flow of discourse;
     b) *Ideographic,* that are used to sketch out in the air the path or motion of thought.
   - *Objective,* those gestures that have an independent meaning from speech and that may be used by the speaker while talking. According to Efron, these kinds of gestures can belong to three different subcategories, which are:

a) *Deictic:* when pointing a visually present object or subject;

b) *Physiographic:* gestures that physically show their meaning. In this subcategory it is possible to find both gestures that depict the form of the visual object, called *iconographics,* and gestures that depict a body action or movement, called *kinetographics.*

c) *Emblematic* or *symbolic:* those gestures whose function is to represent either a visual or a logical object by a form that has no morphological relations with the entity it represents since their meaning is standardized within a specific culture.

The analysis by David Efron can be considered pioneering work. First of all, it is one of the first attempts to offer an insightful and detailed examination of gestures in relation with speech because it recognized that, in order to get a comprehensive understanding of gestural communication, it is necessary to integrate analyses from different perspectives. Furthermore, his analysis of how gesturing can be related to speech from a linguistic point of view is one of the most articulated to date and it is one of the milestones on which the later studies are based.

4.2.2 Paul Ekman and Wallace Friesen: the repertoire of nonverbal behavior

As said before, the analysis made by Efron represents the basis for many studies that followed from the second half of the XX century. One of the most important among these is the study by Paul Ekman and Wallace Friesen who used and somehow expanded Efron's analysis offering a structured category system that allows distinguishing different kinds of nonverbal communicative performances. In their work, entitled 'The Repertoire of Nonverbal Behavior: Categories, Usage, and Coding' (1969), Ekman & Friesen define *nonverbal behavior* as 'any movement or position of the face and/or the body' that is relevant for communication and interaction. Unlike Efron, Ekman and Friesen didn't focus only on the hand and arm movements, but extended their interest also on facial expressions. Their distinctions follow three main considerations for understanding a person's nonverbal behavior: *origin, usage* and *coding*.

The first one, *origin,* refers to whether the action is innate or learned. For example, facial expressions of emotions can be considered heavily shaped by inborn tendencies (1969). The

authors differentiate at least three types of origins that can occur whenever a nonverbal behavior is enacted:

a) The simple relationship between stimulus events that cause a nonverbal activity which is proper of every intact nervous system of any human being. The most obvious example is the one represented by a reflex or any kind of spontaneous reaction, and it seems that facial expressions of emotions are based upon inherited neurological connections common to everyone.

b) The nonverbal action that is common to all the people that is not strictly inherited and inborn but is rather acquired as part of the human experience while interacting with the surrounding environment. For example, no matter the culture, every human being uses hands to bring food to their mouths.

c) Nonverbal behaviors that come from an experience that, as the authors say, 'varies with culture, class, family or individual'. In this third kind of origins, Ekman and Friesen include those behaviors that are learned:

- as part of an instrumental task, in order to master a particular activity such as swimming or learning how to do something in a certain way;
- as part of a social interaction, with a more or less explicit and conscious attention from learner and/or tutor, in order to establish or keep a specific type of social interaction. In fact, imitating postures or facial expressions can be done intentionally if they are made consciously from a person learning to be an actor but they may be occurring with less awareness when a child is imitating the movements of the same-sex parent.

Other kinds of considerations that, according to Ekman and Friesen (1969), are necessary to understand the human nonverbal behavior are those about *usage.* With the term *usage*, the authors refer to regular conditions surrounding the occurrence of a nonverbal action. The issues included in these considerations are various:

a) The relationship of the act with the associated verbal behavior: a nonverbal act can anticipate, match, follow the verbal behavior, or even substitute it; a nonverbal act can illustrate or underline what is being said by the speaker, but it can also contradict the co-occurring words if, for instance, the speaker wants to be ironic.

b) The nature of the external conditions found when the act occurs: this refers to any environmental circumstances that can inhibit or facilitate an act. For example, the set of an

act can be considered an external condition. In fact, some acts are more frequent in a setting than in another one, for example, the same two persons can act as father and son in a home setting and as employer and employee in a work setting.

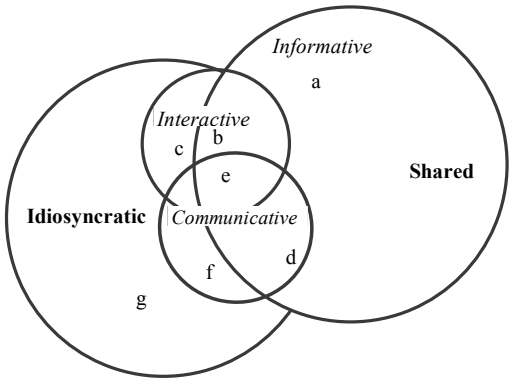
c) Whether the person is acting consciously or not: this refers to the awareness of a person delivering a specific nonverbal message. People, in fact, can be more or less conscious of delivering a communicative message through their nonverbal enactments.

d) The person's intention to communicate or the deliberate use of a particular nonverbal behavior to intentionally communicate something rather than something else.

e) The feedback from the observer of the act: this means the information concerning a nonverbal act that the receiver provides to the sender.

f) The type of information conveyed by the act: this distinguishes between *idiosyncratic*, referring to that information that has a meaning or interpretation only for a single person, and *shared* information, where groups of people agree on how to interpret an act.

The authors then distinguish between three types of interpretations that a nonverbal act can have: *informative,* when the act can be interpreted in a shared way by others but the information delivered is independent of whether the sender intended to convey any information or not; *communicative,* when the non-acoustic parts of the acts are clearly and consciously intended to transmit a message



(a) Much informative behavior is neither interactive nor communicative
(b) Though, interactive behavior may also be informative
(c) But it can also be idiosyncratic (these are acts that influence a specific person, i.e. the acts of a son in regard of his mother, but that have no influence on the other interactants)
(d) Some communicative behavior is only informative
(e) Some other are both interactive and informative
(f) There could be some communicative behavior which is neither interactive nor informative (rare cases in which a mental disorder may affect the communication to the receivers which cannot interpret or interact with the sender)
(g) Much idiosyncratic behavior is neither interactive nor communicative.

Fig. 4 Relationships among types of 'usage' according to Ekman & Friesen, 1969

to the receiver; and *interactive,* whenever a nonverbal behavior clearly modifies or influences the interactive behavior or the receiver. The relationships between these nonverbal behavior modalities are illustrated by Ekman and Friesen in fig 4.

The last of Ekman and Friesen's three aspects about human nonverbal behavior is *coding*, or the principle of correspondence among the act and its meaning. The rule that characterizes the relationship between a certain action and its significance is the code, and this may be intrinsic, when the act stands for its own significant without any kind of semantic separation, or extrinsic, when the act signifies for something else in a way that can be arbitrary or iconic: acts that are arbitrarily coded have no resemblance to their meaning while iconically-coded acts carry similarities to what they are meant to signify. The difference between iconic and intrinsic codes is that the first only resembles its significant while the second is, at least in a part, the significant itself (a kiss, for example, does not only look like a kiss, it is the meaning and the significant of that act at the same time). Referring to both intrinsic and iconic acts, Ekman and Friesen affirm that the ways in which a nonverbal act is related to its significant may be *pictorial,* whenever a movement shows its meaning by drawing a picture of what the person is communicating; *spatial,* that is the relationship in which the movement depicts distance between people, objects or ideas; *kinetic,* in which the movement resembles an action or a part of it; *pointing*, in which some part of the body points to an object, a person or even to an abstract idea; *rhythmic,* in which the movement traces the flow of ideas, or accents a specific sentence. Usually, the rhythmic relationship between act and meaning does not carry any message. Besides pointing, that is always intrinsically coded, these relationships are always iconic, while kinetic relationships may be either iconic or intrinsic, whenever the movement is part of the action it signifies.

Following the considerations about origin, usage, and coding, Ekman and Friesen identified five different categories of nonverbal behavior: *emblems, illustrators, manipulators* (or *adapters*)*, regulators,* and *emotional expressions* (or *affect display*).

1. Emblems. When defining emblems, the authors referred to those nonverbal actions that can be directly translated into words and usually consist of one or two words, sometimes a phrase, but do not change the information conveyed. An emblem may repeat, replace, contradict the verbal behavior, but also provide a separate comment related to the co-occurring speech. According to Ekman and Friesen (1969 1972) and Ekman (2004) emblems are the "only true body language", in the sense that these kinds of movements have a set of specific meanings which are understood by all members of the same culture, subculture or social groups. Since they are socially learned,

emblems are as culturally variable as language so, as it happens with words, the same message may be represented by one emblem in a culture but by a different one in another culture.

People are usually conscious of the fact they are using an emblem, they are aware of when they are executing these kinds of nonverbal acts and can repeat them if asked to. Emblems may be both iconically coded, in the case in which there is a clear resemblance between the act and its meaning, or arbitrarily coded, if there is no similarity at all. As demonstrated by Efron (1941), these types of gestures originate through learning in a specific cultural setting, they do not occur spontaneously and, possibly, have a much higher proportion of shared meaning than idiosyncratic meaning. However, Ekman & Friesen point out that emblems, in addition to their shared meaning, could also have an idiosyncratic level of meaning. Furthermore, emblems are used as communicative acts, although sometimes they can be produced unconsciously. Since they can influence the perceiver's attention or behavior, emblems can often be considered also interactive –in addition to being communicative.

2. Illustrators. The authors include in this category all those movements that illustrate speech, they serve to somehow clarify what is being said and they are closely related to what the speaker is saying in a "moment-to-moment basis". Ekman and Friesen grouped together under this category the five classes of gestures recognized by Efron (1941) and added two more categories as showed in Tab. 1.

Based on Efron's analyses	*Batons*	Movements which time out, accent or emphasize a particular word or phrase, that beat the tempo of mental flow
	Ideographs	Movements used to sketch a path or direction of thought, that trace the flow of ideas
	Deictic	Used to point to a present object or to a person
	Kinetographs	Movements that depict a body action
	Spatial	Acts that depict a spatial relationship
Added by Ekman and Friesen	*Pictographs*	Used to draw a picture of their referent
	Rythmic movements	Depict the rhythm or the pace of an event

Tab. 1 The seven types of illustrators (Ekman and Friesen, 1969; Ekman, 2004)

The first two subcategories of illustrators, batons and ideographs, do not have any independent connotation if analyzed separately from the concomitant speech (Efron, 1941; Ekman and Friesen, 1969) while the other types of illustrators do carry an independent meaning, conveying at least something of the speech content if watched without hearing what is being said. If emblems are almost always intentional and produced with a high level of consciousness, illustrators can be used with less awareness and intentionality, anyway it is not rare to use them consciously (or to learn how to use them in order to get an effective professional speech for example). Most illustrators are informative, since they provide a shared meaning, but some illustrators could be also considered to be communicative when they are used with the intention to convey a message to the receiver, and in some contexts they may also be interactive. All kinds of illustrators are usually coded iconically (when there is resemblance among the acts and their meaning) but in some cases they could be intrinsically coded too, and therefore stand for their own significant, as it happens for the deictic illustrators, for example.

The body parts that are involved when people perform an illustrator gesture are typically the hands, even though sometimes head and/or feet may be used as well. In fact, it is possible to use for example eyebrow movements to accent or emphasize a word and this is usually connected to a change in loudness and pitch (Graf et al. 2002).

3. Manipulators (or *adaptors*). In these movements one part of the body or face manipulates another part of the body, face or an object in some way (some of these manipulations are, for example, pressing, stroking, scratching, biting, etc.). While some of these manipulators emerge to accomplish cleaning, grooming or other kinds of personal care, many others seem to appear without any instrumental goal and seem to simply reflect nervousness or the need of getting more comfortable. Manipulators seem to be at the edge of awareness and people often notice they are producing a manipulator only if someone else asks them what they are doing (Ekman, 2004).

4. Regulators. This category includes those acts that "maintain and regulate the back-and forth nature of speaking and listening between two or more interactants". Regulators tell speakers and listeners when it is their turn to talk, to repeat, hurry up, pay special attention etc. These gestures are related to conversation and are connected to the conversational flow and pacing and do not carry any message content. Regulators seem to be "on the periphery of awareness" since people execute regulative acts unintentionally but, if asked, they can remember them and do them again. On the other side, the observers seem to be sensitive to regulators when they are not present but, at the same time, they do not seem aware of their presence when they see them. According to the

authors, the most common regulator is the head nod, but regulators can also take the form of slight forward or backwards movements, eyebrow raises and many other small nonverbal behaviors.

5. Emotional expressions (or *affect displays*). Which mainly involve facial expressions of emotion. The authors (Ekman and Friesen, 1974; Ekman, 1993) listed seven pan-cultural primary affects: happiness, surprise, fear, sadness, anger, disgust and interest. Each of these affective conditions can be recognized by observers of any culture. So, according to the authors, even if these affective states are pan-cultural, display rules are socially learned probably in the early years of a person's development and they are individually and culturally variable. Hence, Ekman and Friesen affirm that there are well-established social norms about which display rule is the right one according to affect, the characteristics of the displayer (sex, age, role, etc.) and the context in which they have to be expressed. So these rules "specify who can show which emotion to whom and when" (Ekman, 1993, 2004). Emotional expressions can be related to verbal behavior in many ways: they can repeat, qualify or contradict a verbally stated emotion and they can be used as emblems if, for example, a particular social group coded a specific expression so that it is recognized as emblems (that happens with the smile, coded as "happiness" in most cultures). Most affect displays are informative and often have interactive consequences since they usually provoke a reaction of the other interactant. Since much of these kinds of nonverbal behaviors is produced with no intention or awareness, they are usually not communicative. The meaning of emotional expressions is clearly shared, but they also have an important level of idiosyncratic meaning: past experiences, expectations and particular circumstances can provide idiosyncratic information among interactants.

4.2.3 David McNeill

One of the most important contributions to gestures studies in the XX century is the one provided by David McNeill (1992) who, starting from Kendon's contribution to gesture analyses, investigated nonverbal behavior in its correlation with speech, supported by the idea that speech and gestures share a very tight bond and that, therefore, they actually must be seen as inseparable components of the act of utterance.

The types of gestures that McNeill considers for the aim of his analyses are:

1. *Iconics.* According to McNeill (1992) this type of gesture includes those nonverbal acts that carry a close formal relationship to the semantic content of speech, they physically recall the thing they are meant to signify.

   For example, when describing the action of opening a bottle of water, the speaker might appear to hold something in the left hand while turning its imaginary cap in anticlockwise direction with the right hand above the left one. In this way, iconic gestures show that there is a very close connection with speech, which is depicted through hand and body motions and that only together speech and gesture can offer the complete picture of a person's thought. Thus, iconic gestures seem to be both co-expressive and complementary with speech, they are produced together and they also add another semantic level to what is being said.
2. *Metaphorics.* Like iconic gestures, metaphoric gestures are pictorials, but their pictorial content conveys an abstract idea and does not represent a concrete object or event. This kind of gestures, thus, does represent the idea, the concept the speaker is trying to deliver.

Fig. 5. Metaphoric gesture

For example, a metaphoric gesture is the one that is typically done while talking about something as abstract as a conceptor a feeling. For example, fig.5 shows a person talking about the famous cartoon 'Tweety and Silvester', the gesture that he produces is an instance of the metaphor: "the idea of the genre presented as a bounded container supported by the hands" (McNeill, 1992). Anyway, it seems that while this kind of gestures of abstract ideas as physical containers occurs very often in speakers coming from Western cultures, this image does not appear in speakers coming from non-western traditions (as Chinese).

3. *Beats.* Those gestures in which the hand moves along "with the rhythmical pulsation of speech". These kinds of gestures are those that Ekman and Friesen (1969) called *batons* and are those gestures that, unlike *iconics* or *metaphorics,* have the same shape regardless of the content (McNeill & Levy, 1980). A typical beat is performed by a simple movement of the hand or finger, up and down or back and forth, the movement is fast and short and the space in which it is produced is often the peripheral zone of the gesture space. The main distinctive aspect of this type of gestures is that while the other gestures are formed by three phases (*preparation*, *stroke* and *retraction*, discussed below), beat gestures just have two movement phases: up/down, in/out, back/forth, etc.

   Among all gestures, beats are the most "insignificant looking" but they can actually reveal very important aspects of the co-occurrent speech, like the conception of the narrative discourse as a whole and the importance or significance of a word or phrase in its pragmatic content.
4. *Deictics.* These gestures are typically implied to indicate objects and events in the concrete world, but they can also point at something completely abstract as it often

Fig. 6. Deictic gesture

   happens in narratives and conversations. In fact, when the deictic gestures point at some abstract concept, they employ a metaphorical picture in which abstract ideas have a physical locus. For example, fig.6 represents the gesturing of a person while saying "where did you come from before?": the speaker points to space between self and interlocutor; the pointed space, though, is not the space where the interlocutors find themselves but it is an abstract space housing an introduced reference. Deictics are prototypically performed with the index, although any extensible body part or object can be used: a hand, the head, the chin, as well as a manipulated object.

5. *Butterworths.* These are gestures that usually correspond to "speech failure" (as intended by Freedman & Hoffman, 1967; Freedman, 1977). McNeill named this category after Brian Butterworth who first argued the correspondence of these gestures with speech failures (Butterworth & Beattie, 1978; Hadar & Butterworth, 1989). While neither Efron nor Ekman and Friesen included such group in their categorization, McNeill thinks that they have to be considered since "they occur as part of an effort to recall a word and/or find and appropriate sentence structure" and meanwhile they also tell conversation partners that we are not yet done speaking. In narratives, though, these gestures do not appear that often. Typically, they might take the form of a hand waving or grasping while the speaker is trying to recall a word. Hence, while they appear there is usually no co-occurring speech.

Tab. 2[3] shows how the two previous categorizations can be compared to the categorization proposed by McNeill in 1992:

Efron (1941)	**Ekman & Friesen (1969)**	**McNeill (1992)**
Physiographics Kinetographics	Kinetographics Pictographs	Iconics
Ideographs	Ideographs Spatials	Metaphorics
Deictics	Deictics	Deictics
Batons	Batons Rythmics	Beats
-	-	Butterworths

Tab. 2. Gesture categorization comparison, Efron, Ekman & Friesen and McNeill

[3] Taken and adapted from McNeill, 1992; Table 3.1 p. 76.

4.2.4 Kendon's continuum

In order to set the stage for gesture studying, McNeill (1992) presented and ordered gestures that were first described by Adam Kendon (1988) and called them, in his honor, *Kendon's continuum*. In Kendon's continuum, McNeill described the gesticulation in its relationship with language.

In this way, Kendon's continuum is important to sort out gestures of fundamentally different kinds. In fact, according to McNeill, it is not possible to just talk about gestures in general but it is essential to distinguish among different categories. The main focus of McNeill's research is on *gesticulation* in which gestures are produced as "idiosyncratic spontaneous movements of the hands and arms accompanying speech, and that almost never occur in speech absence" (McNeill, 1992). Reading the continuum from right to left, the correspondence between gestures and speech gets closer. In '*speech-linked' gestures*, gestures are grammatically assimilated into the utterance; in the sentence "I don't think he's crazy I just think he's [gesture resembling someone drinking]", the gesture actually fills the grammatical slot of an adjective which, in this case would be "drunk". *Emblems* are conventionalized gestures in which meaning is shared within a specific culture or social group, they can appear even without speech since they carry they own meaning. Italians, for example use a wide variety of different emblems that cannot be understood by people that do not share the same culture. At the same time, the same emblem might have different connotations in different cultures: for example, the "ring" gesture, in which the tips of the index and thumb are joined in order to represent a circle while the rest of the fingers are kept straight, is widely recognized as "Okay" but it is actually considered obscene in countries like Brazil and Iran (in which it is roughly equivalent to the middle finger gesture). In *pantomimes*, instead, hands portray objects or actions in order to describe a story or situation without speech. Pantomime is used instead of language but it differs from emblems because it is not culturally coded, and can actually be used and understood even by interlocutors that do not share the same culture. At the other extreme of the continuum McNeill lists *sign language*, which is actually made of lexical words represented as signs. Sign languages, such as ASL or LIS, are actually considered verbal

languages, have their own linguistic structures and have evolved without the need of being coordinated with speech.

4.2.5 Speech interaction with gestures – Pragmatic meanings or rhythmic markers

Usually, the meaning of a gesture is strictly connected to its form, but there are cases in which gestures forms and meanings are disconnected from each other, for example in the case of gestures that are used to indicate something about the speaker's attitude to the referential meaning or to help the listener interpret the framework in which the speaker wants that utterance to be understood. To indicate these gesture functions (and in order to distinguish them from the *referential* ones) Kendon (2004) uses the term *pragmatic function* relating them to Efron's *logical-discursive* category. According to Kendon, there are four main kinds of pragmatic function:

1. *Modal function:* the gesture modifies the frame in which what is being said has to be interpreted. In other cases, the modal function of a gesture indicates whether speakers are talking about a hypothesis or a fact they believe is concrete.
2. *Performative function:* the gesture is used to indicate the kind of speech act or interactional move a person in engaging in. For example, if a person performs a gesture moving a palm up-open hand towards their interlocutor, this movement may be indicating that what is being said is being 'offered' or 'proposed' to the interlocutor.
3. *Parsing function:* the gesture is used as a punctuation of speech, or is marking its logical components. Beats, for example, are described by McNeill (1992) as simple rhythmic hand movements that can mark new information in contrast to given information in a specific speech context. This kind of gestures can also function as a cohesive action that allows the speaker to show logical connections between different parts of the whole discourse. Other gestures can convey the contrast between different temporal references (Kendon, 2004; Valbonesi et al. 2002): a simple gesture can indicate whether the speaker is referring to background information, to the exact moment in which the discourse is developing, or to a future moment. Another important role of gestures occurring in natural conversations studied by Kendon (1995) is that they can actually help to distinguish discourse marking functions such as differentiating a topic from a comment or help specify which is a fundamental part of the speech and which is not.

4. *Interactive function:* gestures can be used as a way to indicate the recipient of a specific utterance, who claims the role of speaker (even if they are actually not speaking in that exact moment) and when they are passing the turn to talk to another interlocutor, for example by pointing at them. The interactive functions of gestures have often been acknowledged (Kaulfers, 1931; Goodwin, 1981; Streek and Hartege, 1992) but their important role for what concerns different aspects of a conversational interchange management has been recognized and systematically discussed only in the last decades. Bavelas et al. (1992) not only recognized this function but actually distinguished a separate gesture category, which she called "interactive gestures", referring to those gestures that seem to cross-reference the matter of a speaker's utterance to the main argument of the conversation, to show the participant's understanding of what is being said by the speaker, and manage the turn distribution in a communication exchange (Bavelas et al 1992, 1995; Bavelas, 2008).

On the other hand, gestures can actually provide also a visible indication of different levels of discourse structure and have a helping role in the management of the interaction between speakers. In fact, besides accompanying a verbal utterance, gestures may actually be a part of its referential content. In the example shown in Fig. 3, in the utterance "When you knock something down/ you don't throw it away/…", the gestures "knock down" and "throw away" are defined as *referential* because they make actual reference to the actions they describe. Various attempts have been made to classify referential gestures (see, for example, Mandel, 1977; Kendon, 1980; Poggi & Magno Caldognetto, 1997; Calbris, 1990) and, even if they present differences from one another, they all share a division between three main kinds of strategies that are used to perform referential gestures:

1. *Modelling:* whenever a body part is used as if it was a model for some object. For example, a hand can be shaped so that it resembles the object or a relationship with the object the speaker is talking about.
2. *Enactment* (or *pantomime*): when the speaker uses a pattern of action that has features in common with some actual pattern of action that is being referring to. For example, a speaker that is talking about cutting a tree might move their hands as if they were actually holding an axe.

3. *Depiction:* when the hands, usually previously shaped in different ways according to what is being depicted by the speaker, are engaged in creating an object in the air, sculpting or drawing the shape of something.

The recognition of referential gestures requires an understanding of the contexts in which they are produced and of the relationship with them; these contexts are usually provided by the co-occurring speech. In fact, it would be impossible to be sure of the meaning of the axe enactment presented before, if the speaker did not clarify the context verbally.

2/ Gestures in relation with speech

1. Introduction

Gestures and language differ from each other in a number of fundamental dimensions. Yet they are closely linked. Kendon's gesture microanalysis, made through audiovisual recording technology of the relationship between speech and body movement, reveals that speech and gestures are produced together, and that they must be considered as two aspects of the same communication process where speakers combine both spoken and gestural expressions in a single action (Kendon 1972, 1980a; McClave, 1991; McNeill 1985, 1992, Nobe 1996).

Co-speech gestures seem to fulfill a number of functions, and may in fact be multifunctional (Holler & Wilkin 2009, 2011; Holler et al, 2011; Galati & Brennan, 2014) Gestures have been shown to facilitate speakers' cognitive processes during speech production; for example, they seem to help speakers conceptualize, retrieve lexical items, manage cognitive loads, organize information into syntactic constituents. Gestures also seem to be planned and produced with the addressee's needs in mind, and so play a role in communication. For example, speakers produce more and larger gestures when they see the recipient, than when they do not (e.g., when they are talking over the phone) (Mol et al. 2011). Speakers' gestures are also affected by common ground, that is the amount of knowledge that is shared between the participants in a spoken interaction. It has been shown that assuming common ground causes speakers to use less words in their narratives than when no common ground can be assumed (because in the first case speakers can rely on their interlocutors to understand implicit references); on the other hand, common ground produces an increase in the use and extent of gestures during speech, possibly to enhance communication with the interlocutors (Holler & Wilkin 2009; Holler et al, 2011; Horton & Keysar, 1996; Galati & Brennan, 2014; McKinnon, & Prieto, 2014).

Gestures and speech happen to be temporally synchronized in a very close way and often have overlapping meanings. Yet, they express these meanings in completely different systems (McNeill, 1992). The combination and coordination of gesture and speech allow both speaker and recipient to fully express and understand many details in a way that would not be possible to communicate

in any other way. A non-coordinated speech and gesture utterance, in fact, might bring to misunderstandings. For example, let's simply think of a speaker indicating downwards while talking about the ceiling, this would probably confuse the listener.

There are many links between speech and gestures that lead to the assumption that they must be considered within a unified conceptual framework as two aspects of the same concept. McNeill (1992) lists 5 of these linkages:

1) *Gestures occur during speech (or instead of it).* While emblems and pantomimes may be delivered independently from speech, iconics, metaphorics and beats are almost always accompanying it.

2) *Gestures and speech are semantically and pragmatically co-expressive.* Usually, gestures that accompany utterances also present semantically close related meanings and the same pragmatic function. For example, iconics usually appear together with verbal utterances that depict concrete objects or events while metaphorics are produced with utterances that refer to the pragmatic structure of the discourse as a whole.

3) *Gestures and speech are synchronous.* Very often, the verbal utterance is temporally aligned with the semantically corresponding gestural utterance. In fact, the stroke (the most significant phase of a gestural action) frequently appears as contemporaneous to its equivalent linguistic segment. Such synchrony implies that the speaker is delivering the same meaning through these two different channels, but at once, in a single performance. Kendon (1990) also affirms that it seems that gestures are often temporally aligned with speech: for example, in a telling of the story Little Red Riding Hood, a speaker used the gesture of a slicing hatchet right when pronouncing the word "slice" demonstrating that, at least in this case, gestural action and speech must have been organized together. Furthermore, in this particular example, gesture and speech actually co-operate together to provide a more specific meaning to what is being said: in fact, the "slicing action" can be done in a number of different manners, but the right one is being given by the speaker while gesturing the exact action they are referring to, delivering also meanings beyond those expressed linguistically (Kendon 1990).

4) *Gestures and speech develop together in children.* For what concerns gestures, children usually begin their gestural activity with the use of concrete pointing and some kind of iconics and, sometimes, emblems. They develop the use of the other kinds of gestures

much later: iconics, beats, metaphorics and last of all, abstract pointings. The interesting thing is that gestural development seems to follow the same path of speech development: as children develop their speech abilities from referential focus to descriptive elaboration and, eventually, to discourse structuring, they also develop their gesture abilities: from a concrete deictic, to descriptive iconics and, eventually to discourse referring gestures with metaphorics, beats and abstract pointings (Haviland, 1993; McNeill, 1993).

5) *Gestures and speech break down together in aphasia.* Broca and Wernicke aphasias are two of the main disturbs of speech, but they do not only affect speech production: they affect gestures too. The first one, Broca's aphasia, consists of a "relatively intact ability to use referring terms but a radically impaired ability to combine terms into larger grammatical units" (McNeill, 1992). Gestures of these kinds of aphasics are equivalent to what happens to their speech: even if they use a lot of iconic gestures, their gesturing lacks of metaphorics and beats. In Wernicke aphasia, speakers present fluent speech but a loss of semantic coherence and connection. As a result of such disturb, Wernicke's aphasics present gestures that are fluid and large but that cannot be interpretable since they lack of coherence between one another. Also, while Broca's aphasics present a large number of iconics but almost no metaphorics and beats, Wernicke's aphasics use a lot of metaphoric and beat gestures but do not use a lot of iconics, demonstrating that these syndromes affect speech and gesture production in parallel ways.

2. Speaking and Gesturing. When Words are Complemented by Gestures

It is known that words are often accompanied by gestures. Speakers, though, do not gesture every single time they speak. As it happens with the spoken language, the use of gestures varies in relation with the linguistic, interactional, and social context (Kendon 1997, 2004). What elicits gestures, though, is a matter that has received very little attention. To understand in which circumstances speakers use gestures, Cohen & Harrison (1973) compared situations in which interlocutors could or could not see each other and showed that speakers, when giving route directions, used fewer gestures when they couldn't see the person they were talking to. Some further studies, as the one by Rimé (1982), showed how, in comparison to when the speakers could see each other, when talking with a barrier between them, speakers only slightly reduced the amount of gestures they implied. Moreover, it has been demonstrated that when speakers are

not able to see each other while talking, they actually avoid using any 'interactive' gestures, meaning those that refer to some aspects of conversing with an interlocutor; although, they keep on using 'topic' gestures which are those that express the content of an utterance (Bavelas et al., 1992; 1994).

Gestures are employed by the speaker to help the listener understand the meaning and the possible connotations of what is being said. However, anecdotal evidence shows that people gesture even when they are engaged in a phone conversation, not being able to see their addressees. Studies have shown that when the speakers are not allowed to see their interlocutors and are forced to talk to someone on an intercom or through a partition, i.e., their visibility varied, their total amount of gestures does not drop to zero (Alibali et. al, 2001; Bavelas et al. 1992, 2008; Cohen & Harrison, 1973; Emmorey & Casey, 2002; Rimé, 1982, de Ruiter, 1998). All these experiments lead to the assumption that, besides being used to help the receiver understand the meaning that the speaker wants to deliver, gestures might also have a cognitive function for the speakers themselves. Furthermore, it seems that gestures, as well as intonation, are integral to the face-to-face signals compound of a dialogue, so that it would result difficult to produce speech without gesturing, and that, briefly, the presence of dialogue seems to have an effect on speakers' gesturing, independent of visibility. This might explain why speakers gesture significantly more when talking on the telephone rather than when they are engaged in a monologue (Clark, 1996; Bavelas, 2008). Since it has been demonstrated that gestures are present in a wide number of communicative contexts, when the interlocutor is invisible or even absent, gestures are thought to support speakers by providing aid to verbal formulations and lexical retrieval, as well as by helping them to keep complex concepts in mind (Butterworth & Beattie, 1978; Krauss et al, 1996; Goodwin & Goodwin, 1986; Morrel-Samuels & Krauss, 1992; Rauscher et al, 1996; Rimé & Schiaratura, 1991).

3. Gestures in L1 and L2 Acquisition

3.1 L1 acquisition

In the language development of children, gesture is often considered to have an important and precursor role. In fact, it is considered to be the first communicative tool that children use to be understood by others, a tool that is gradually replaced by speech. The main example of this is

represented by *deictic* gestures: young children initially point at objects and then, as they grow older, they learn the words referred to the objects they want to point at and finally do not need to use deictic gestures anymore (Butterworth, 2003; Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005; Lock, 1978; Lock et al. 1990; Tomasello et al., 2007; Pizzuto & Capobianco, 2005; Vygotsky, 1962). There are two main perspectives about how gestures are related to speech in first language acquisition. A first approach claims that language and speech are not connected (Chomsky, 1972) but instead are two independent and autonomous systems (Butterworth & Beattie, 1978; Krauss et al. 1996) and that gestures and speech do share interactive processes but only with gesture affecting speech, and not with speech affecting gesture (Feyereisen 1997). The second perspective is more recent and claims that gesture evolves together with speech (Kendon, 2004; Masur, 1990; McNeill, 1992). Gesture and speech would, in this case, interface one another through the interaction of spatial and verbal representations (Kita, 2000; Özyürek & Kita, 1999; Özyürek, 2002).

Speech	Vocalization			Conventional speech
Gesture	Gesture	Gestures and divergent gaze	Combination of gestures	Complex use of gestures

Tab. 3 L1 development (Masur 1990)

Gestures evolve with age and, while growing, children implement more gesture categories. The first gesture category that is used by children is the concrete deictic one, which is the only one used before the age of two (McNeill, 1992). Before the age of three, children start to use depicting gestures, and tend to use extended pantomimic movements of the whole body (McNeill, 1992). The rest of the gesture categories, such as beats, abstract deictics and metaphorics tend to appear after the age of five and their development reaches completeness around the age of 12 (Cassell, 1988; McNeill, 1992). In general, the use of representational gestures (representational and metaphorical) decreases with age, whereas beat gestures tend to increase (Freedman et al., 1986).

Age	4	10	14
Gesture use	Substitution of speech	Together with speech	Subordinate speech

Tab. 4 Gesture use in relation to age (adapted from Gullberg, 1998)

3.2 L2 acquisition

When people decide to learn a foreign language, they usually get through a very complex and dynamic process that gradually makes them vary their level of proficiency and takes them to manage that language (Larsen-Freeman, 1991).

In L2 communication, L1 gestures appear to have an effect on L2 gestures at all stages of language development. In fact, L2 acquisition is characterized by processes of transfer and interference of gestures from the L1 to the L2 that should be studied, together with verbal language, as part of the interlanguage (Gullberg, 2006, 2008b; Pika et al. 2006; Brown, 2007; Brown & Gullberg, 2008; Nicoladis, 1999; Nicoladis et al., 2007; So, 2010).

Gestures may have an important role in L2 acquisition process, but the importance of this aspect of communication has been generally underestimated. Studies that have investigated the role of gesturing in L2 acquisition are usually focused on the overall gesture behavior of L2 learners and are normally concentrated on the study of emblems, for instance, suggesting the use of specific gesture inventories in language classrooms in order to better introduce the culture and gestural styles belonging to the taught language (Hayes & Green 1971; Gullberg, 1998).

Other empirical studies have been carried out on the possibility of gesture and speech elicitation in order to understand what kind of information gestures provide or add to speech (Marcos, 1979; McNeill, 1992, Goldin-Meadow & Alibali 1995; Tellier, 2008). In this way, it has been possible to better determine how mental representations work. For instance, Marcos (1979) studied the gesturing of English/Spanish bilinguals and found that they generally used more gestures in the language they dominate the less, with significance found for beats and self-adaptors. These results not only apply to Spanish or French (Marcos, 1979; Sainsbury & Wood 1977), which are known to be high-gesticulation languages, but they apply also to cultures that are considered to

be gesture-discouraging. In fact, it has been found that also Japanese learners of English increase their use of gesture in L2 production (Kita 1993; Nobe 1993), with a significant increase of the beat and representational gesture categories (Nobe, 1993). Goldin-Meadow & Alibali (1995) investigated the learning process of children that were asked to accomplish Piagetian conservation tasks and they found that, while learning the task, children produced mismatched gestures and speech, with the two features conveying different information: while saying the objects were not the same size, their gesture showed that they thought the objects were actually of the same size. According to Goldin-Meadow (1999, 2000), the mismatch between speech and gesture applies to adults too and can show whether L2 learners are in transition from thinking in their native language to thinking in the foreign language.

On the other hand, there also are some studies that found no increase in gestures in the second language, as it happened with Chinese (Si-Qing, 1990) or Finnish (Valokorpi, 1981) that actually found that speakers did not change their individual gestural behavior when switching to their L2. In a study on international adopted children, De Geer (1992), found that they prefer to add complementary nonverbal behavior to speech as a communicative strategy when they arrive in their new country. As they start to develop the new language and to gain proficiency, their nonverbal communication decreases.

Further investigations have examined the role of gestures at different levels of L2 proficiency (Gullberg, 1998, 2003, 2006, 2008; Özyürek, 2002; Stam, 2006, 2008, 2016). Gullberg (1998) saw that Swedish speakers used more abstract pointing gestures more frequently in their L2 (French) than when speaking their native language in cases in which they could not use nominal pronouns. Gullberg's results show that abstract gestures compensate for the low proficiency of the subjects. Özyürek (2002) investigated whether subjects changed their gesturing when switching from one language to another, focusing on how the same speakers change their gestures as they use a language that is typologically different from their native language. Specifically, Özyürek analyzed gestures of Turkish subjects that used English as L2 at different proficiency levels and that were asked to tell a story to different listeners. The study showed that the same subjects change their way of gesturing when speaking an L2 and that this changes as they change the way of expressing information at each proficiency level. The findings by Özyürek confirm those by Gullberg (1998) revealing that subjects use gestures as a compensation strategy; however, both authors also conclude that there is an interaction between spatial and verbal representations in L2 production.

McCafferty (1998, 2004) compared the gesturing of a Taiwanese learner of English to the one of an American graduate student and reported that the Taiwanese student made a marked use of iconic gestures whereas the American student seemed to consider the use of iconic gestures as a teaching and an illustrating tool. In a further study (2004) McCafferty examined the use of gestures to understand whether the Taiwanese student was using gestures in order to acquire the L2. During the videotaping of the subjects McCafferty (2004) explains that "[…] the American participant initially established the use of iconic and other forms of gesture as explicit teaching tools, one use of which was to enhance comprehension through illustration of the verbal component. It did not take long before the L2 participant began to imitate these behaviors. However, during the sequences in question, the Taiwanese student's spoken English was quite clear, so it does not appear as though the gestures were necessary to help clarify meaning for his interlocutor." The conclusion of McCafferty's study provides evidence that representational gestures carry a communicative function but also that gestures were used by the Taiwanese to provide a greater degree of accuracy and were identical to speech. This might suggest that in L2 gestures can lead speech, helping to build the structure of the verbal channel.

Further studies on L2 communication, lead to the assumption that L1 gestures have an effect on L2 gestures at all stages of language development. In fact, L2 acquisition is characterized by processes of transfer and interference of gestures from the L1 to the L2 that should be studied, together with verbal language, as part of the interlanguage (Gullberg, 2006, 2008b; Pika et al. 2006; Brown, 2007; Brown & Gullberg, 2008; Ortega, 2009; Nicoladis, 1999; Nicoladis et al., 2007; So, 2010). Some studies suggest that bilingual speakers might gesture more than monolingual speakers because gesturing helps them formulate their spoken message and is a way to compensate for the reduced proficiency in their L2 (Nicoladis et al. 2007; Sherman & Nicoladis, 2004) In addition, speakers with low levels of competence might use more L1-specific gestures than speakers with higher levels of competence (Nicoladis, 2007). L2 speakers' greater use of gestures than L1 speakers might be explained on cognitive grounds, that is, due to the cognitive complexity that speaking a foreign language requires (Kita, 2000).

Contextual factors such as task expressiveness, nervousness, as well individual factors might also affect L2 speakers' gestures. Nicoladis et al. (2007) examined the relationship between gesture use, L2 proficiency level and task complexity in a story recall task. They found only weak evidence supporting the idea that increased task complexity leads to increased gesture use, and suggest that gesture use might also be related to expressivity, as well to the speaker's gender.

Wessel-Tolvig (2016) investigated the role of the gesture conceptualization in manner and path in Danish learners of Italian. The objects of his analyses co-verbal behavior in studies of SLA to in order to better understand the conceptualization that learners of another language employ during their language development. Different languages, in fact, carry different lexical conceptualizations of motion and path. His study lead to the conclusions that, at least for the domain of motion, learners are able to reorganize semantic representations as in the L2 pattern and to adapt to the target language.

4. Gestures and Narration

Narration is an important feature of speech. When people are engaged in discourse and narration their gestures are related to the specific narrative level they are using. Gestures are distributed among three different levels of narrative (Hopper, 1979; Hopper & Thompson, 1980; Gullberg, 1998):

a) *Narrative* (storyline): this is the level in which the narrated story advances and is presented in sequences and is temporally ordered.
b) *Meta-narrative* (background): which is formed by all the comments on the storyline that are presented with no link with the temporality of the story. In other words, this is the case in which the story and its structure becomes the referent of the narration.
c) *Para-narrative* (the situation): which are comments on the situation in which the narration is performed, like for example when talking about the relationship between two interlocutors.

Across these three levels, gestures seem to be regularly distributed (Levy & McNeill, 1992; McNeill, 1992; McNeill & Levy, 1993).

Iconic gestures depict the concrete objects or actions of what is said by speech and, thus, this kind of gesture is typically closely linked to the semantic content expressed in the narrative level by the speakers. As found out by McNeill & Levy (1980), iconic gestures tend to depict whole scenes which, through gestures, are represented in several more aspects than they are with words, since the choice of verbal expressions is constrained by grammar aspects of the language (McNeill & Levy, 1980).

Metaphoric gestures are more often employed at the meta-narrative level, since they are used to convey metaphors through an iconic depiction. In narration, these gestures are used to show and describe the story as an object of reference (Gullberg, 1998).

Beats do not depict any concrete aspect of the situation described by the co-occurring speech and, for this reason, are almost exclusively used as a transition from one narrative level to the other. Thus, they seem to be more frequent at the boundaries between different utterances (Gullberg, 1998). In L2 production, beats seem to be used to underline self-correction whenever the speakers recognize their speech problem and try to fix it (Gullberg 1998).

Pointing or deictic gestures occur at all levels of narrative, whenever direction or orientation are mentioned. The most common deictics in narratives are the abstract pointing gestures. Concrete deictics are not so commonly used, since usually the speaker is narrating or referring to something which is not present at the moment they are speaking. Abstract deictics can therefore be used to accomplish different functions (McNeill et al., 1993). At the proper narrative level, these gestures are used to introduce referents and to locate them in space (Levy & McNeill, 1992). At the meta-narrative level, they mark new events and, at the para-narrative level, they show the relationship between the speaker and their recipients (Gullberg, 1998; Marslen-Wilson et al., 1982).

As reported by Gullberg (1998), when they are used in second language narratives, gestures can be used to exploit different strategies of communication. Her experimental research leads to the conclusion that learners use iconic and referential metaphoric gestures to help themselves find and retrieve lexical items: by depicting the word with their hands, speakers get to elicit the word they need in their second language. Gullberg (1998) also found that deictics, besides helping the speakers retrieving words (as iconic and metaphoric gestures), can also be employed to create redundancy and keep coherence in learners' narrations (for example by making depicted reference to an imaginary spatial map). According to Gullberg, then, learners tend to over-mark reference in L2 through the use of abstract deictics, sometimes combined to iconics.

3/ Cognitive and communicative functions of gesture

1. Introduction

McNeill (1985, p. 353) provides an example to illustrate the gesture-speech interaction:

'she chases him out again'
[hand, gripping an object, swings from left to right]

McNeill (1985) affirms that the sentence in the example above is well formed and therefore the gesture cannot be considered as a repair or some other transformation of the sentence. McNeill's argument is that "to get the full cognitive representation that the speaker had in mind, both the sentence and the gesture must be taken into account" (McNeill, 1985).

The theoretical perspective proposed by McNeill contrasts the theory presented by Butterworth and Hadar (1989) that considers speech and gestures as two separate channels. Butterworth and Hadar (later expanded in Hadar & Butterworth, 1997) claim the possibility of other kinds of relationships between gesture iconicity and speech. They report the utterance:

> 'when certain problems can be raised'
> [hand starts to rise on the word 'certain']

In the example above, the *hand rising gesture* seems quite redundant with respect to speech. Another interesting thing about this example is given by the fact that the *hand rising gesture* begins before the verbal expression to which it refers to. This seems to show that "the speaker knew what the word would be, or at least had a pretty good idea, well before he uttered it" (Butterworth & Beattie, 1978). Moreover, the temporal asynchrony between gestures and speech commonly occurs in spontaneous speech above all when hand gestures are associated with low frequency lexical items (Butterworth & Beattie, 1978). These findings led the authors to conclude that gestures might be the result of planning processes of speech: the speakers know in advance

the semantic content they intend to convey and, while lexical retrieving, they produce a gesture that actually helps their cognitive process.

2. Gesture, Cognition and Thinking for Speaking

The appearance of co-speech gestures is often seen as the consequence of activating patterns of action (seen as gestures) while formulating meanings in lexical syntactic forms (Rimé and Schiaratura, 1991). To support this claim, Rimé and Schiaratura (1991) advanced three observations: first, after a verbal face-to-face interaction people are usually unable to remember their interlocutors' hand-gestures. Second, they affirm that some experimental studies (Rimé, 1982) indicate that comprehension is not compromised if the interactants cannot see each other. Third, subjects are not able to guess the speech content to which gestures refer. Thus, according to Rimé and Schiaratura, co-speech gestures seem to be just a by-product of the process of verbal articulation and may just play a helping role for the speaker, leaving the possible communicative function of gestures as merely incidental.

For what concerns iconic and representational gestures, in 1989 Butterworth & Hadar published a study in which they developed the theory that iconic gestures in speech may be due to aspects of lexical search and, moreover, that such gestures play an important functional role in lexical retrieval. In fact, the main point of their claim is that these kinds of gestures are hardly understandable without the co-occurring speech. In this light, gestures do not help addressees' comprehension of what is verbally uttered, but they would rather work as a speak facilitating tool for the addressers.

In 1997, Hadar & Butterworth developed their model for 'Gesture and the facilitation of lexical retrieval' (see fig. 6). Through this model, their intent is to explain how the conceptual processing automatically activates imagery. According to them, a visual image mediates between conceptual processing and the generation of iconic gestures: the visual image facilitates lexical retrieval in three different ways:

a) by 'focusing conceptual processing'
b) by 'holding core features during semantic reselection'
c) by 'directly activating word forms in the phonological lexicon'

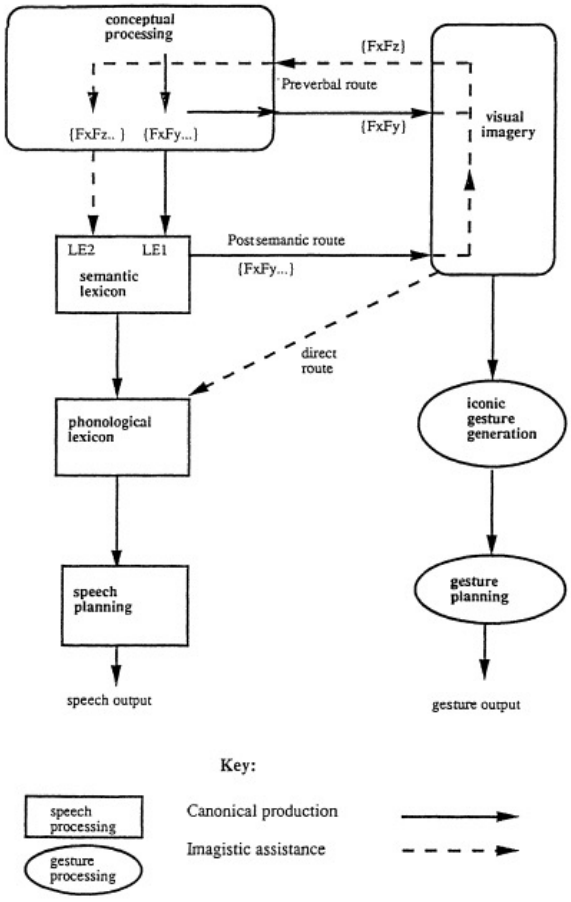


Fig. 7. The relation between iconic gestures generation and speech production (Hadar & Butterworth, 1997)

In the figure above (fig. 7.) Hadar & Butterworth (1997) assume a two-stage model of word retrieval.

The conceptual processing creates or selects a set of semantic features that have to be linguistically performed (indicated as Fx, Fy,…). These features, besides matching with lexical items, may also activate visual images through the preverbal route. According to Hadar & Butterworth (1997), a visual image can activate the conceptualization process, and thus facilitate the word-finding processes. Hence, the visual image can be translated into semantic features and, thus, into iconic gestures. Occasionally the translation process will not be exactly the same as the

original feature that evoked the image. Context or intention may determine which aspect of the visual image is salient and activates a gesture. For example: when speaking about a table, its squareness will surely elicit a different gesture from its flatness (Hadar & Butterworth, 1997).

One of the most debated hypotheses supporting the fact that the main function of gestures is to help speakers is the Lexical Retrieval Hypothesis (Rauscher et al., 1996; Krauss et al., 1996). According to this hypothesis, the production of a gesture keeps certain spatio-dynamic images activated and, in turn, keeps a set of conceptual features activated during the retrieval of the lexical item.

A further analysis to prove the hypothesis that gesturing facilitates lexical retrieval was made through a Tip Of the Tongue (TOT) study (Beattie & Coughlan, 1999). The authors elicited TOT states, in which a person is sure about the word they want to say but cannot temporarily access it. The subjects were presented with definitions and were asked to deduct which was the word they referred to. Only half of the subjects were allowed to gesture, while the rest were asked to keep their arms folded during the experiment. Contrary to the authors' predictions, that expected gestures to help subjects resolve the TOT states, the main finding of this study was that significantly more TOT states were solved by the non-gesturers, meaning that the presence of gestures actually lowered the probability of solving the TOT. To explain these results, de Ruiter (2003) proposes his theory of *channel compensation mechanism* according to which gesturing lowers the probability of resolving the word finding difficulty because it actually compensates for the interruption occurring in the speech channel, meaning that the more difficult to find the word is, the more the gesture channel will be activated.

Other studies about the cognitive function of gestures are those focusing on gestures that help speakers to organize their thinking for speaking. Kita (2000) presented an alternative view to the classical Lexical Retrieval Hypothesis with his Information Packaging Hypothesis. According to this theory, representational gestures are involved in the *conceptual planning* of messages, helping speakers to package information into verbal units. Kita (2000) defines this hypothesis in these three main points:

a) The production of a representational gesture helps speakers organize spatio-motoric information into packages suitable for speaking.

b) The trigger of representational gestures is spatio-motoric thinking, and it provides an alternative informational organization that is not accessible through analytic thinking and thus helps the speaking action.

c) Spatio-motoric thinking and analytic thinking have access to different sets of informational organizations but, during speech production, their ways of representation tend to collaborate and be coordinated.

Hence, the main idea in the Information Package Hypothesis is that speakers have access to a wide range of possible mental organizations since they can take advantage of both the analytical and the spatio-motoric thinking.

Alibali, Kita & Young (2000) tested the Information Packaging Hypothesis with children through two different tasks that required similar lexical access, but different information packaging. The first task was an *explanation task*, based on the Piagetian Conservation task[4], in which the children had to explain the reasons why two items had or did not have the same quantity. The second experiment consisted in a *description task* in which the children had to describe in which way two items looked different. In the results, the authors reported that in the explanation task, more than in the description task, the children performed a higher number of gestures conveying the perceptual dimensions of the objects, and of gestures with a different semantic from the accompanying speech. These results, led the authors to the assumption that gesture is involved in the conceptual planning of speech.

The Information Packaging Hypothesis seems to support the idea that gestures help speakers choose in which way they can organize their own perception. Thus, the *information packages* are meant to help speakers breaking down their perceptual selection into verbal units (Kita, 2000). Hence, gestures seem to help speakers represent knowledge in verbal form.

Alibali & Kita (2010, 2017) further investigated the role of gestures perception and the reason why speakers produce gestures. The authors tested the hypothesis that gestures can facilitate the conceptual planning of speaking. According to this hypothesis, gestures promote thinking about perceptually available information. To verify this hypothesis, Alibali and Kita (2010) analyzed

---

[4] Jean Piaget (1954), in order to understand children cognitive development, created seven conservation tasks to be able to test a child's ability to see that some properties are conserved or invariant after an object undergoes physical transformation.

the effect of gesture prohibition on children. Children are very likely to concentrate on perceptually existing information during conservation explanations. Thus, the experimental group performed two sets of conservation tasks: while in the first one gesture was allowed, in the second set it was prohibited. The gesture production of the experimental group was compared to the performance of children in a control group, which completed both tasks with gestures allowed. As a result, the authors found that during the explanation of the Piagetian conservation tasks, if gestures were prohibited, children were more likely to concentrate on information that was not perceptually present. On the other hand, when gestures were allowed, they were more likely to focus on perceptually present information. Hence, it seems that the presence of gesturing could actually activate speech more highly about perceptually present information. According to Alibali & Kita (2010), this correlation would explain the participation of gesture in the conceptual planning for speaking. In fact, when gesture was not allowed, children seemed to be less involved with the objects in front of them and, rather, seemed to be likely to express other kinds of information that was not related to the present (i.e. how the objects looked before the transformation, how they could be transformed, etc.). The authors found that children expressed more *resource-intensive* explanation when gesture was allowed, meaning gesture might help manage capacity demands. In fact, when gesture was not allowed, children used *resource-lean* explanations, which they could express without the use of gesture as memory assisting tool. In this way, these findings confirm those by Butterworth & Hadar, 1989; Krauss et al., 1996; Rauscher et al., 1996 and thus lead to the assumption that gestures actually facilitate word-accessing: gesture prohibition did not let children access the spatial words they needed to express perceptually present explanations, so that they shifted the explanations to those that did not require gestures.

3. Beyond cognition. The Communicative Function of Gestures

Even though a bulk of research shows that the main function of gestures is to help speakers's cognitive processes, many studies have addressed the idea that co-speech gestures may be produced for the benefit of the recipient.

According to Kendon (1985) "gesticulation arises as an integral part of an individual's *communicative* effort". However, studies investigating the hypothesis that gestures play an

important role in communication have not provided conclusive results (cf. Bull, 1983; Kendon 1983, 1985, 2004).

A number of studies, that can be considered as the bases of this approach, have shown that the addressees' comprehension is greater when they can see the speakers' gestures (Berger & Popelka, 1971; Graham & Argyle, 1975; Rogers, 1978; Riseborough, 1981).

Berger and Popelka (1971) asked their subjects to write down what they heard when a sentence was told them with a soft voice by a speaker that was standing at a distance of, more or less, 1 meter from them and compared the precision of what the subjects reported according to whether the sentences were accompanied by gestures (mostly emblems) or not. The results led to the conclusion that the accuracy of hearing the spoken utterances were twice as high when the speech was combined with a gesture.

Graham and Argyle (1975) asked speakers to describe abstract line drawings to a small audience of listeners who then tried to reproduce the drawings. Only for half of the descriptions, speakers were allowed to gesture, while for the rest they were asked to keep their arms folded. The researchers found that audiences of the non-gesturing speakers reproduced the figures somewhat less accurately. Although this finding is consistent with the proposition that gestures are communicative, it is possible that forbidding speakers to gesture could affect their speech production. In fact, speakers who had the permission to gesture may have produced better verbal descriptions enabling their listeners to reproduce the figures in a more accurate way.

Rogers (1978) presented 60 subjects with video-recordings of speakers explaining to an interlocutor short scenes that they had just watched or saying something about abstract topics. The speakers in the video were naturally and spontaneously producing gestures accompanying what they were saying. After watching the video recordings, the subjects were asked a number of questions designed to understanding the accuracy of their comprehension and their understanding of different grammatical aspects of the descriptions (i.e. agent, action, location, recipient of action, verbs, complements, etc.). The observers saw the videos in different conditions: with the only sound available, with sound and image, and with sound and a modified image in which the speaker's lips were hidden. In each of these conditions the sound was altered using four different levels of white noise, in order to compare different sound conditions: ranging from very difficult to hear, to very easy to hear. The results led Rogers (1978) to report that there was a greater understanding of the descriptions in the two visual conditions rather than in the

audio-only condition. Since comprehension got better results in the condition in which the lips of the speakers were hidden than in the audio-only condition, the conclusion is that hand gestures actually help the observers' understand. Furthermore, Rogers discovered that when it was more difficult to hear what the speaker was saying, there was a greater difference in comprehension between the visual conditions as compared to the audio conditions.

A further investigation on the effect of gesture in comprehension was done by Riseborough (1981), who conducted three experiments that led her to the conclusions that when the subjects are asked to guess what is someone is talking about, they guess it more precisely if they can see the gestures that accompany speech. Riseborough also found that when subjects are asked to remember a list of verbs a person is saying (on a video), they remember them better when the verbs are accompanied by well-defined movements rather than when they are accompanied by vague or no movements. Also, Riseborough (1981) added a condition in which two different levels of noise were presented, and she found that the more difficult it was to hear the speech, the stronger was the contribution of the well-defined gestures to the accuracy of comprehension.

Another kind of experiment on the role of gestures in communication is the one centered on the judgements of co-speech gestures in a condition of speech absence as those by Feyereisen, van de Wiele, and Dubois (1988); and by Krauss, Morrel-Samuels, and Colasante (1991).

Feyereisen, van de Wiele, and Dubois (1988) video recorded two lecturers while giving talks to university students and selected segments in which iconic and batonic gestures occurred, they chose 32 gestures for each speaker (16 iconic and 16 batonic). These selected recordings were showed to three different groups of 20 students each, first without sound and then with sound. For the first and second panel, the authors asked the students to discriminate iconic and batonic gestures and their judgements were analyzed in order to see how much they agreed with the classifications made by the experimenters. It was found that the students' judgements agreed more with those made by the experimenters when they had access to sound. However, a proportion of gestures were judged the same regardless of whether there was sound or not. Therefore, the gestures produced by the professors included some that were consistently recognized as gestures that conveyed meaning, regardless of the presence of speech, and some others that were perceived less consistently as their statuses could change whether accompanied by speech. When the gestures were presented without sound, students (first and second panel) were asked to guess the content of the utterance produced by the speakers and what they said was

collected and presented to the third panel. Thus, the authors presented to the third group a sample of those gestures that had been judged consistently as iconic by the other two panels as well as another gestures sample that was instead incoherently judged by them. This group panel was asked to indicate one of the three given verbal utterances to the gesture they were seeing. They could choose among the correct words, those that had actually been said by the lecturers, they could also choose plausible words, those that had been most frequently guessed by the students of the previous panels, and finally they could indicate implausible words, those that had been guessed only once in the previous sessions. The experiment results indicated that, in respect to iconic gestures, the third panel matched more often plausible and correct words to them, choosing the plausible words more often than the correct word. Feyereisen et al. (1988) concluded that gestures produced by speakers actually do convey semantic information, but that it seems to be a very general one. Also, the authors concluded that some gestures suggest what words were said in relation to them, while others just evoke a range of possible meanings rather than a precise word.

Another important study that investigated the observers' interpretations about the meaning of co-speech gestures is the one by Krauss, Morrel-Samuels, and Colasante (1991). They chose a selection of video recordings of narrative discourses in which the speaker used gestures and submitted it to different panels and under a series of different conditions. The kind of gestures that were presented to the subjects were those that are defined *illustrators* by Ekman and Friesen (1969). First, the authors presented the gestures without sound to a group of judges who were then asked to choose between the actual lexical affiliate of the gesture or a random lexical affiliate. The panel resulted to be more inclined to choose the actual lexical affiliate rather than the random chosen one. In another condition, Krauss et al. (1991) asked another panel to simply guess the verbal expressions that could be associated to the gestures in the videos and to write them down. The experimenters showed that the judges were able to interpret and guess the gestures' lexical affiliate, being the interpretations really close to the actual verbal expression. Krauss et al. (1990) also found that gestures make contribution to the semantic category assignment. In fact, a further panel was able to assign gestures to a specific category, choosing between five given categories with a certain degree of accuracy, even if less precisely than when gestures were presented combined with the co-occurring verbal utterance. Hence, the study by Krauss et al. (1990) leads to results that are coherent with those reported by Feyereisen, van de Wiele, and Dubois (1988) and to the conclusion that, even though they do not communicate as

much as speech, conversational hand-gestures do convey semantic information and, thus, they do contribute to the aim of communication.

Even if the studies cited above somehow confirm that gestures actually do play a communicative function, the authors' attitude to the communicative import of co-speech gestures' is not clear. In fact, even though these studies produced evidence that co-speech gestures convey some semantic information related to the co-occurring verbal expression, their authors tend to conclude that the role of gestures have a mainly occasional and incidental role in communication. Anyhow, they affirm that some gestures do have a primarily communicative function, as it may be the case for *emblems*, *pointing* gestures and *interactive* gestures distinguished by Bavelas et al. (1992).

Another studies on the possible functions of gestures are those based on observation in more or less natural situations. In order to organize and interpret the different ways and circumstances of gestures use, Kendon (1985) was able to provide a starting point for considering systematically the ways people employ gestures. In this way, Kendon (1985) was able to indicate that there are various situations in which gestures help the listener understand the actual semantic meaning of a verbal utterance. Moreover, Kendon (1985) also described the ways in which speakers supplemented their speech through the use of gestures. Gestures, for example, were observed to be used as a means of completing an expression that was not socially accepted to be said out loud. Another example of gesture usage described by Kendon (1985) is the supplementation of speech in cases in which ambient noise did not permit the listener to hear well what was being said by the speaker. Further instances were noticed by the author in which gestures were used by the speakers as an additional dimension of meaning of what was being said, such as indicating the shape, size, location of something, or even pointing at an object or location in order to identify what the speaker was talking about.

Besides the observational studies conducted by Kendon (1985), a number of further studies were based on the close analysis of conversations in natural settings (M. H. Goodwin, 1980; C. Goodwin and M. H. Goodwin, 1986, 1992; Moerman, 1990; de Fornel 1992; Streeck and Hartge, 1992; Streeck, 1993). These studies showed that there are occasions in which speakers actually do take advantage of gesturing to understand or to deliver the proper meaning of what is being said.

Some studies support the idea that kinetic actions may be used to regulate the visual attention of recipients. C. Goodwin (1986) reported examples in which *non-gestures*, such as a *manipulators*

or *adaptors* (following the categorization of Ekman and Friesen, 1969) in which the speakers touch their face, may function in a way to make the recipient look away or get distracted. Instead, in a study by Heath (1992), it was found that gestures help the listener bring their attention back to the speaker when distracted. Furthermore, Heath (1992) also showed that listeners can actually understand the meaning of an utterance in the moment the speaker gestures, without waiting for its linguistic affiliate. One of the examples that Heath (1992) reports involves a doctor explaining to a patient that a medicine will "damp down" a symptom while making a number of downward movements with his hand. Then the doctor says "they help sort of you know to dampen down the inflammation", completing the several downward movements before saying *you know*, and it is here that the patient started nodding with his head as to confirm he understood what the doctor meant. M.G. Goodwin (1986) found a similar example when she noticed that a recipient was able to understand the meaning of a sentence the moment the speaker gestured, even if the sentence itself was not concluded yet: the speaker, in fact, while describing the size of an object, paused in the middle of the sentence, spread her hands and held a "size specifier" gesture. Even though she had not concluded the sentence, the listener started nodding, showing they understood the content of the whole utterance.

Streeck and Hartge (1992), investigated the possible use of gestures as a way to introduce a talk, concluding that the appearance of these gestures is perceived by the recipients as evidence of something to come and works in parallel with verbal turn prefaces described in conversation analyses (Schegloff, 1980; Goodwin and Heritage, 1990; Streeck and Knapp, 1992; Cowley, 1998). In particular, Streeck and Hartge (1992) analyzed conversations of Ilokano speakers (the third most-spoken native language of the Philippines) and their use of a facial expression (that actually seems to recall the mouth position when about to pronounce the vowel /a/) as a forewarning of a talking turn to come. Together with this turn preface facial expression, the authors also analyzed a hand gesture that seem to have a similar function: the gesture is formed by the hand palm held upward and that is usually performed when the speaker wants to communicate they have more to tell.

From the considerations above, it appears clear that gestures and speech form part of the same process and that they provide a basis for some cognitive processes that lead to speak but at the same time also do convey information to the addressees, playing a communicative function. Yet, the extent to which gestures contribute to speech cognitive or to communicative functions is not clear. Some studies (i.e., Rimé and Schiaratura, 1991; Krauss et al, 1991) lead to the conclusions

that the communicative functions of gestures are mainly merely incidental and occasional. Nevertheless, many other studies (i.e. Kendon, 1986, 1983; Goodwin, 1986; Bavelas, 1992, 1995) report that recipients actually show a better understanding of an utterance when speech is accompanied by gestures, but also that some gestures are produced more frequently when the speaker is interacting with their addressees or is engaged in a conversation with someone. Therefore, these studies lead once again to the assumption that speech and gestures are part of the same process. Thus, gestures are produced as part of the speaker's cognitive process that steps along with the verbal articulation. However, at the same time they can provide some semantic content that helps the communicative function of gestures, and addresses the listener to the right comprehension of the utterance itself. Thus, there is the need to clarify what are the functions of gestures and if some gestural categories are used more in the cognitive or the communicative function.

3.1 The role of representational gestures in communication.

As seen in the precious section, whether and how gestures generally contribute to communication seem to be a complex issue. The communicative function of gestures seems to be necessarily analyzed in relation to the different gesture categories. In fact, if the majority of studies are in general agreement about the fact that *deictic* gestures are often produced with the aim of communicate, on *representational* gestures there is a more complex debate. *Representationals*, in fact, share a strong relationship with the semantic aspects of the co-occurring speech since they often represent concrete or abstract objects, traits, manner, paths or activities.

Most studies aimed at defining whether gestures have a communicative function have investigated how listeners' comprehension of an utterance might be influenced by the presence of gestures (Krauss, Morrels-Samuels, & Colasante, 1991; Krauss et al, 1995; Feyereisen et al, 1988), or how listeners, in order to understand a narrative, normally integrate speech with the information expressed by co-speech gestures (Beattie & Shovelton, 1999a, 1999b, 2003; Beattie et al., 2014; Cassell, McNeill, & McCullough, 1999; Gullberg, 2003; Holler & Beattie 2002, 2003; Holler, Beattie & Shovelton, 2009).

Beattie and Shovelton (1999) investigated the range of information contained in co-speech gestures. Using a structured interview, the authors tried to measure the amount of information

perceived by respondents exposed to clause-length clips containing depictions of a cartoon story. Their analysis led them to discover that whenever subjects were able to see the iconic gestures as well as hear the speech, there was an increase of 10.8% in overall accuracy in their responses, meaning they perceived significantly more precise information about certain aspects of the original clips of the cartoon than when they could only hear the verbal depictions. Anyhow, in a condition in which respondents could see the representational gesture without being able to hear the speech, there still was a significant amount of information being conveyed. Moreover, Beattie and Shovelton (1999a; 1999b), also discovered that the positive effect of gestural communication was significant only with respect to some specific semantic categories, such as the iconic gestures referred to relative position and size of objects. This study seems to confirm the theory proposed by McNeill (1985) that iconic gestures convey important information in interpersonal communication, demonstrating that receivers actually gather information from representational gestures as much as from speech.

In a following study, Holler, Shovelton & Beattie (2009) tried to further investigate whether, in a face-to face context, iconic gestures really do contribute to the communication of semantic information. Since all the evidence on the impact of iconic gestures as a communicative source comes from studies in which subjects were exposed to video-recorded samples, the authors felt it was also necessary to understand the role these gestures had in face-to-face situations. To do so, they compared the amount of semantic information that recipients received in contexts in which gestures and speech were presented on video and when they occurred in face-to-face contexts. During a face-to face conversation, besides speech and intentional gestures, people use a wide variety of non-verbal behaviors, but none of these are as linked to the semantic content of speech as much as the iconic-representational gestures. Since their previous works (Beattie & Shovelton, 1999a, 1999b, 2002) found that iconic gestures that relate to position and size communicate information, the authors concentrated on the same gesture categories. In this case, Holler, Beattie and Shovelton (2009) used four conditions:

- 1st: the participants watched and listened to cartoon narrations in a face-to-face context;
- 2nd: the participants watched and listened to video-recorded cartoon narrations in a TV screen;
- 3rd: the participants only listened to the video-recorded cartoon narration (since the monitor was turned off);

- 4th: the participants only watched the video-recorded cartoon narration (since the audio was turned off).

Since the iconic gestures selected as stimuli had been shown to be highly communicative in the previous studies (Beattie & Shovelton, 1999a, 1999b, 2002), the authors decided to make an actor deliver the chosen gestures, so that they could obtain evidence that these gestures were still communicative even if not produced spontaneously. The aim of the research was to clarify whether iconic gestures are communicative in various contexts or not. The results led to the conclusion that iconic gestures, namely those conveying size and position information, communicate effectively both when seen on a TV screen and when presented in a face-to-face situation. However, iconic gestures seemed to communicate significantly more information in the face-to-face condition than in the video mediated condition. The study also showed, in line with previous studies, that 'gestures do communicate a considerable amount of semantic information even in absence of speech' (Holler, Sholvelton & Beattie, 2009).

Thus, the research by Holler, Shovelton & Beattie confirms that iconic gestures tend to accompany speech, and convey semantic information to the receivers even when speech is not audible. While previous studies (e.g. Bettie & Shovelton 1999a, 1999b; Riseborough, 1981; Rogers, 1978) found that gestures conveyed considerably less information than speech, in this study the authors showed that information was equally well communicated by the two modalities. This was probably due to the fact that they chose gesture stimuli that they knew were very communicative and that they only considered those two semantic categories (iconic of size and position) that had been shown to reliably communicate in previous investigations employing video presentations. Thus, it is not appropriate to generalize the results to the whole iconic gesture category by saying they communicate as much information as much as speech does. Rather, this study provides an important insight on the communicative effectiveness of two kinds of iconic gestures, namely those related to size and position.

Only a few studies have investigated whether iconic gestures are actually intended by the speaker to be communicative. Besides the studies of Cohen & Harrison (1973) and Bavelas et al. (2002, 2008) that showed how face-to-face interaction (even when the interlocutor is not visible) bring to a greater gesture frequency and a use of different types of gestures, Graham and Heywood (1975) focused on the effect of gesture prohibition on the speech content and found that it causes an increase in the number of words when the speakers had to describe spatial relations and, at the

same time, a decrease in the use of deictic expressions. These results lead to the assumption that gestures are also used to convey information that might not be expressed verbally. Later on, Melinger & Levelt (2004) investigated whether speakers intentionally use a specific kind of iconic gestures (which they call *iconic tracing gestures*) in order to convey part of their message, thus showing that these gestures are part of the communicative strategy of the speaker. To do so, the authors analyzed the spoken descriptions of a picture in which there were networks of colored circles organized along a path. The subjects were asked to describe the picture to a visible interlocutor and to identify the color of each circle and the spatial relationship it had with the closest circles. The aim of the study was to understand whether the inclusion of gestures is related to verbal directional omissions and their frequency: if the speakers use gestures instead of speech in order to convey a message, then the use of gestures should be related to the omission of information through speech. If the speakers do not use gestures in a communicative way, then there should not be any relationship between the occurrence of gestures and the verbal omissions of the information. The authors collected 480 picture descriptions made by 30 university students, native Dutch speakers. In the descriptions they collected, Melinger & Levelt (2004) chose to let the speakers free to produce gestures or not. The results of the authors' analyses support the hypothesis that gestures are meant by speakers to supplement and complement their speech. In fact, they found that some speakers, taking advantage of the face-to-face interactional context, divided the information load between the verbal and the gestural modalities, meaning that gestures were meant to be communicative. Moreover, Melinger & Levelt (2004) also found some differences between the verbal descriptions produced by gesturers and those produced by non-gesturers. In fact, gesturers produced more directional omissions than non-gesturers and they also produced different kinds of verbal omissions than non-gesturers; specifically, the change of direction omissions occurred only when there was a visible co-occurring gesture. Finally, the authors observed that whenever speakers delivered a picture overview with gestures at the beginning of their description, they then had the tendency to leave out subsequent directional terms. A first hint of the iconic tracing gestures being often intended to be communicative is given in those circumstances in which some of the subjects chose to divide the necessary information between speech and gestures. In fact, if a necessary information is only provided through gestures, then the gesture is considered by the speaker as the proper way to express that information. The fact that some speakers chose to describe the pictures through both verbal and gestural modalities, and that there seems to be an interrelation between the linguistic omissions and the use of co-occurring gestures with the same semantic content, provides evidence for the

speakers' use of iconic gestures as an intentional communicative tool. Campisi & Özyürek (2013), for example studied the use of iconic gestures and focused on the fact that when the context changes, speakers tend to modify the way they employ gestures in quantity and size. The findings by Campisi & Özyürek showed that adults increased the rate of iconic gestures when explaining the usage of a machine to children, these results support the idea that iconic gestures can actually be used with a clear communicative intent.

4/ About speech prosody and its relations with gestures

1. Prosody

When people speak, they organize their speech through the use of features like sound, syllables, lexicon, grammar, syntax. These though, are only a few of the many components that, together, allow speakers to convey the verbal messages they want to communicate. Human speech, furthermore, cannot be fully represented as segments and sequences of phonemes, syllables or words. When carefully listening to natural speech, it is possible to distinguish more features that provide melodic and rhythmical properties to speech. For example, segments and syllables can be produced more or less long and pitch can be produced in a more or less varied way. Prosody, thus, can be defined as a cluster of features that do not belong to the segmental part of speech, such as intonation, pitch range, timing within intonation phrases, stress, duration and so on. In this thesis, the central aspect that will be considered is the pitch and its variation as part of communicative strategy by the speakers, and the influence that some features such as disfluencies and pauses have on subjects' speech rate and fluency.

In summary, then, prosody is formed by two main components: metric and intonation. While the metrical component concerns sentence stress and duration (Goldsmith, 1990; Hayes, 1995; Liberman & Prince, 1977; Vihman, 1991), the intonation component refers to changes in pitch (or fundamental frequency, F0) across the utterance (Ferreira, 2007) and this means that different kinds of tones lead to different intonation contours (Beckman, et al. 2002; Bolinger, 1986; Ladd, 2008).

Prosody plays an important role in communication since, for example, it allows speakers to disambiguate meanings, emphasize important words and communicate both at the semantic and at the pragmatic levels (Curl et al., 2006; Levis & Pickering, 2004).

Generally, prosodic typologies are not simple to categorize and every language has its own prosodic characteristics. Humans, however, can discriminate languages based on prosodic cues as suggested by studies that have investigated whether humans can identify languages by hearing synthesized simulation of different prosodies (Komatsu, 2002; Ramus et al., 1999) and suggest that prosody plays a role in language discrimination.

2. On Intonation

In general, intonation is referred to as the "pitch contour of an utterance" (Ahrens, 2005) and it is defined as the movement of pitch due to the changes in the fundamental frequency (F0) in the speech production (Cruttenden, 1997). In other words, F0, or fundamental frequency, refers to the acoustic correlate of the modulations in the vocal fold vibrations; whereas the way in which the variations in F0 are perceived by the listeners is commonly defined as *pitch* (Ladd, 2008; Vaissière, 2005; Halliday and Greaves, 2008). Thus, phonetically speaking, intonation is measured through the analysis of pitch modulation and pitch range (see §3) (Bolinger, 1986; Vaissière, 2005).

Fundamental frequency depends on the number of vibrations of the vocal folds, and this measure (in Hertz) can be expressed in cycles per seconds. At a different vocal fold speed corresponds a different F0 height: the faster the vibrations, the higher will be F0 (Ladefoged, 2003).

As reported by Wells (2006) speakers use intonation with different functions:

1. *Attitudinal function:* through the use of different tones, intonation is used to express emotions and attitudes;

2. *Grammatical function:* intonation gives a structure to speech and renders what punctuation does in writing;

3. *Focusing function:* it is also called accentual or informational (Wells, 2006) and is the function through which intonation helps distinguishing between new and old information in an utterance or to make the interlocutor focus on a precise chunk of information. This function is achieved through the use of tonicity. For example, in the sentence "let's meet

in front of the house tomorrow" if the words "the house" are used with a different tone, it appears clear that that is the is new or more important information to focus on.

4. *Discourse and cohesive function:* it is the way in which intonation shows which sequences of sentences go together in discourse, either to contrast or to cohere what is said in the utterance. This function works as the sentence and paragraph division in a written text. Furthermore, this function also signals whether the utterance is ending and, thus, has an important role in giving turns to the interlocutor;

5. *Psychological function:* through the use of tonality, intonation helps the interlocutors' perception and memorization;

6. *Indexical function:* intonation may mark personal or social identity. For example, a certain intonation may be one of the features that help identifying the romantic nature of a relationship between two interlocutors.

In summary, intonation has an important role in the communicative interaction, since it acts as a structuring and organizing feature of speech, for example it is used by speakers to indicate the end of an utterance and thus helps the interactants give and take turns in communication (Jin, 1990; Selting, 1995). Moreover, it is also used to express different moods and attitudes towards what it is being said (Mozziconacci, 2002; Wells, 2006).

3. About Pitch Range

Pitch range is defined by Clark (2003) as "the range of pitch employed by a particular speaker at a particular time and be specified by a minimum and a maximum pitch".

It is known that in many languages some meanings and emphasis are created by means of variations of the fundamental frequency (or F_0) of the human voice. The range over which these variations may occur and are perceived is called pitch range. Even though the use and production of pitch range may vary depending on language and sociocultural and/or sociophonetic factors (Van Bezooijen, 1995), it is generally recognized that a voice that is heavily modulated and thus presents a wide pitch range, will sound lively and energetic; whereas a voice characterized by a narrower pitch range will most likely sound as monotone (Mennen, 2007, 2012). Consequently,

pitch range and pitch variation offer a good measure tool to analyze speakers' perceived liveliness in discourse and narrations both in the first and second language production (Hincks 2004, 2005; Mennen, 2007). In order to evaluate the pitch range in this perspective, Hincks (2004, 2005) introduced an important framework of analysis to compare speakers' liveliness over long stretches of speech. In her analyses, Hincks (2004) considered the normalized standard deviation of F_0, and found that a value of pitch variation, which she called Pitch Variation Quotient (PVQ), highly correlates with perceived liveliness of the speakers, though only weakly with speakers' proficiency level. Pitch variation appeared to be a stronger perceptual cue to liveliness in male speech than in female speech. She concluded that pitch variation may not be the only measure of speakers' liveliness since rhythm and intensity are also measures of liveliness, but it can be considered as an important one.

All languages present differences in pitch ranges. Furthermore, pitch range can correlate with emotions and commitment. Specifically, while a wide pitch range can correspond to excited and involved participation, a narrow pitch can sound flat, boring or detached (Mennen et al., 2007, 2012). Even if pitch modulation varies across cultures, it is possible to recognize some major trends that are considered universal. Ladd (2008) recognized the following universal characteristics of pitch:

1. Pitch tends to drop at the end of an utterance, whereas it does not drop or even rises when the utterance is not completed yet even if there is a pause;
2. In questions, pitch is higher probably due to the utterance incompleteness (that will be completed by the recipient's answer);
3. Pitch peaks are used on words that, in the utterance, are considered more important than others.

4. Pitch in L2

When learning foreign languages, speakers often focus on the grammar, lexicon and syntactic features of the language they are learning. Prosody often happens to be considered secondary in the L2 learning process. In a review of previous studies, Mennen et al. (2012) collected a list of

errors upon which students with different language backgrounds often stumble during the production of L2 English prosody:

- L2 is generally characterized by a narrower pitch range (Backman 1979; Jenner 1976; Willems 1982)
- Speakers often have problems with the correct placement of prominence (Backman, 1979; Jenner, 1976)
- There may be a replacement of pitch rises with falls and vice versa (Adams and Munro, 1978; Backman, 1979; Jenner, 1976; Willems, 1982)
- Speakers may produce incorrect pitch on unstressed syllables (Backman, 1979; Willems, 1982)
- L2 speakers may start their utterance with a too low pitch (Backman, 1979; Willems, 1982)

A number of studies have focused on the idea that L2 speech may be characterized by limited pitch variation and a narrower pitch range than L1 speech (Hincks, 2004, 2005; Mennen, 2007; Mennen et al., 2012; Aoyama & Guion, 2007; Pickering, 2004; Traunmüller & Eriksson, 1995; Ullakonoja, 2007). It is possible, in fact, that prosodic information is managed differently by native and non-native speakers due to their different levels of proficiency in the L1/L2. For example, as suggested by Jenkins (2002), rather than on prosodic information, L2 speakers may rely more on segmental information to get their meanings across since they are lacking the suprasegmental knowledge that native speakers can rely on when communicating. An analysis of pitch range in L2 was also carried out by Zimmerer el al. (2014) who ran an experiment with French and German speakers. Their results suggest that speakers tend to produce a smaller pitch range in the respective L2. This was found to be true for both groups of speakers. The explanation that Zimmerer and his colleagues give is in line with what suggested by Jenkins (2002), that is, when speaking L2, speakers are less confident in their productions, therefore, they concentrate more on segments and words and subsequently refrain from realizing more native-like pitch range. Differences in pitch range in L1 and L2 may also be more prominent in particular speaking styles, such as formal presentations (Hincks, 2004, 2005; Johns-Lewis, 1986) during which non-native speakers may be particularly concentrated on conveying their semantic information while they neglect the prosodic aspects of their utterances.

4.1 Measuring global pitch range

A framework for measuring global pitch range cross-linguistically was first established by Ladd (2008) then elaborated by Patterson (2000) and finally by Mennen et al. (2007, 2012). Within this framework, a number of measures are used to quantify differences in pitch level, that is considered to be the speaker's overall pitch height or register; and pitch span which is defined as the speaker's range of frequencies in a verbal utterance. These include F_0 max, min, mean and median, as well as linguistic measures, linked to specific linguistically-defined landmarks in the F0 contour.

Furthermore, different types of gestures seem to have different effects on the production and perception of prosodic prominence. For example, beats (that is, sudden, baton-like, up and down, or back and forth movements that tend to reflect discourse structure by marking important words and phrases) have been shown to increase the acoustic duration, energy, and values of F_0 and F_1 of the pitch accent of the words co-produced with them; thus, it seems that beat production also increases the perceptual prominence of the co-produced words. In this way, beats function as facilitators of the listeners' task in communication (Krahmer & Swerts, 2007).

5. Prosody and Gestures

There is general consensus that gestures are temporally aligned with prominent parts of speech (Birdwhistell, 1970; Kendon, 1972, 1980, 2004; Bull & Connelly, 1985; Loehr, 2004, 2007, 2012; Esteve-Gibert et al., 2014; Esteve-Gibert & Prieto, 2013). Yet, the actual nature of this temporal relationship is not clear. Many investigations have provided contradictory evidence as to whether gestures and prominence are coordinated at the level of the focused word (Butterworth & Beattie, 1978; Roustan & Dohen, 2010), the lexically stressed syllable (Loehr, 2004, 2007; Rochet-Capellan et al. 2008), or the syllable with an intonation peak (De Ruiter, 1998, 2003; Nobe, 1996).

An important contribution to the understanding of the relationship and alignment between intonation and gesture was given by Kendon (1972) who associated a gesture hierarchy to the intonation hierarchy and who claimed that a gesture phrase generally coincides with a *tone group* or *tone unit*. A tone group is defined by the British School of Intonation as "the smallest grouping

of syllables over which a completed intonation tune occurs" (Kendon, 1972). In fact, Kendon claimed that the gesture stroke (the part of the gesture that can be considered the most salient one) generally occurs just before or exactly at the onset of a stress syllable. In further studies focused on African languages, Creider (1986) investigated the alignment of stressed syllables and gesture and described differences in hand movements between languages that seem to be conditioned by the different prosodic systems of the languages themselves. For example, he reported that in the Luo language, beats are timed with a nuclear tone and that a beat at the end of a falling nuclear tone would signal the end of the utterance (and thus give turn to the interlocutor).

Another important contribution on the gesture and intonation relationship was given by Bolinger (1983, 1986) and McClave (1991). Bolinger claimed that these two features are actually the representations of the same system, one is visible and the other is audible and that intonation is strictly linked to gesture, more than grammar. In his view, these two channels together express speakers' emotional states, moving in parallel, rising with emotional tension and lowering with relaxation. McClave analyzed video conversations and annotated gesture with the McNeill classification. On the same time, McClave also annotated tone units using pauses and pitch movements or changes as boundaries and found that gestural phrases align with tone unit boundaries and that holds and strokes (of gestures) usually last less if there are more gesture within the same tone unit. Thus, the results got by McClave suggest that speakers somehow know in advance that they will produce more gestures together with a certain concept. Furthermore, McClave (1991) also confirmed the hypothesis that beat gestures are temporally aligned with stressed syllables.

Zellers et al., (2016) investigated the role that prosodic and gestural cues may play in combination with one another to help interlocutors achieve smooth turn transition in Swedish. In a study about the speech and gesture timing, Valbonesi et al. (2002) found that gestures tend to occur during fluent speech and that gesture strokes generally align with stressed syllable.

Many studies focused on the shared features between gestures and spoken languages (Krahmer & Swerts, 2009; Mol et al.2011, Loehr 2012; Wagner et al, 2014). For instance, Swerts and Krahmer (2005) and Krahmer and Swerts, (2007, 2009) investigated the role of acoustic and visual prosody in the detection of epistemic information. The authors showed that there are well

defined visual cues that delineate the speakers' feelings and consciousness and that allow the addressee to estimate the mutual knowledge they share.

Esteve-Gilbert & Prieto (2013) who ran an investigation about the coordination between the peaks of pointing gestures and of intonation in words produced with different stress patterns in Catalan (as for example, monosyllables, tronchees or iambs). The results of this experiment lead to the conclusion that gesture and F0 peaks used to co-occur, this means that, for example, the peak of the gestures occurred at the middle of the stressed syllable when the word was a iamb but at the end when they were tronchees. In this way, the authors showed that both intonation and pointing gestures are linked to the prosodic phrase. In a further investigation Esteve-Gilbert et al. (2014) expanded the research also on the alignment between the prosodic structure of speech and head movements and concluded that head gestures do align differently according to where the stressed syllable is found in the word.

6. Pitch and Gestures

The studies on intonation and gestural relationship have been mainly focused on the temporal alignment of certain gestures (generally beats) with stressed syllables and intonation units.

Little attention has been paid to the relationship between speakers' global pitch range and use of gestures. Anecdotal evidence suggests that there the speakers' pitch variation in speech and their use of gestures might be related. For example, monotonic speakers generally do not gesticulate much, while animated speech is often characterized by a considerable amount of gesturing. It is likely in fact that speakers convey paralinguistic meanings both through their voice and their gestures.

In public speaking classes teachers tell students to use a 'lively' voice and to accompany 'purposeful' gestures when delivering a speech, emphasizing the idea that congruence in language, voice and gestures is at the basis of successful communication; it is suggested that while variations in the speaker's pitch range (and intonation, rhythm and volume) help listeners follow the information flow, gestures can contribute to maintaining the listeners' attention by providing them with a visual channel, in addition to the audio channel, that helps them follow what the speaker is saying. However, there is still little empirical evidence to support the methods

used in teaching public speaking. The present work aims at contributing data to fill this gap in the literature.

Assessing whether speakers' global pitch range and use of gestures are related is important also in L2 studies. In L2 speech, the speakers' combined use of non-native pitch and gestures (together with the use of non-native language structures) affect communication. Mennen et al. (2007, 2012) found that Southern Standard British English speakers have a higher and more varied pitch range than Northern Standard German speakers. As a result, Germans may sound "bored" or "unfriendly" to British listeners (Gibbon, 1998, 2009), and British voices, especially the female ones, may sound "over-excited" (Eckert & Laver, 1994) or even "aggressive" (Gibbon, 1998) to German listeners. Similarly, the use of frequent, broad, full arm, animated gestures during speech may be common and/or accepted in some linguistic communities, but may be considered distracting and cause annoyance to the listener in some other linguistic communities. On the other hand, too little gesturing may put off interlocutors that are used to greater gesturing in conversations (Axtell, 1991; Efron, 1972; Ekman and Friesen, 1969; Graham and Argyle, 1975; Okada and Brosnahan, 1990).

The few studies that have investigated the relationship between speakers' global pitch range and gestures have not found that this relationship is significant. Hoetjes et al. (2014) compared the speech characteristics of subjects that were asked to describe a tie knot with or without using gestures. The hypothesis was that not being able to gesture increases the cognitive complexity of the speech task, and this affects the subject's speech by making it less fluent and more monotonic (i.e., characterized by more filled pauses, smaller pitch range, lower intensity, lower speech duration, and lower speech rate). The authors also looked at the number of speakers' attempts (or repetitions), under the assumption that the first attempt is cognitively more difficult than the following ones. The results showed no effect of ability to gesture on the number of words produced, speech duration, speech rate, production of pauses, or pitch range. The number of attempts had an effect on the number of words produced, speech duration, and production of filled pauses, which all decreased after the first attempt, showing that, with repetitions, speakers become more fluent and use fewer words; but pitch and speech rate did not change significantly.

Instructing subjects to gesture, or alternatively not to gesture, is a common procedure in gesture studies. This procedure has often been used to investigate various aspects of the co-production of gestures and speech (Parrill et al. 2016). However, it is not clear that instructions to gesture

(or not gesture) will change speakers' gesture rate or type (Parrill et al. 2016). Thus, it is possible that it may not be the best procedure to elicit variations in speakers' global pitch range or other speech characteristics in relation to gestures. Because prosodic variation is used to express paralinguistic meanings, a better way to elicit speakers' variations in pitch might be to encourage speakers to be communicative in their speech, as is actually done in public speaking classes. This is tested in the present investigation.

Part II | Experimental analyses

5/ Does communicative effort influence pitch and gestures?

Experiment 1

1. A note on terminology

In the tradition of gesture studies the term *speech* is commonly used to refer to the semantic meaning of words and syntactic construction of utterances. In the tradition of phonetic studies, instead, *speech* is used to refer to speech sounds and their segmental and suprasegmental characteristics. In the present work, which draws largely from experimental phonetic studies, the term *speech* will be used to refer to speech sounds and their acoustic correlates. In particular, the main focus of the analyses about speech production will be on prosodic correlates such as fluency, disfluencies (corrections, repetitions), pauses (respiratory, filled and silent), speech rate and articulation rate.

2. Introduction

Gestures can have communicative and cognitive functions. Yet, how these two gesture functions are related to speech and prosody is worth further exploration. A specific speaking style can be linked to a specific way of using overall pitch range, but is the overall pitch range linked to the way speakers gesture?

For example, public speaking professors usually teach students to use a 'lively' voice when delivering a speech. This means that students should speak with a varied intonation, rhythm and volume. In fact, by varying intonation, rhythm and volume speakers can highlight those points of their discourse that they believe to be important and, on the other side, deemphasize others. At the same time, the way speakers emphasize or deemphasize certain parts of their speech helps listeners follow the information flow and focus their attention on what is considered important by the speaker and on the speaker's message (Hincks, 2004; 2005). Besides training students to vary their voices, public speaking classes also highlight the importance of body language. In fact,

students are usually told to use an "open" body position and to use gaze and gestures to emphasize the important parts of their speeches. In general, though, there is a lack of scientific research that confirms the real effectiveness of an integrated use of gesture and speech when the purpose is to be communicative.

The relation between speakers' global pitch range and their use of gestures is important also in L2 studies. In fact, for second language learners, speaking (above all if in front of an audience) involves planning their speech in a language that is not their native one, which means that they have to retrieve lexical items and plan their discourse structure together with the right intonation and gestures. Of course, all these factors may affect communication.

As for gestures, the use of frequent, broad, full arm, or energetic co-speech gestures may be common or accepted in some linguistic communities, but may be considered distracting or annoying in some other linguistic communities. It is known, for example, that there are cultures in which gesturing is much more accepted than in others. Italians, for example, are widely known to be expert gesticulators, in fact, many of the studies on gestures actually focused on the southern Italian way to gesture. Kendon himself studied and described the way southern Italians gesture (see Kendon, 1995, 2000). At the same time, though, the lack of gesturing may confuse or distract recipients that are used to greater gesturing in conversations (Axtell, 1991; Efron, 1972; Ekman and Friesen, 1969; Graham and Argyle, 1975; Okada and Brosnahan, 1990).

Most studies that have investigated the relationship between speakers' global pitch range and gestures have not find any statistical significance between them. Hoetjes et al. (2014) did not show any effect of increased cognitive load (when subjects were not allowed to gesture) on the number of words produced, speech duration, speech rate, production of pauses, and pitch range. However, the number of attempts had as an effect the decrease of the number of total words, speech duration, and production of filled pauses. The results led to the conclusion that, with repetitions, speakers become more fluent and use fewer words; at the same time, though, pitch and speech rate did not change significantly. Instructing subjects to gesture, or alternatively not to gesture is a procedure that has often been used to investigate properties of the co-production of gestures and speech (Parrill et al. 2016). However, it is not clear if instructions to gesture can actually modify gesture rate or type of gestures used by speakers (Parrill et al. 2016).

The direct instruction to gesture (or not to gesture) may not be the best procedure to elicit variations in global pitch range or other speech characteristics employed by speakers. In fact,

since prosodic variation is normally used to convey paralinguistic features, a better way to elicit speakers' variations in pitch may be to ask speakers to be communicative in their speech, as is actually done in public speaking classes.

In this perspective, the first study presented in this work aims to investigate global pitch range and use of gestures as produced by Italian speakers of English L2 before and after they are asked to be communicative in their speech. The hypothesis is that, if there is a correlation between prosody and gestures, explicit instruction to be communicative may incite speakers to use a different pitch range together with a different way of gesturing.

To test this hypothesis, an experiment was designed to compare global pitch range and use of gestures in the L2 (English) production of 8 Italian speakers of English as a second language. Speech in Italian is also examined but is was collected only for comparison reasons. What is expected is that when speakers have the intention to be more communicative in the L2 they will use a more varied pitch range and more gestures. Other speech and gesture characteristics, as articulation and speech rates, speech disfluencies and type of gestures used by the speakers are also analyzed.

The experiment aimed to verify the existence of a correlation between subjects' speech suprasegmental features and gestures in a story-telling task, and precisely under the assumption that when the subjects repeat a task in a foreign language, this will be less cognitively complex, and be characterized by greater fluency, as shown through higher articulation rate and higher speech rate.

This experiment aimed to test three main hypotheses:

1. When performing a task in the L2 (English), the cognitive effort required by speaking the L2 will appear as a narrower pitch range, less fluent speech and increased gesturing in the L2 as compared to the L1;

2. If the subjects repeat the task in English, this repetition will decrease the cognitive effort required by the speakers to perform the task; this will show up as more fluent speech.

3. If speakers are instructed to be communicative, this will have an effect on their pitch and gestures when they repeat the task. The expected effect on the second repetition

in English is that it will cause both greater and more varied pitch and gesturing than in the first repetition. The possible variation in gesture type usage is to verify.

3. Subjects and Materials

8 native Italian speakers were recruited. All subjects were female, mean age 23.5, all Italian L1 speakers (from the North-East area of Italy) and students of a *Public Speaking* class, taught by professor M.G. Busà at the Department of Linguistic and Literary Studies (DISLL) at the University of Padova, Italy. The selected students were English L2 learners and had a B2 level of the CEFR, since this was the minimum requirement to attend the Public Speaking class. All the selected subjects came from the same regional area, Veneto. The students in the class who were not selected for the experiment participated as audience.

At the beginning of the experiment, all subjects were asked to read Aesop's fable "The Fox and the Crow" in English to themselves (see Appendix 1). They were then asked to tell the fable to their classmates, first in Italian and then, after about 30 minutes, in English. A week later, they told the story in English again, but this time they were urged to try to be as communicative as possible while telling the story. All students were naïve to the purpose of the experiment. The experiment took place at the beginning of the course (October) so that the students were not yet familiar with common techniques used in public speaking. The students were told that their video recordings would be used by the teacher to assess their initial competences in public speaking to be compared with videos that would be (and in fact were) taken at the end of the course. Both times, speakers were video-recorded by the teacher. Each recording lasted about 90-120 seconds.

4. Procedure

The subjects were asked to repeat the fable they read in front of their classmates. The room in which the experiment took place (Fig. 8) was the same room in which they were used to attend the *public speaking* classes.

Subjects were free to read the fable a few times. No time constraint was imposed on the reading and memorizing phase. On average, students read the fable twice or three times before they felt

sufficiently prepared. Once they felt comfortable with the content of the fable, they had to stand up in front of the rest of the classmates (who were not selected as subjects) and tell the story. The first time, no particular instruction was given to the subjects but that of telling the story. The subjects told the story first in one language and then in the other. Half of the subjects were asked to perform the task in L2-L1 order, whereas the rest did it in L1-L2 order.

The subjects stood in front of their audience who were all seated at their desks and were asked to listen carefully to the person who was talking. The camera was set at 3 meters distant, and in order to facilitate the annotation procedure and to keep the video material as neutral as possible, a whiteboard was placed right behind the subjects. The teacher, after turning on the recording camera, sat at the back of the class. This was done in order to avoid increasing the pressure subjects might feel being recorded by a person behind the camera. In the following figure (fig.8) a schematic representation of the experimental set-up is presented:

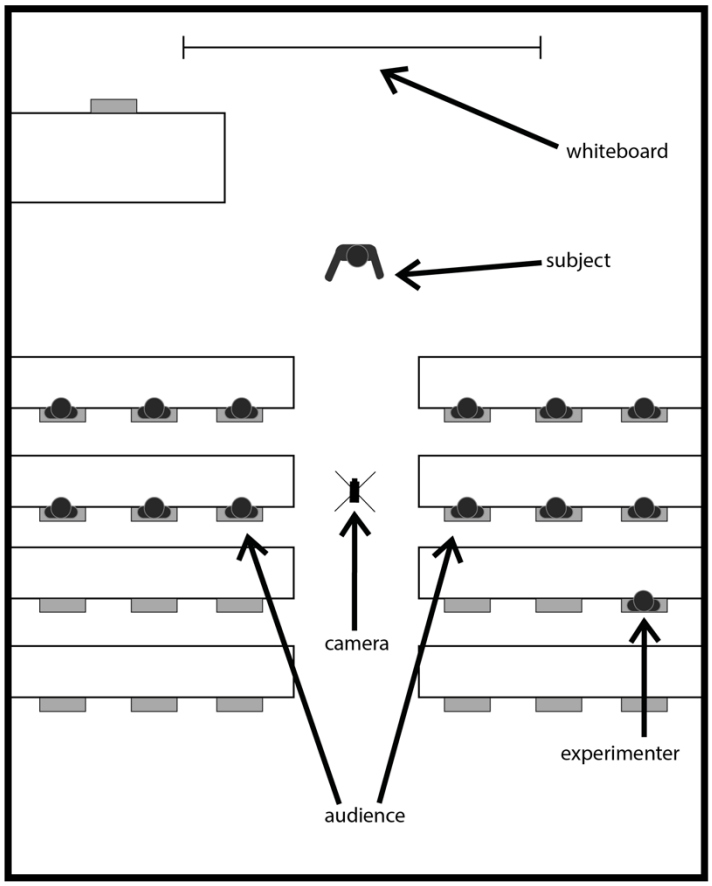


Fig. 8. Experimental set

The same set-up was kept on the second recording, about 7-10 days later, in which the subjects were asked to repeat the story in front of their audience. Again, the students were free to read the story as many times as they wanted before repeating the experiment. As for the first time, they all generally read the story twice or three times. This time they were asked to try to be as communicative as possible. The students were recorded following the same procedure as in their first performance, but this time they narrated the story only one time, in English.

5. Methods and Analysis

The corpus consists of three experimental conditions:

1. Italian, produced at time1, which will be referred to as *Italian t1*;
2. English, produced at time1, which will be referred to as *English t1*;
3. English, produced at time2, which will be referred to as *English t2*.

The analysis was carried out in three main phases: a first examination of the audiovisual material focusing on the subjects' gestural behavior; a prosodic annotation and analysis; a gesture annotation and analysis.

To be able to analyze the audio signal, it was necessary to extract the audio from the videos materials. To do so, the AVC freeware software (http://www.any-video-converter.com/) was used. The audio files were then imported into Praat (www.praat.org). Pitch was measured setting the pitch floor to 75 Hz, and the ceiling to 500 Hz, since all the speakers were female.

Using Praat textgrids, the following data were annotated:

a) *words*: the orthographic transcription of each pronounced word;

b) *respiratory pauses:* every time the subjects took a breath was annotated as PR;

c) *silent pauses,* i.e., all silences longer than 100ms in duration. Unlike respiratory pauses, silent pauses are usually not used for the physical need to inhale, instead they are used more or less voluntarily and often have a grammatical function (Zellner,1994). Silent pauses may also occur when speakers are unsure of what they are about to say. In this

investigation, silent pauses were considered when appearing during speeches in which the subject hesitates to say something, or when they need to access and retrieve some lexical item, especially in the case of L2 speakers. All these phenomena were annotated as PS;

d) *filled pauses*: all those vocalizations or vocal phenomena that have no semantic meaning such as 'uhm', 'ehm', 'mmm', and denote the length of the delay of upcoming speech (Clark & Fox Tree, 2002; Zellner, 1994) were annotated as filled pauses (PP 'pausa piena'); full pauses were excluded from the words count.

e) *disfluencies,* i.e., any vocal interruption of speech that does not add any propositional or semantic content to the utterance. Disfluencies were subcategorized in *repetitions* and *corrections*. *Repetitions* (annotated as REP) were considered as all those reiterations of the word/s the speaker just said. In spontaneous speech, repetitions generally involve a first instance of the repeated word, a possible silent pause, a second instance of the repeated word, and the continuation of the utterance (Rangarajan, & Shrikanth 2006; Tree, 1995). *Corrections* (annotated as CORR) were considered as all those verbal mistakes subjects did while delivering their speech. That is, every time the speakers stopped to correct a word they considered wrong or inappropriate for correcting it with the one they considered more appropriate. Corrections involve a wrong word, a possible silent pause and/or a repetition, followed by the correct word. For the purpose of the total word count and the calculation of Speech Rate, repetitions and corrections were counted as words.

The following measures were taken:

- Total duration of speech;
- Number and duration of pauses (silent and respiratory);
- Duration of articulated speech (Tot.Dur.Speech–Tot.Dur.Pauses);
- Articulation rate (n.words/duration articulated speech);
- Speech rate (n.words/total duration speech) (Pettorino 2003);
- Total number of words;
- For pitch range: F_0Min, F_0Max, F_0Mean, St.Dev. (Ladd 2008; Patterson 2000; Mennen et al. 2007, 2012).

6. Gesture data

The last phase of the data analysis was carried out using the multimodal annotation software Elan (https://tla.mpi.nl/tools/tla-tools/elan/ as in fig.9) through which each co-speech gesture in the three experimental conditions (Italian t1, English t1 and English t2) was annotated.

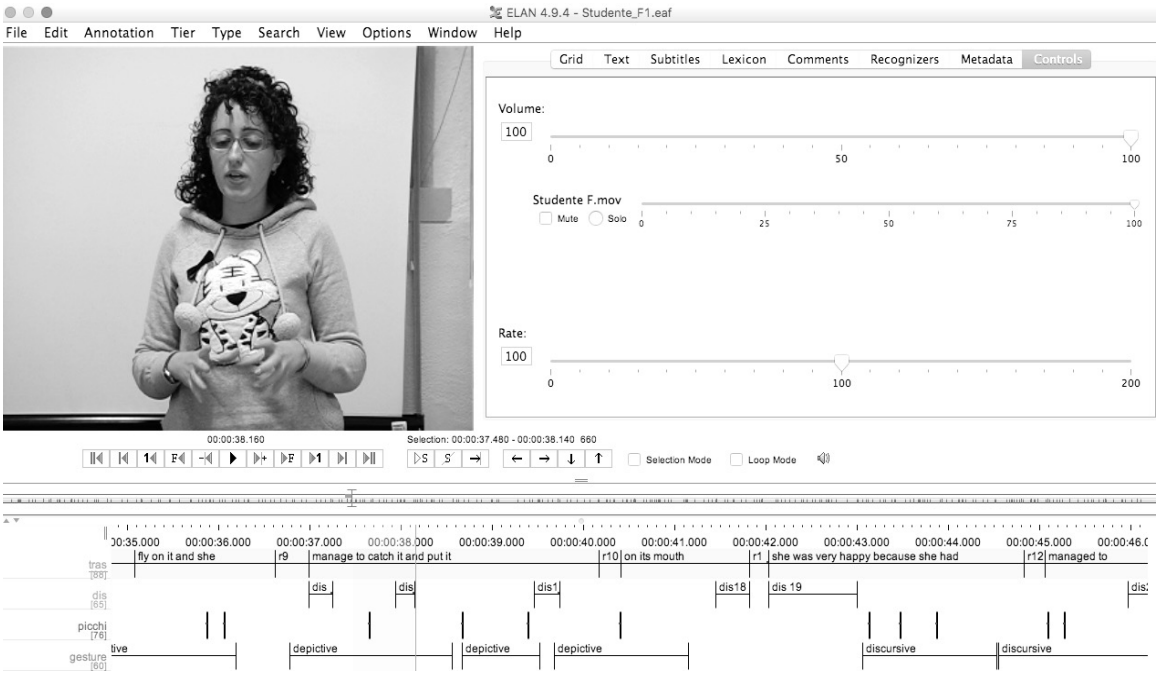


Fig. 9. Annotation in Elan, Speaker F is performing an iconic gesture in her first narration in English

In order to classify the co-speech gestures occurring in the narrations, the classificatory system purposed by David McNeill (1992) was chosen. McNeill considers gestures and speech as part of the same system, and gestures are considered to be the reflection of the common representation that gestures and speech share. Furthermore, McNeill's classification is probably the most used and it has been successfully applied to a wide range of studies. This allowed a better comparison with gestures studies. The categories that McNeill identified are:

- *Iconic*: that physically represent the object of speech. When a subject uses an iconic gesture for 'ball', their hand will probably resemble a rounded (bouncing) object. In the specific case of this experiment, an iconic gesture that often occurred was the subject describing "the piece of cheese" with one or both hands shaping a triangular object.

- *Metaphoric*: when hands depict some abstract concept with their hands. When talking about a feeling, for example, speakers can keep their hands in front of them as if they are holding something in their hands.
- *Deictic*: also called pointing gestures, are used to indicate something physical or abstract. The pointing gesture can occur with the index or an open hand indicating something. For the aim of this research, no distinction was made between these different hand configurations.
- *Beats*: rhythmic gestures that usually accompany speech but have no semantic meaning.

Fig. 10. Example of iconic gesture. The subject is talking about the crow stealing the cheese with its beak (see Appendix 1 for the whole story)

Deictic gestures were divided into two categories: *abstract pointing* and *concrete deictic.* The difference between these two is that whereas the first one is used by the speaker to point at something that is not actually present in the room, the latter is used to indicate present objects.

To annotate the gestures, they were divided into three main categories: *representational, discursive* and *emblems.* The first type, *representational*, included iconic, metaphorical and deictic gestures. The second type, *discursive*, included beats, and, in addition to the categorization of gestures made by McNeill, also a *vagueness and approximation gestures* category was included. As described by Vincze and Poggi (2013), these kinds of gestures "generally use a basic and easy handshape (open hand, curve fingers, generally no protruded fingers), and curve movement trajectory; they are generally repeated, they have the shape of a circle or a cyclic form, and involve movements of outward rotation (as opposed to the oscillation of approximation gestures), with low muscular tension and high fluidity. They are sometimes accompanied by eyes looking upward or sideways, typical of someone who has not yet found the right concept, or by a grimace with lips lowered" (Vincze & Poggi, 2013). More specifically,

Vincze & Poggi (2013) distinguished these gestures into four types: *vagueness*, characterized by a "fuzzy-round" movement; *approximation*, in which the hand marks oscillatory movements of the head and the hands; *word search* described as a rotating movement; and *hastiness gestures,* which they describe also as "jerky". Generally, *vagueness* and *word search* share the rotating movement, but whereas in vagueness the rotation is looser, it becomes rapid when it is employed for *word-searching*.

A similar distinction was made also by Ladewig (2011) who distinguished and described *cyclic* gestures, as recurrent gestures characterized by a "continuous circular movement of the hand, performed away from the body, i.e. outward. The hand remains in situ, i.e., it is not moved forwards or sidewards during the performance of the circular motion" (Ladewig, 2011). According to Ladewig (2011), the main function of *cyclic* gestures is word or concept retrieval, but it is also used to accompany descriptions, requests, and sometimes while illustrating lists or enumerations. For the aim of this study, all the distinctions made for these kinds of rotating and cyclic movements were not considered and all the gestures that fulfilled the descriptions above were considered part of the *vagueness and approximation gesture,* and thus, of the *discursive* category.

One category was annotated separately: *emblems*, which are those gestures that are culturally shared and whose meaning is usually understood only by the components of a certain social group (Kendon, 1980, 2004). Emblems are highly coded gestures whose meaning cannot be understood by everyone and that can occur independently from speech. This makes them a category on their own. The present task was not designed to elicit this kind of gestures but, since they are highly cultural, they could possibly appear anyway. An emblem that often occurred, for instance, is the Italian emblem that means "yesterday" or "time ago", it is performed by folding the arm and moving the hand as if throwing something behind the shoulder (Caon, 2010). This kind of gesture was performed by the speakers both when speaking Italian and English, at the beginning of the story, while saying sentences as "once upon a time".

In the following table, all the category distinctions are summarized:

MAIN CATEGORIES	TYPES OF GESTURES
Representational	Iconic
	Metaphoric
	Abstract Pointing
Discursive	Pointing
	Vagueness
	Beat
	Other (cyclic/ other idiosyncratic gestures)
Emblems	Emblems

Tab. 5 gesture macro-categorization

Speakers' gestures were annotated, counted and distinguished by the macro-categories just illustrated. Since the aim was to investigate whether representational gestures have a communicative function, the gestures were directly assigned to the macro-category they belonged to, (i.e. representational, discursive, emblems) as showed in tab.5. After counting the number of gestures occurring for each category, gesture rate was calculated for each language (Italian and English) and the time in which the story was told (time 1, time 2) with the formula Total N.gestures/Total N.words * 100 (Nicoladis et al., 2007). In this way, it was possible to relate gesture rate with speech rate, as well as to allow a comparison of the number of gestures in the first and second repetition.

6.1 The gesturing styles: some examples from the corpora

Since this experiment was mainly focused on the representationality of the gestures and their role when the subjects increase their communicative effort, the different types of gestures were divided into three main categories (representational, discursive and emblems). During the annotation phase of this investigation, the gestures, as said before, were assigned to one of the three macro-categories. It is important, though, to mention that each subject typically have their own gesticulation style and that this changed according to their personality traits, sense of confidence and feelings. The three different corpora (Italian t1, English t1, English t2) that were

analyzed for this experiment include three different times in which each speaker told the story. As for the representational macro-category, in each experimental condition there had been a few representational gestures that happened to occur more often than others. This occurred mainly with those gestures that matched with those that were perceived by the subjects as the most salient points of the fable, these can be summarized in the following list (for the whole fable, see Appendix 1):

a) the crow was flying around
b) it picks up the piece of cheese
c) the window sill
d) it went on the top of the tree
e) a fox arrived
f) the fox saw the cheese (and asked the crow to sing)
g) the crow opened it beak
h) the cheese fell down
i) the fox caught the cheese (and walked away)

In general, the subjects did not show significant changes their gesturing style from the first to the second narration in English and neither they presented any particular different style when speaking their native language. A difference that was perceived, though, was the fact that in their second narration the gestures they used seemed to be slightly wider (with a possible use of a greater gesture space) and with more precise movements. This could be explained with the gain of confidence and greater memorization of the tale, features that can have caused a lighter cognitive load and less nervousness in front of the camera.

The first within the most common representational gestures made by the subjects is *a) the crow was flying around.*

As in the example (fig. 11), in her performances in English, Student C produced a co-speech gesture while she was speaking about the crow flying while looking for food. In her first attempt, while saying "[the crow] flew around[5] looking…", the speaker used a very rapid movement from the upper left periphery to the center of the gesture space, she kept the hand open conveying the

[5] The underlined words correspond to the position of the gesture in the utterance

meaning of the crow gliding down to reach the food. In her second attempt, the second repetition,

Fig. 11 Student C in her first (on the left) and second telling of the fable in English

the subject uses a different type of gesture to convey the same sense. In fact, even if the co-occurring speech does not change much from the first time she told the story (this time she said "flying around looking for food"), her gesture seems to be wider and longer: the subjects rises her hand in the upper left space while her hand draws some circles in the air for a couple of times.

Another example of representational gesturing style is the one reported in fig. 12. One of the salient points of the story was the fact that after spotting it, the crow picks a piece of cheese in its beak. A clear example of the gesture that the subjects performed is the one showed below:

English 1 *English 2* *Italian*

Fig. 12 Student B while conveying the sentence "it picks up the cheese" in the three experimental conditions

In this example, it is possible to notice one only difference in the gesturing style of this speaker. In fact, she actually seems to perform the exact same gesture in both her first telling in English

and her telling in Italian: with her left hand in the center-center gesture space, she mimics the beak of the bird while grabbing something small (the cheese). In her second performance in English, though, her gesturing style changes: she does not use the left hand anymore, but rather prefers to use her right hand and to use a completely different gesture space: this time she uses a wider movement that goes from the upper right space, down to the lower right space of gesturing. This kind of changes in the gesturing style from the first and the second performances in English happened in other occasions, with other gestures and other participants too. These results might let room for speculations about the way speaker perceive their communicative style. Since the gesturing style did not change much within the first time they told the story in English and Italian in which the subjects were only asked to repeat the story, what could have made them change their actual representational gesturing style might have been the *communicative* instruction that was given to them before performing their second task in English.

For what concerns those gestures that were annotated as part of the *discursive* category, the gestures that more often happened to occur were those that rhythmically accompanied the telling of the fable and the cyclic and vague gestures that often occurred together with hesitancies, disfluencies and approximations:

Sp.F- English 1 *Sp.E - English 2* *Sp.B - Italian*

Fig. 13 Beat Gestures (annotated discursive) in the three different conditions

Beat gestures (in fig.13), are those repeated movements without stroke phase that typically accompany the rhythm of speech (McNeill, 1992, 2005). In this investigation, these kind of gestures were considered and annotated as part of the discursive macro-category. In the examples in fig. 13 were reported a few of these gestures.

In the first image in fig. 13, speaker F is performing a beat gesture right after a hesitation. The co-occurring speech in fact is: “it is [disfluency] an animal” where the beat gesture underlines the word *animal.* Instead, in the second picture, speaker E is marking every single word in the phrase “so he starts singing” by hitting her left hand on the right one, more precisely on the tips of her right fingers. In the third picture, student B is tracing an oblique line up and down on the bottom part of her gesture space while she pronounces the words “la volpe inizia a lusingarlo” (translated: the fox starts to flatter him). Her hands, in this case, underline the word *lusingarlo.*

Other examples of gestures that were annotated as part of the *discursive* category are the cyclic (Ladewig, 2011) and the vagueness and approximation gestures (Vincze & Poggi, 2013).

Sp. E - English 1

Sp. D -English 1

Fig. 14 Cyclic, vague and other pragmatic gestures

The two examples reported above (fig. 14) are taken from the performances of two students that were both engaged in their first attempts in English. Similar gestures occurred also in the second repetition in English and in the attempt in Italian, though, they happened more frequently in English (at time 1). In the first picture, Student E is performing a cyclic gesture in the center-center area of her gesture space while saying “because she is a [pause] clever animal”. By producing such a gesture Student E conveys an approximation, as if she was not completely satisfied with the word she chose but that probably that word is the one that gets closer to the meaning she wants to convey. The picture on the right of fig. 14, instead, shows a gesture that is characterized by more complex movements. In fact, the hands of Student D are both engaged in two different actions: with her right hand she produces some small irregular movements from the center-center space to the lower right periphery of her gesture space, meanwhile she performs a sort of a beat movement with her left hand which she keeps open and rigid, while making a very

little movement from the center to the left. In this case, Speaker D was very nervous and shy while she was pronouncing the phrase “fly down to [pause] catch the cheese” so probably her cognitive load was very high and she produced the gesture of vagueness as a sort of lexical retrieval tool. In summary, all these kind of gestures often carry a clear pragmatic or meta-narrative meaning but were considered part of the discursive category because they did not carry any representational feature linked to the object of the discourse.

The last distinction that was made in order to annotate the gesture occurrences of this experiments is the one dedicated to *emblem* gestures. The present task was not designed to elicit this kind of gestures but, since they are highly cultural, they could appear anyway. An emblem that often occurred, for instance, is the Italian emblem that means “yesterday” or “time ago”, it is performed folding the arm and rotating the hand as if throwing something behind the shoulder (Caon, 2010). This kind of gesture was performed by the speakers both when speaking Italian and English, at the beginning of the story, while saying sentences as “once upon a time”.

Sp. B English 2 *Sp. F English 2*

Fig. 15 Emblem occurrence. In this case, both Student B and F are performing the same emblem "once upon a time"

As shown in fig. 15, the emblem gesture that was used the most by all the subjects was the one co-occurring with the phrase “once upon a time there was a crow”. It actually occurred in every condition, but it occurred the most in the English t2 condition when subjects were asked to perform a more communicative narration. The use of emblems was not statistically significant and this experiment was not designed to elicit this gesture category. Yet, these results give an input to a further investigation about the use of emblems in story-telling.

7. Results

7.1 About Pitch

Tab. 6 shows the results of the pitch values for the three experimental conditions, averaged across the subjects. Whereas in Italian t1 and English t1 the values are similar, subjects show generally higher values in English t2. Higher Mean pitch, higher Max pitch, and higher standard deviation indicate that the speakers performed their narration with a much more varied F_0 when speaking for the second time in English than in the other two conditions.

	Italian t1	**English t1**	**English t2**
Mean pitch	201.0	199.0	228.2
Min pitch	148.5	149.6	146
Max pitch	445.	432.1	475.4
ST DEV	42.0	42.6	57.1

Tab. 6. Pitch values for the three experimental conditions

In order to evaluate the differences in pitch variation between the three experimental conditions, the procedure followed Hincks (2004, 2005). Hincks measured a value of pitch variation which she called Pitch Variation Quotient (PVQ) which is obtained by normalizing F_0 dividing the standard deviation of pitch by the mean.

In order to calculate the PVQ, following a procedure indicated in Hincks (2004, 2005), pitch listings were extracted from each audio file. First of all, it was necessary to remove all the outliers, which may be the result of external noises or other disturbances. In fact, the software Praat gets a list of every pitch value in the audio files, but some of them can be the result of background noises or other features that actually are not produced by human voices. Pitch listings were checked and all the possible outliers were consequently removed.

After mean pitch and standard deviation were calculated, the data were normalized dividing the standard deviation of F_0 by the mean (PVQ = St.Dev.F_0/F_0Mean). Unless specified otherwise, the statistical significance of the results was tested with within-subject ANOVAs.

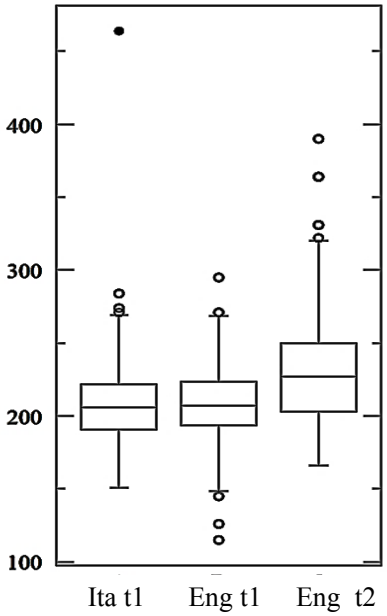


Fig. 16. Anova results for PVQ values. The second repetition in English results as significantly more varied

Fig. 16 shows the PVQ data for the eight speakers in the tree conditions. The data confirms what is shown in Table 1, that is, all speakers varied their pitch more in English t2 than in English t1 or Italian t1. For all speakers the PVQ of Italian is comparable to the PVQ of English t1, showing that the speakers did not use a very varied pitch in Italian and in their first repetition in English; the speakers' pitch is more varied when they repeat the story in English the second time. The difference in PVQ values in the three experimental conditions was highly significant: $F(2, 779) = 49.96$, $p < 0.01$ at a between-subject ANOVA test. Furthermore, the results of a t-test confirmed that the condition that is significantly different from the others is indeed English t2 (as shown in Fig.11).

7.2 Speech data

Table 7 shows the data relating to the characteristics of subjects' speech in the three conditions.

	Ital. t1	**English t1**	**English t2**
Total dur. speech (ms)	84.93	96.23	92.39
Total n. pauses (sil, filled, resp.)	23.25	35.75	37.5
Total dur. of pauses (sil, filled, resp.)	9.14	12.86	13.33
Total dur. articulated speech (ms)	75.78	83.36	79.05
N. of words	208.62	189.75	222.37
Articulation rate	2.75	2.27	2.81
Speech rate	2.45	1.97	2.40

Tab. 7. Speech data in the three experimental conditions

The Italian t1 condition shows the shortest total duration, whereas English t1 has the longest duration, and English t2 has values between the other two datasets. Even if they can give a hint of the fact that the first time speakers have to narrate in English they probably have more difficulties, these differences do not result to be statistically significant.

The results show also that speakers make fewer pauses in Italian t1 than in English t1 and English t2. There is a significant difference between the condition in which speakers narrated the story in Italian and both the times they spoke in English (English t1 and English t2) ($F(2, 21) = 49.96$, $p < .0.01$). Instead, there is no significant difference between the English t1 and the English t2 conditions.

Speakers' pauses last significantly less in the Italian t1 condition than in both English t1 condition ($F(1. 7) = 14.189$, $p < 0.01$) and English t2 condition ($F(1. 7) = 10.732$, $p < 0.05$), whereas the difference in the duration of pauses between English t1 and English t2 is not significant. As expected, the results show that the subjects are more fluent in Italian than in either English t1 and English t2. Thus, the second repetition in English does not significantly change the subjects' level of speech fluency. This result could suggest that fluency does not benefit much from repetition. The significant difference between Italian t1 and English t1, though, is lower than the one between Italian t1 and English t2. This could mean that fluency in English improves with the second repetition, getting closer to the fluency speakers have when they narrate in their native language.

Analyzing the number of words, it is evident that the English t1 condition shows the smallest number of words, English t2 has the greatest number of words, whereas Italian t1 shows values in between. The difference in number of words is significant only in the case of the English t1 and English t2 datasets ($F(1.7) = 12.699$, $p < 0.01$).

To better show how speakers employ pauses in the three conditions, Figure 17 shows the relative ratio of pauses and spoken words in the three experimental conditions: In Italian t1 subjects produce longer speech before they make a pause; in English t1 speech between pauses is the

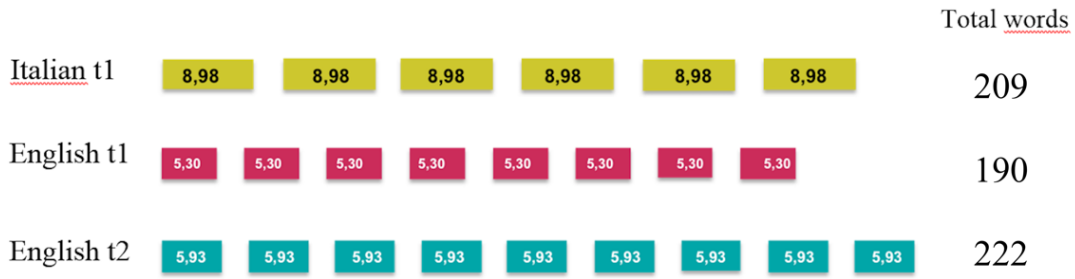


Fig. 17. Comparison of the frequency of co-occurrence of pauses in relation to spoken words in the three datasets

shortest; in English t2 the speech is slightly longer but still shorter than in Italian. When narrating in Italian, in fact, speakers employ on average almost 9 words before making pauses of any kind. This does not occur when they deliver their speech in English. In fact, whereas the first time they produce on average 5.3 words before pausing, the second time they pronounce on average almost 6 words (and even produce a higher number of words overall). The subjects' fluency improves with repetition but, at least in this case, the improvement is not enough to reach the native language values. Some differences within subjects were found also in the duration of articulated time. The data are not statistically significant but they show a trend, since they once again show the effect that repetition and the communicative intent have on subjects' speech. Articulation time is the time speakers use to actually produce the sounds through the movements (i.e. of the tongue, lips, jaw, vocal chords and all the speech organs) that allow them to deliver the words they intend to say. Articulated speech does not include any kind of silence or pause and thus it is given by the total duration of speech minus the total duration of pauses (and it is calculated in this thesis with: Tot.Dur.Speech–Tot.Dur.Pauses). The results show that in the Italian t1 condition, the speakers used less time than both the times they narrated the story in English. However, the time they used to tell the story decreased in their second repetition. In fact, the average articulated time in English t1 is 83.36'', and 79.05'' in English t2. Again, the values in English t2 seem to get closer to those showed in the speakers' native language (that was 75.78'') and, thus, these data suggest that the repetition and communicative intent might have a positive influence on second language even if between the first and the second experiment there is a time gap of one week.

Two important values for the identification of differences in speech are the articulation rate and the speech rate. Speech and Articulation rates are two prosodic features that are defined as "the number of output units per unit of time" (Tsao, Weismer & Iqbal, 2006) but it is commonly recognized that whereas Articulation rate does not consider speakers' hesitations and pauses in determining the pace at which segments are produced, Speech rate actually does include everything is said or communicated by the speakers and thus includes features as hesitations, pauses, fillers, and so on.

Both articulation and speech rates were calculated. The articulation rate results show lower values in English t1 than both Italian t1 and English t2. When speakers narrate in English for the first time their articulation rate is 2.27, rather lower than their articulation rate in Italian t1 (which was 2.75) and this difference, in fact, is significant ($F(1.7) = 6.413$, $p < 0.05$). When speakers narrate the fable for their second time in English, their articulation rates increases and it has an average value of 2.81 The difference between English t1 and English t2 is highly significant ($F(1.7) = 15.48$, $p < 0.01$). Since the values of English t2 get higher, there is less difference with the Italian t1 values and so it is not significant. Anyhow, even if their difference is not statistically significant, the articulation rate of English t2 shows values that are even slightly higher (2.81) than those presented in Italian t1 (2.75), and this could be explained as the result of both repetition and the higher communicative intent of the speakers.

Similarly, for speech rate, English t1 has lower values than both Italian t1 and English t2. The first time the subjects spoke in English they told the stories with a speech rate that had a value of 1.97, showing that their fluency was really low compared to the one they had when speaking in their native language (2.45). In English t1, speakers had a very low speech rate because the cognitive difficulty of the task was higher and it caused many filled pauses and hesitations (in fact, their articulation rate was 2.27). When delivering their speech in English for the second time, their average speech rate value reached 2.40, and got closer to the native language one of 2.45. Thus, there is a significant difference between Italian t1 and English t1 ($F(1.7) = 10.269$, $p < 0.05$), and between English t1 and English t2 ($F(1.7) = 12.436$, $p < 0.01$), while the difference between Italian t1 and English t2 is not significant.

In conclusion, the reported data about speech show that in the first repetition in English the subjects speak at significantly slower articulation and speech rates than in their native language, but with the second repetition in English they reach values that are not too dissimilar from the

native ones. When they speak for the first time in their L2, subjects are confronting a cognitive load that is higher than in their L1: they not only have to recall and structure the story in their minds, but they also need more time to retrieve the right words and to pronounce them in a language which is not theirs. One week later, when subjects were called to repeat the experiment, they were allowed to read the story again (in English) and, they were asked to be as much communicative as they could. This time their fluency increased, they had higher values of speech and articulation rates, which mean they had less hesitations, made less pauses and mistakes, but also that they were able to speak faster. Generally, the second time they spoke in English, their speech values got closer to those subjects had when speaking Italian. This means that, in their second repetition, speakers' cognitive load was reduced thanks to the repetition (they remembered the story and did not need to retrieve the words as much as in their first performance) and this probably let them concentrate on being more communicative. In fact, subjects did not just reduce their mistakes and hesitations (higher speech rate) but were also able to increase the number of words and speak faster (higher articulation rate).

7.3 Occurrence of gesture types

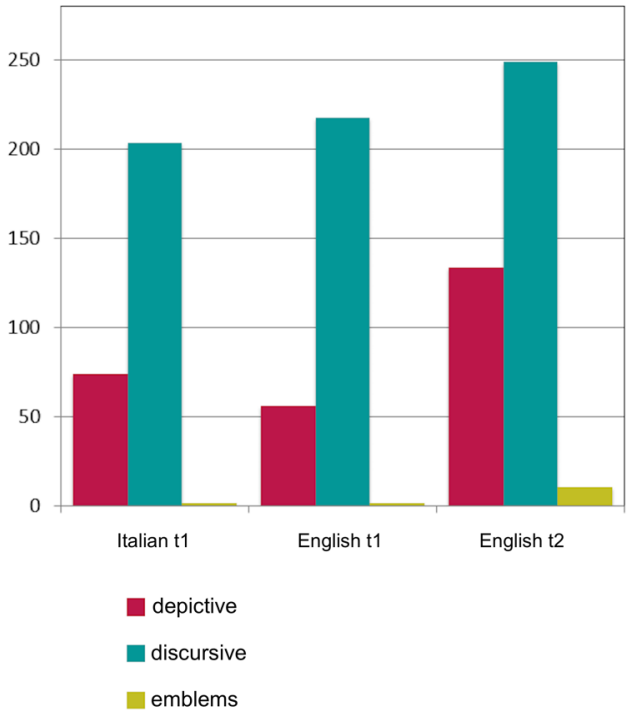


Fig. 18 gesture occurrences (in raw numbers) for every macro category

Gestures were counted as explained in §5. The comparison of gesture occurrence is visible in fig. 18. For each experimental condition, the three macro-categories are presented.

The turquoise columns show the overall occurrence of *discursive* gestures for each experimental condition. When they spoke Italian, their mother language, subjects used less discursive gestures than they did in the other two conditions. Indeed, though, the experimental condition in which subject used discursive gestures the most is the English t2, the one in which they were asked to increase their communicative intent. The English t1 condition presents values between the other two, but closer to those occurring in the subjects' native language. Statistically, though, the Italian t1 condition does not differ significantly from either the English t1 or the English t2 ones, nor is there a significant difference between the English t1 and the English t2 experimental conditions.

The reddish columns show data referring to the representational macro-category. In general, the experimental condition in which *representational* gestures were used less is the English t1 one, whereas the English t2 condition presents the highest number of *representational* gestures. Surprisingly, in the Italian t1 condition, subjects used more representational gesture than they did in their first performance in English. Since the cognitive load is higher in the L2 task, a relevant higher number of *representational* gestures was expected. Anyway, it is not possible to ignore the cognitive difficulty subject may have ran into and that it is not due to the foreign language: they were talking in front of a camera, recalling a story they had just read. Thus, the similarity between the occurrence of representational gestures in these two experimental conditions might be just an aleatory incidence, since the statistical analysis shows no significant difference between these two conditions. Instead, a statistical difference was found in the comparison between Italian t1($F(1.7) = 8.607$, $p < 0.05$) and between English t2 and, even more, between English t1 and English t2 conditions ($F(1.7) = 17.775$, $p < 0.01$). These results suggest that, with a lower cognitive effort and instructed to be more communicative, speakers tend to use a higher number of representational gestures: in order to help their audience to visualize the content of their narrations, they describe the scene they talk about through the use of more iconic and metaphorical gestures.

The lime-yellow columns are the ones referred to *emblems.* In these three specific conditions, subjects did not employ a significant number of emblems. Anyway, there are a few considerations that are worth to mention about the use of this specific kind of gesture. In fact, the emblem

gestures were sometimes used as a communication strategy in the second experimental condition of English (English t2). Among the 8 subjects, only a couple of them actually used emblems as a communicative strategy, and increased the number of emblems in their second repetition of English. One of them (Student F), for example, did not use any emblem in Italian t1 nor in English t1 but instead used 7 emblems in the English t2 condition.

In conclusion, comparing the gestures occurrences from a global perspective, it is clear that there is an overall higher number of gestures in the English t2 than in the other two experimental conditions. The students, to be more communicative and, thus, to help their recipients to better understand and visualize the content of the story, probably increased their overall gesticulation. All three categories, in fact, show an increased number of gestures. Out of all three categories, though, the one that increases the most is the *representational* one. This result strongly suggests that, when asked to increase their communicativeness, subjects employ a more "visual" strategy by incrementing the iconicity of their gestures. This kind of strategy is very similar to the one found by Campisi and Özyürek (2013) on adults that had to explain to 12 years old children, novice adults and adults how to make coffee. Campisi and Özyürek found that, when adults had to explain to children the usage of the coffee machine and the procedure to make coffee, they increased the rate of iconic gestures. In the present experiment, instead, the subjects were talking to their peers (their classmates) but the urge to result communicative probably modified their way of gesturing in a way similar to the strategy people use with children.

7.4 Relation of pitch and gestures

One of the hypotheses of this experiment was that if speakers are instructed to be communicative, their output will show effect on both their pitch and gestures at the repetition of the English task (English t2). To understand if the expectations were positively confirmed, it is necessary to compare the PVQ value and the Gesture Rate in the three experimental datasets.

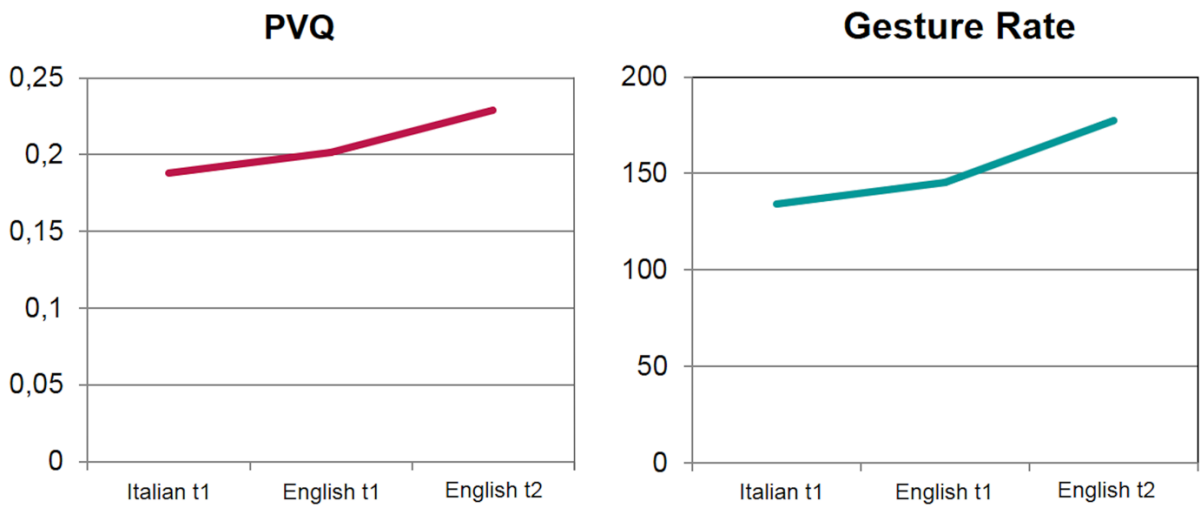


Fig. 19. Comparison between the pitch variation and the Gesture Rate

Figure 19 shows the comparison between the pitch variation and the Gesture Rate values for the three datasets.

On the left, the figure shows data referring to Pitch Variation Quotient. When the subjects told the tale in Italian and in English for the first time, they had a similar PVQ, even if English, compared to Italian, is commonly known as a language characterized by higher overall pitch range (Busà, 2010; Urbani 2013). L2 speakers, though, show narrower pitch ranges than L1 English speakers (Busà & Urbani, 2011) and this is well represented by the data referring to the English t1 condition. In English t2, instead, the students show higher pitch variation and have higher values of PVQ.

A similar pattern is presented in the graphic on the left. When speaking Italian, gesture rate (which measures the relation between gestures and word pronounced) is lower than in English t1, but the difference is not significant. The significance is reached when comparing the two different English datasets. In fact, the only data that is significant is the difference in the gesture rate of representational gestures in English t1 and English t2 ($F(1,7) = 11.946$, $p < 0.05$). This means that, the second time they are telling the story, the subjects employ a wider number of representational gestures than the first time and that representationals are the only category to significantly differ in the two datasets.

8. Discussion

This experiment was designed in order to examine the productions of Italian speakers of English L2 who were asked to read Aesop's fable "The Fox and the Crow" and to tell its content in Italian and then twice in English at two different times. The main object of investigation was the global pitch range, speech characteristics and gestures of the speakers' productions and, above all, the effects that the instruction to be communicative may have on subjects. Urging subjects to be communicative might in fact influence and/or modify the way they tell the story and gesticulate during their narrations.

The results show that the subjects used very similar pitch range values in their first telling of the story in Italian and English. In this case, this result suggests that speakers transfer their pitch values from Italian to English, and does not provide support to previous findings which claim that L2 speech may be characterized by limited pitch variation and a narrower pitch range than L1 speech (Aoyama & Guion, 2007; Busà & Urbani, 2011; Hincks, 2004, 2005; Mennen et. al. 2007, 2012; Pickering, 2004; Traunmüller & Eriksson, 1995; Ullakonoja, 2007). Differences in pitch range between L1 and L2 have been explained as due to the fact that native and non-native speakers process prosodic information differently because of their different levels of competence in the L1/L2. For example, non-native speakers, instead that on prosody, may rely more on segmental information to get their meanings across, since they lack the amount of extra-linguistic knowledge that native speakers can rely on when communicating (as proposed by Jenkins, 2002). It is possible, however, that differences in pitch range between the L1 and the L2 may be more evident when non-native speakers are focusing on getting their meanings across (for example in formal presentations or public speaking tasks) than on the way in which they deliver the content. Thus, the second time the subjects were called to tell the fable in English (experimental condition English t2), with the instruction to be as communicative as possible, may have caused them to use a significantly more varied, wider, pitch.

The data about fluency and speech characteristics of the narrations show that the subjects, as expected, are more fluent in Italian than in English. In fact, Italian t1 showed the highest speech rate value. The English t1 condition shows the lowest values, suggesting that the speakers' cognitive difficulty was high and it caused them to produce many filled pauses and hesitations. The English t2 condition, instead, showed higher speech rates. In the English t2 condition, subjects improve their speech rate, hence they improve their fluency. Yet, they continue to

produce speech that has more and longer pauses than in Italian. In fact, in Italian t1 students produced longer stretches of speech before they made a pause, whereas the speech between pauses in the English t2 condition, even if longer, is still shorter than in Italian.

In their second repetition in English (English t2), however, subjects reached values that are not too distant from the native ones. In English t1, subjects deal with a higher cognitive load than in their native language: besides having to remember and structure the story in their minds, they also need to retrieve the right lexicon and to pronounce words in a foreign language. Thus, the subjects articulate and speak more slowly in the first repetition in English than in Italian, possibly due to cognitive and linguistic problems with the task in the L2. In English t2, a week later, the subjects' fluency increases, they have higher speech and articulation rates, which means they have less hesitations and are able to speak faster, even if the way they produce pauses did not change significantly. They manage to reach articulation and speech rate similar to those of the native language, showing a significant improvement on their first repetition, and this might be due to the decreased cognitive effort required for the task. This probably let them focus on being more communicative. In fact, subjects did not just reduce their mistakes and hesitations (higher speech rate) but were also able to increase the number of words and speak faster (higher articulation rate).

The increased communicative effort that subjects employed in English t2 can be also seen through the number of words that they used. The subjects, in fact, produce a higher number of words in the second repetition in English than in either of the other two experimental conditions. It is known that, because of the *common ground* effect (Clark, 1996), when speakers make reference to information that is shared with the interlocutor they tend to use fewer words and less informative utterances. Thus, the expectation was to find a lower number of words in the second repetition in English than in the first. This different outcome might be interpreted as the result of the speakers' attempt to be more communicative: to better deliver their story, the subjects might have used more words in order to add details to the fable they were telling.

On the gesture side, in this experiment subjects appeared to gesture more in the second repetition in English than in the other two tasks. If the subjects in the present investigation had used gestures simply with a cognitive function, they would have used most gestures in English t1, when the task was most complex cognitively, and fewer gestures in English t2. Instead, the increase in gestures in the second repetition in English shows that the subjects are using gestures with a

communicative function. Thus, the students, to be more communicative and to help their recipients to better understand and visualize the content of the story, probably increased their overall gesticulation. All three gesture categories, in fact, show an increase in number. Out of the three categories, though, the one that increases the most and whose increase is statistically significant is the *representational* one. This result suggests that, when asked to focus on being communicative, subjects' strategy is to focus on the represenationality of their gestures. Interestingly, in the second repetition in English the speakers also use the most varied pitch range, confirming the hypothesis that speakers' communicative intentions cause an increase in both gestures and pitch range.

The present study shows that asking subjects to be more communicative in their speech is also effective in bringing about a change in number and type of gestures and in gesture rate. In this investigation, the increase in gestures concerns primarily representational gestures, confirming previous findings that L2 speakers use more iconic gestures with a communicative function, especially when the interlocutor is in sight (Bavelas et al. 2008; Cohen & Harrison, 1973). Whereas representational gestures are usually considered to have a self-oriented function (i.e., help speakers in their cognitive processes, as suggested by Alibali et al., 2017 and Kita, 2010, 2017) the present investigation shows an instance in which these gestures also have a communicative function, at least for L2 speakers.

Public speaking classes urge students that to be more communicative and to enhance listeners' comprehension they should vary their intonation and employ meaningful gestures. However, studies showing the actual effectiveness of this kind of instructions are lacking. Through this investigation it is shown that the repetition of a task can bring about substantial changes in the speakers' verbal and non-verbal patterns: besides modifying the way they use words, subjects can in fact change their speech and articulation rates, pitch range, gesture type and rate, and deliver a better performance.

The results of this study are relevant also to L2 instruction. Even though students' fluency may be related to speakers' proficiency and thus be hard to improve in the short period of a week's time, besides repetition, the explicit instruction to try to increase communicativeness can significantly lead to a change both at the verbal and non-verbal levels in the L2.

Finally, this study showed the importance of the type of instruction it is given to the subjects. Whereas previous studies (Parrill et al. 2016) did not disclose any significant result when the

experimenters asked to their subject to gesture more, the present study showed that, instead, providing students with instructions to be communicative can be particularly effective in bringing about the change.

6/ Does repetition help to be more communicative?

Experiment 2

1.Introduction

The results of the first experiments led to the conclusion that if a person is asked to be communicative, they will increase both their pitch variation and the total number of gestures produced, with a significant increase in representational gestures.

However, these results do not exclude the possible effect of other variables, first of all of task repetition. In fact, it is possible that the fact that the subjects repeated the story twice, which facilitated them in telling the story, let them concentrate on a livelier delivery of their speech.

Therefore, it was considered necessary to verify the effect of task repetition on pitch and gestures. Thus, an experiment was designed to verify if task repetition can cause speakers to use a more varied pitch range and a change in their use of gesturing.

Previous studies (Parril et al. 2016) showed that asking people to gesture did not have any significant effect on gesture elicitation. Since my previous experiment showed that when asked to be more communicative people actually gesture more and use more representational gestures, I decided to follow my investigation assessing what happens when the subjects are not given any specific instruction.

The first experiment, moreover, was designed with the purpose of understanding how subjects use gestures and pitch from a L2 perspective (L1 was considered only as control). In the second experiment the study was further expanded to L1 to test whether task repetition would still have some communicative effect in L1 and L2 when no instruction is given to the speakers. Speaking their L1, Italian, would present the subjects with less difficulties (less lexical retrieval effort, due to the greater memorization of the story and greater proficiency). This needed to be verified, since it was not part of the hypotheses tested in the previous experiment.

Thus, in the second experiment the following hypotheses were tested:

1) When telling the story in the L1, subjects will be more fluent in the second repetition, with less disfluencies and pauses. Since no specific instruction is given and they are simply asked to repeat the story in their native language, no significant differences on pitch and gestures are expected;
2) When subjects repeat the story in English, this repetition will decrease the cognitive effort on the subject and this will show up as better fluency than in the first repetition due to the greater ease of the task;
3) With no instruction except that of repeating the task, the subjects will not change their pitch and gestures patterns, since they do not aim to be more communicative.

The task the speakers were asked to perform was very similar to the one presented in the previous study, with a few exceptions that allowed for a better control of the variables. This time, speakers were asked to watch a short cartoon of the fable "The Fox and the Crow" and to tell it in front of a small audience while being video-recorded. The subjects were asked to repeat the task one week later in front of a different audience. They were given no instruction as to how they should communicate the story.

2. Subjects and Materials

10 native Italian students of the University of Padova took part in the experiment. All of them were L2 learners (B2 level of the CEFR). All subjects were female, mean age 24 and all coming from the same regional area: Veneto (North-East of Italy).

The experiment was divided into two sessions; the second session took place one week after the first one. In the first session subjects were asked to watch a cartoon representing the fable "The Fox and the Crow" by Aesop. The choice to change the kind of stimuli (written in experiment 1, visual in experiment 2) was done because a video could better help visualize the story. The video was adapted from a cartoon for children available online. In fact, the original video was taken by the YouTube channel for children *Pinkfong! Kids' Songs & Stories!* and originally lasted 2'32". The whole original version of the cartoon is available at this link: https://www.youtube.com/watch?v=vt3HP4VWuH0&t.

The video was manipulated and cut using the software Adobe Premiere Pro CC so that, except for a background music and onomatopoeic noises, it presented no other sound or dialogues. However, a brief written summary of what the subjects were going to see was added to the cartoon just before its beginning. The summaries were added in order to guide the subject and make them focus on some parts of the story. In this way, their story-telling would be more homogeneous. Two versions of the video were made: one having the summary of the story in Italian and the other showing it in English, both lasted around 1.50".

3.Experiment procedure

The experiment involved two main steps:

1. The subjects were exposed to the video in Italian and English. They were asked to watch it twice in order to memorize it better.
2. once they watched the video, they were asked to stand up and tell the story they just saw. The subject told the content of the story to an audience formed by two people who did not see the cartoon and did not previously know the story (see Fig. 20). The subjects were video-recorded during their speech.

After a short time, usually about half an hour, they repeated the two steps, and told the story in the other language (Italian or English).

The experiment was repeated one week later following the same steps. The subjects first watched the video twice, in the language they did not use the first time, and then told the story to an audience that they thought did not know the story. The audience was formed by two people, either confederates or other students, that changed each time the subject repeated the narration.

All subjects were naïve about the real purpose of the experiment: they all were told that the aim of the study was to verify mnemonic effects on storytelling.

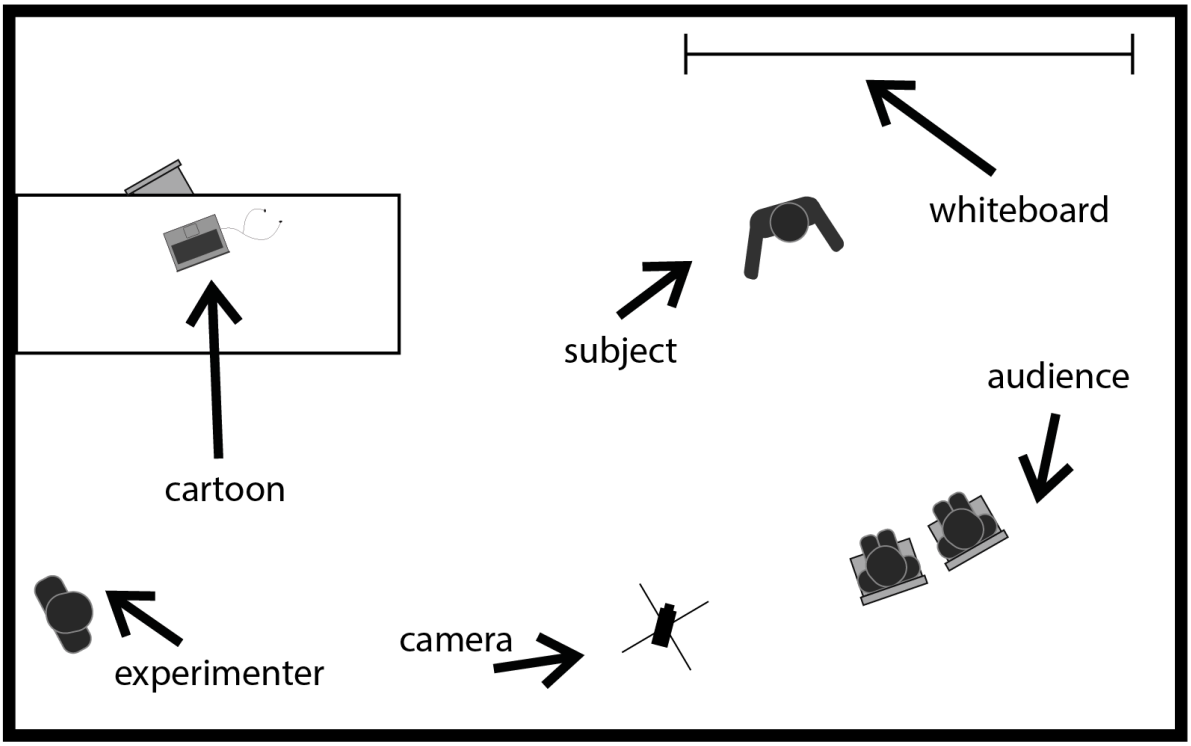


Fig. 20. Experimental set

In order to avoid the *common ground* effect (Holler & Wilkin 2000; Holler et al, 2011; Galati & Brennan, 2014) the audience was changed every repetition and for every language so that, in fact, every speaker told the story to four different audiences.

The corpus consists of four data sets:

- Italian – produced at time1, will be referred to as Ita t1
- English – produced at time1, will be referred to as Eng t1
- Italian – produced at time 2, will be referred to as Ita t2
- English – produced at time 2, will be referred to as Eng t2

The order in which the students told the story was not the same for everyone. Of all the ten subjects, while the first five started in Italian and concluded in English, the other five did the exact opposite. The week after, when they had to repeat the experiment, the groups switched the languages so that the first five now started with the English story and concluded with the Italian one.

4. Methods and Analysis

As for the first experiment, the analysis followed three main steps:

1) A visual examination of the audiovisual material, focused on the prosodic, gestural and verbal behavior of the ten subjects;
2) A prosodic and speech analysis made with the Praat software;
3) A gestural analysis made with the Elan software.

5. Prosodic and Speech data

The audio signal was extracted from the videos using the AVC software, which is freeware and it is downloadable at http://www.any-video-converter.com/, and then analyzed with Praat, software for speech and phonetic analysis.

As for the previous experiment, pitch was measured with the floor set at 75 Hz and the ceiling at 500Hz, since the speakers were all female. Pitch was analyzed through the extraction of the pitch listings with Praat. Once all pitch listings were extracted from Praat, they were ordered in a list from lowest to highest. Usually the distance between two pitch points is very small (for example three contiguous pitch points could be: 166.102 Hz; 166.112 Hz; 166,136 Hz…) but normally there are some outliers, that could be, for example, the result of external noise or microphones. These are generally very different from the contiguous pitch points. All the outliers were deleted manually.

Praat textgrids were used to annotate the following:

a) *Words:* the orthographic transcription of each pronounced word;
b) *Respiratory pauses;*
c) *Silent pauses:* as for the previous experiment, all silences longer than 100ms in duration were annotated as silent pauses;
d) *Filled pauses*: all those vocalizations or vocal phenomena such as ‘uhm’, ‘ehm’, ‘mmm’, that do not have a semantic meaning, but they are rather used to signal the length of the delay of upcoming speech (Clark & Fox Tree, 2002);

e) *Disfluencies:* as for the first experiment, all the vocal interruptions of speech that did not add any propositional or semantic content to the speech data were considered disfluencies. The sub-categorization was maintained, and *repetitions* and *corrections* were annotated with the same criteria than in the previous experiment: all the reiterations of the word/s the speaker just said, that were often interrupted by a brief silent pause, were considered repetitions, while all the verbal mistakes followed by a possible short silent pause and the correct word (or more appropriate word according to the speaker) were considered corrections. Both corrections and repetitions were counted as words in the total word count.

Considering all the data analyzed and annotated with Praat, the following measures were taken:

- Total duration of speech;
- Number and duration of pauses (considering both silent and respiratory pauses);
- Duration of articulated speech, which is the total duration of speech without the pauses
- Articulation rate (n.words/duration of articulated speech);
- Speech rate (n.words/total duration of speech – following Pettorino 2003);
- Total number of words;
- Pitch range: F_0Min, F_0Max, F_0Mean, St.Dev
- Pitch Variation Quotient (PVQ): following Hincks (2005), after the extraction of the pitch listing from each audio file, data was normalized dividing the standard deviation of F_0 by the mean.

6. Gesture data

The videos of the four datasets were analyzed with the software Elan (as in fig. 21 below) which is free and downloadable at https://tla.mpi.nl/tools/tla-tools/elan/.

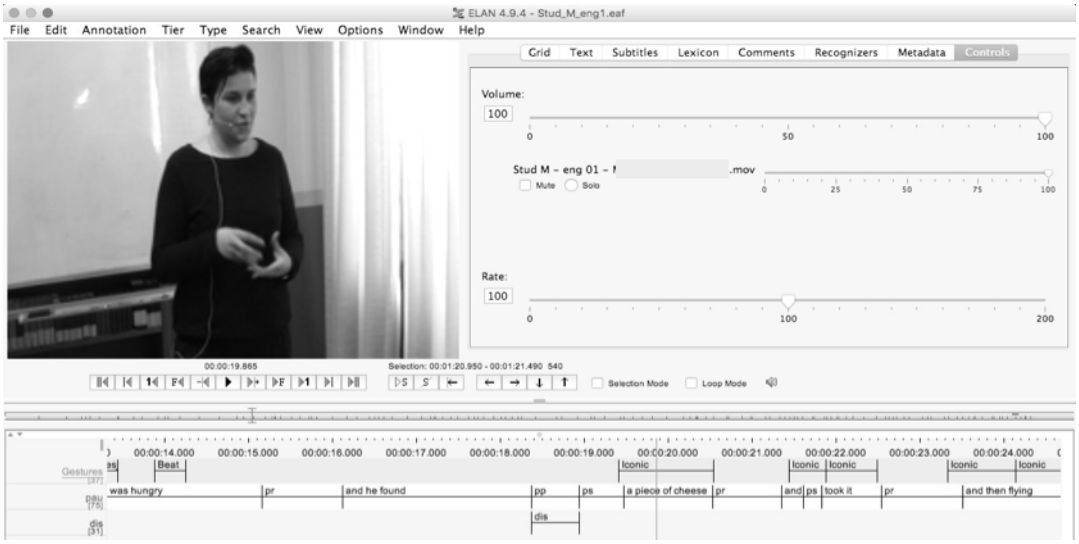


Fig. 21. Example of Elan annotation

To be able to compare the results with those obtained in the first experiment, the gestures were annotated with the same criteria, divided into three main categories:

1. *Representational* gestures: including iconic and metaphorical gestures. All those gestures that physically shape an object, action or even an abstract idea. For example, when talking about the crow dropping the piece of cheese, subjects easily made an iconic gesture by mimicking the cheese while falling down and tracing its path to the floor. At the same time, when speaking about an idea, a subject could make a metaphoric gesture by keeping their hands as if they were holding or manipulating a metaphoric object that actually represents the abstract concept they are talking about (McNeill, 1992, 2005; Kendon, 2004).
2. *Discursive* gestures: including beats, vagueness and all those pragmatic gestures that do not carry any clear iconic or metaphorical meaning.
3. *Emblems:* those gestures which meaning is shared within a certain culture. This kind of gesture was annotated in order to verify possible gestural transfers that could occur when switching from one language to another.

Whereas in the first experiment the gesture annotation was directly done using the final categorization, and having no distinction between specific gesture types did not permit a further investigation about the actual use of each gestural category (for example, it is not possible to say in which percentage the *discursive* category is formed by beat gestures), in this study the annotation was done distinguishing more categories and collecting them into broader categories. In this way it can be possible to recognize differences also inside the main categories (for example it will be possible to understand whether the *vagueness* gestures behave differently in the second repetition). This (tab. 8) is how gesture categories were annotated and aggregated into broader categories:

MAIN CATEGORIES	TYPES OF GESTURES
Representational	Iconic
	Metaphoric
	Abstract Pointing
Discursive	Pointing
	Vagueness
	Beat
	Other (cohesives / not iconic)

Tab. 8 Gesture categorization

All those gestures that were not related to speech, like for example nervousness gestures or self-adjusting gestures (like scratching, adjusting their clothes, etc…) were not considered. Whereas in the previous experiment *emblems* were barely used, in the current experiment they did not occurred at all. For this reason, in this chapter, *emblems* won't be mentioned.

Gestures were finally counted and summed into the two main categories, i.e., representational and discursive. Gesture rate was then calculated with the same formula as in the previous experiment: total n. of gestures divided by the total number of words multiplied by 100 (Tot.Gestures/Tot.Words*100, following Nicoladis et al., 2007). In this way it will be possible to compare speech rate to gesture rate, as well as compare the gesture rates of the first repetition to the one of the second repetition in each language.

6.1 Gesturing styles and annotation criteria. Some examples from the corpora

One of the aims of this experiment was to verify the possible changes that repetition itself could cause on gestures. Whereas the first investigation presented in this thesis (chapter 5) aimed to analyze whether speakers who were asked to be more communicative changed their narrations (in both the prosodic and the gestural feature) here, the intent is to investigate the possible communicative effects that repetition have on the speaker's narrative deliveries. For these reasons, the categorization of gestures did not change: the main analyses were concluded on the three macro-categories that were defined for the previous experiment. One of the main differences between the two experiments is the type of stimuli. Whereas in the first experiment the subjects were asked only to read the fable, in the present investigation the subjects were asked to watch a cartoon twice (Appendix 4) and this might have guided, and thus influence, their gesturing. Anyhow, each student had their own gesturing and narrating style even if the stimuli was somehow less imaginative than in the first experiment. A few examples are reported below.

The first category considered in the annotations is the *representational* one, which is formed by three different types of gestures: iconic, metaphoric and abstract deictic. These three categories are associated by the fact they all represent or indicate a physical or an abstract object (or subject) of the narration.

St. G – English 1

St. G – English 2

St. C – Italian 1

Fig. 22 Students G and C performing iconic gestures

In the example above (fig.22) the iconic gesture for "piece of cheese" is shown. This specific gesture occurred very often and it was annotated as *iconic* because it carries a very close formal relationship with the semantic content of what it is being said by the speaker. In all the three

examples the speakers (Students G and C) were talking about the crow spotting a piece of cheese while flying or holding the piece of cheese in its beak. In the first and second pictures (in fig.22) from the left Student G is pronouncing the words "the piece of cheese[6]" in both her two English repetitions whereas in the third picture, Student C is saying "aveva il pezzo di formaggio nel becco" (translation: [it] had the piece of cheese in the beak) while performing her first narration in Italian. In these examples, the students all chose the same gesture space (see McNeill, 1992, explained in chapter 1, §3) which is their center-center gesture space.

The representational macro-category includes also other two kind of gestures: metaphoric and abstract deictic.

St. A – Italian 1 *St. A – Italian 2* *St. A – English 1*

Fig. 23 Student A while producing metaphoric and abstract deictic gestures

In the example above (fig. 23) are reported three examples of metaphoric and abstract deictic gestures produces by Student A in her narration in Italian (both times) and English (first time). In the first picture from the left, Student A is saying "il corvo, tutto inorgoglito dai complimenti" (translation: "the crow, very flattered by the compliments") and, while saying so, she performs a complex gesture made with both her hands in the central part of her gesture space, moving her hands upwards and shaking them both. This kind of gesture was considered as metaphoric because it is depicting a feeling or emotion (being flattered) with the hands as if they were concretely showing the feeling itself. Another example of metaphoric gesture performed by the speakers is the one corresponding to the concept of "very much": in this specific case, Student A is saying "il corvo, che era molto affamato" (translation: the crow, who was very hungry"), her

[6] The underlined words correspond to the position of the gesture in the utterance

hands, in this case, move from the central part of her gesture space down to her lower gesture space while her arms depict a rather big circular movement, carrying the meaning of something which seems to be "big". In this case, this metaphoric gesture acts as an enforcement of what the speaker is saying, conveying the sense of a very strong hunger (that the crow feels). The third example reported in fig. 23 is the occasion in which Student A performed a gesture that was annotated as abstract deictic in her first narration in English. While she is pronouncing the words "flies to a tree" her left hand moves as to point with her index finger to something that looks higher than her: the tree where the crow flies to. These kind of gestures were considered as abstract because the object they are pointing to is not present in the concrete context in which the speaker is narrating: by pointing to that imaginary tree, Speaker A is depicting and representing something that has to be imagined by the recipients.

The other macro-category that was considered in this investigation is the *discursive* category. Were considered part of this category all those gestures that did not share any physical nor abstract resemblance with what the speaker is talking about in that moment, but that are rather used as rhythmic markers or to convey pragmatic meanings such as approximation and vagueness (following Vinzce & Poggi, 2013). As it will be explained in the following paragraphs (§ 7.3) these kind of gestures occurred the most in the speakers' first narration in English since they are linked to the hesitations that might occur when narrating, but they still occurred in the other experimental conditions and some of them are reported below in fig. 24:

St. C – Italian 2

St. G – English 2

St. C – English 1

Fig. 24 Discursive gestures

In Fig. 24, three different discursive gestures are reported. In the first picture from the left, Student C is speaking for her second time in Italian and the phrase she is pronouncing is "così il corvo fece [disfluency] iniziò a cantare" (translation: "and so the crow did [disfluency] started singing"). In this case, right after hesitating, the speaker produced some circular movements with her right hand while the other hand held the position. In the second picture, the one in the middle, Speaker G is performing a more complex gesture with both hands while delivering her narration in English for the second time. The phrase she is pronouncing while performing this gesture is: "after it took it he goes on a tree, to to [repetition] eat it" and while she is saying "to eat it", she produces a circular movement twice in the central part of her gesture space, with her right hand, while the left hand just makes a couple of small movements back and forth. This gesture pragmatically informs the listener that the speaker considers the second part of the phrase as obvious but that she still needs to clarify the reason why the crow goes to the tree. The third example reported in Fig. 24 is Student C while, at the beginning of her narration says: "and…[silent pause] it tells that once upon a time" and, while pausing, her hand produces a circular clockwise movement. This discursive gesture seemed to be used by Student C as to help herself finding the right content of the story.

There is a category that, despite being considered, did not occur in this experiment: the *emblem* category. Since a few emblems were found in the previous experiment I expected to find the same in the present investigation. This result, or luck of result, on emblem gesture might suggest that the use of emblems (even if rare) in experiment 1 (chapter 5) could have been the result of the more communicative effort made by the speakers when they were asked to improve they communicativeness.

7. Results

7.1 About Pitch

The table below shows the results of the averages of the pitch analysis measures for the four experimental conditions:

	Ita t1	Ita t2	Eng t1	Eng t2
Min pitch	157.1	158.7	156.1	155.8

Max pitch	479.0	461.8	492.3	484.1
Mean pitch	212.0	208.1	207.7	206.7
ST. DEV	43.4	42.3	47.4	52.9
PVQ	0.21	0.21	0.23	0.25

Tab. 9. Pitch values in the four experimental conditions

The first interesting result is that while PVQ does not change in the two repetitions in the speakers' mother tongue, it varies when they speak English. In fact, their initial 0.23 PVQ becomes 0.25 one week later, when speakers repeat the task. A within-subject ANOVA test shows that while in Italian there is no significant difference in PVQ in the two repetitions, this difference is significant in English ($F(1,9) = 36$, $p < 0.01$). This means that, when speaking English, students improve their pitch variation in the second repetition. It is true that, compared to other languages such as Italian or German, English prosody is characterized by a more varied overall pitch range (Busà, 2008; Mennen, 2007; Mennen et al, 2010; Urbani, 2013). Furthermore, Italian L2 speakers seem to have a narrower pitch range and less variable pitch than native speakers of English (Busà & Urbani, 2011). Thus, the results of this experiment could show that repetition has an effect in L2 prosody. In fact, while the PVQ (Fig. 25) is not changed in the second repetition of Italian, things change when speakers repeat the story in English: their second repetition has a more varied pitch.

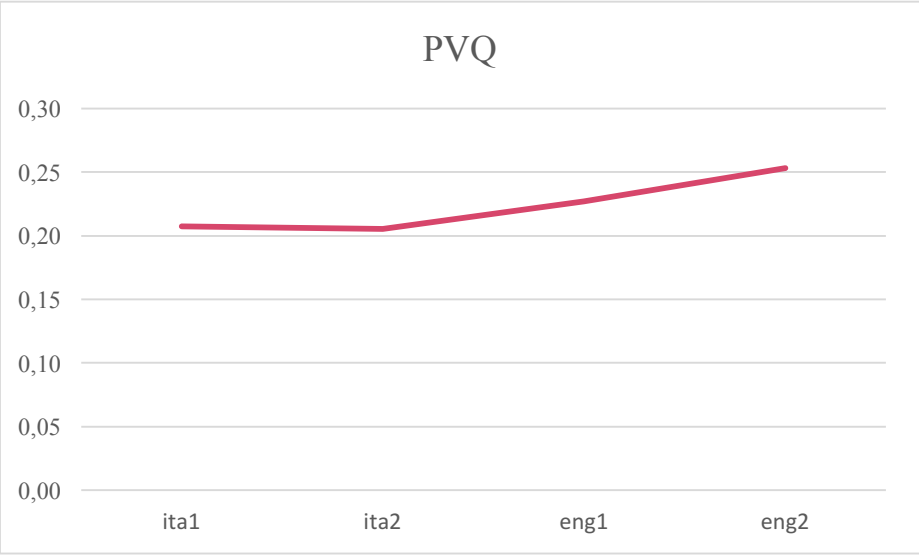


Fig. 25 PVQ in the four conditions

These results could be used to integrate the findings of Zimmerer et al. (2014) who analyzed German and French speakers and found that both groups of speakers had a tendency to produce a smaller pitch range in the respective L2s. Zimmerer et al. (2014) suggest that their finding can be explained as due to the minor confidence of speakers in their L2 productions that would force them to concentrate more on segments and words while refraining from realizing pitch in a more native-like range. The present study could actually prove that, with repetition, speakers gain that confidence and, as the word retrieval process becomes easier, they can concentrate in delivering a more native-like prosody with their speeches.

7.2 Speech data

The following tables (tab. 10 and 11) show all the speech data. The first one, tab.10, shows the average durations in seconds, the second, tab. 11, reports the quantities of occurrences of all the collected data.

AVG. DURATION (in seconds)	ITA1	ITA2	ENG1	ENG2
Articulated Speech	39.6	38.5	48.1	47.4
Silent pauses	1.6	0.1	4.8	3.8
Filled pauses	1.9	1.4	6.4	5.4
Breaths	4.3	4.5	9.2	6.2
Tot pauses	7.4	6.0	20.4	18.9
Tot story dur	50.8	45.5	69.5	61.8
Corrections	0.0	0.4	1.9	1.1
Disfluencies	3.8	3.2	10.0	7.9
Repetitions	0.6	0.2	1.2	0.8
Total Disfl.	5.7	4.6	16.4	12.3

Tab. 10 Average duration of speech features

Avg. OCCURRENCIES	ITA1	ITA2	ENG1	ENG2
Articulated Speech	14.8	12.70	33.20	25.80
Silent pauses	4.3	2.7	14	9.1

filled pauses	3.4	2.6	11.8	9.6
Breaths	8.3	7.8	14.1	12.4
Total Pauses	16	14.6	43.3	33.9
Words	125.8	128.7	128.6	132.6
Corrections	0	0	2	1
Disfluencies	9	7	23	17
Repetitions	1	2	4	4
Total Disfl.	13.2	9.8	35.5	27.1
Speech Rate	2.48	2.85	1.86	2.16

Tab. 11 Average number of occurrencies in speech and speech rate

First of all, the average number of words is not too different in the 4 repetitions: speakers use fewer words when they tell the story in Italian the first time (125.8); the use the same number of words when they speak Italian the second time and English the first time (128.7) and use a greater number of words when they speak in English the second time (132.6). However, these differences are not statistically significant.

The total duration of the articulated speech, without pauses is on average longer in English. In fact, even if the number of words used does not differ much, the storytelling lasted on average more in English t1, while the Italian t2 it has the shortest total duration. In both languages the total time that speakers used to tell the story to their audience decreases after one week. In fact, while in Italian the mean duration goes from 50.68 to 45.15 seconds (5.4 seconds less), in English it shifts from 69.25 to 61.28 (7.7 seconds less). Statistically, there is no significance between the first and the second repetition in each language. Yet, comparing the total duration from a cross-linguistic point of view, both differences result relevant. In fact, the within-subjects ANOVA showed that the difference in duration between Ita t1 and Eng t1 is highly significant ($F(1,9) = 17.434$, $p < 0.01$) and so is the difference between Ita t2 and Eng t2 ($F(1,9) = 14.558$, $p < 0.01$).

The same pattern is visible when comparing the total duration of pauses. Pauses decrease in time from t1 to t2 in both languages, and their cross-linguistic difference is even more evident. If in Italian the total duration of pauses (silent, and breaths) is around 7.84 seconds the first time and 6.40 seconds the second time, in English their average duration is 20.64 the first time and 18.19 seconds the second time. Pauses are not simply longer in English, but they are quantitatively

more in English than in Italian. In fact, the total number of pauses decreases from 16 to 14.6 in Italian, and in English they decrease from 43.3 to 33.9 seconds. In summary, pauses decrease in both cases, but in English their reduction is greater, in both duration and quantity. Statistically, there is no significant difference between the first and the second repetitions in Italian and this means that the fluency does not change much in the speakers' native language. There are statistically significant differences between the first and the second repetition in English (at the ANOVA test $F(1, 9) = 6.687$, $p < 0.05$ for the average total duration of the pauses, and $F(1,9) = 9.693$, $p < 0.05$ for the average number of pauses) leading to the conclusion that the speech fluency gets better in the second repetition. When comparing data from a cross-linguistic point of view, the within-subject ANOVA test revealed highly significant differences both between Ita t1 and Eng t1 ($F(1,9) = 29.899$, $p < 0.01$ for the average total duration, and $F(1,9) = 41.048$, $p < 0.01$ for the average total number of pauses) and between Ita t2 and Eng t2 ($F(1,9) = 13.475$, $p < 0.01$ for the average total duration, and $F(1,9) = 20.243$, $p < 0.01$ for the average total number of pauses).

A further indication of a better fluency in the second repetition of English is given by the data about disfluencies (the sum of full pauses, corrections and word repetitions). In fact, they decrease in number and duration in both languages, but it is in English that their reduction reaches significance ($F(1,9) = 5.136$, $p < 0.05$).

The results above are reflected in the speech rate data too. In Italian, speech rate (the relation between the words and the time used in speech) increases only of a few decimals (from 2.48 to 2.85), and this difference is not significant. On the other hand, speech rates of Eng t1 and Eng t2 show an increase (from 1.86 to 2.16) that appears statistically significant ($F(1,9) = 6.275$, $p < 0.05$). The fluency in English is significantly improved when students repeat the task after one week from their first narration. Cross-linguistically, the differences between Ita t1 and Eng t1, and between Ita t2 and Eng t2, are both relevant.

In conclusion, these results let assume that the cognitive effort changes considerably in the two language contexts: while repetition do not influence the speech rate in the students' native language, it seems to have an important effect in their L2, helping them to reach a speech rate that is closer to that of native speakers.

7.3 Gesture data

Fig. 26. Gesture sequence in which the Speaker L is saying "the cheese falls"

The gesture analysis was based on the occurrence and gesture rate of gesture types collected under two macro-categories, discursive and representational, as explained above (see paragraph 6).

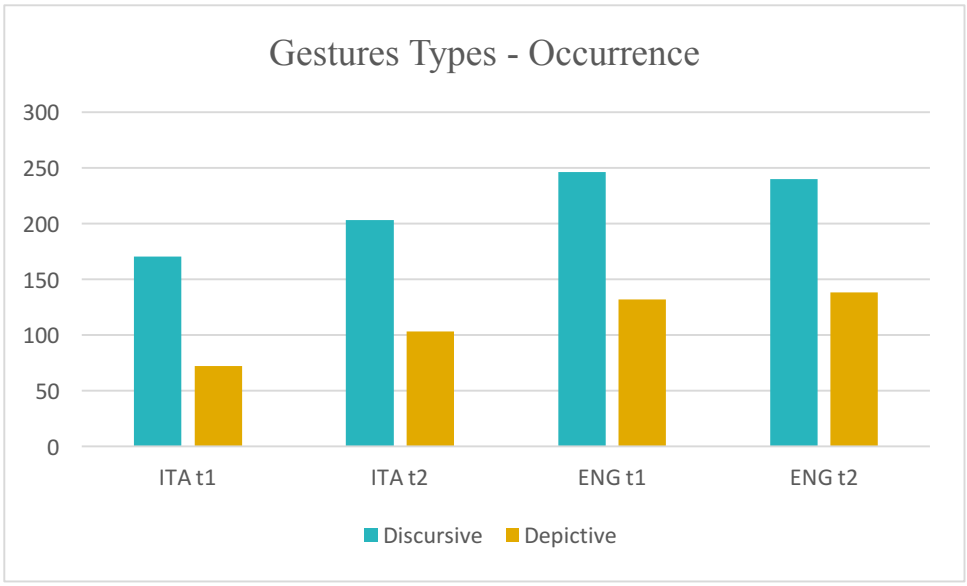


Fig. 27. Occurrencies (in raw numbers) of the two macro-categories in the four conditions. Emblems gestures did not occur.

Fig. 27 shows the occurrences of each gesture category in the four datasets. In general, speakers used much more discursive gestures in all four datasets. Overall, speakers gestured more when they narrated the story in English (see tab. 12). Surprisingly, while the total count of gestures increased in Italian (from 242 in Ita t1 to 306 in Ita t2), the total number of gestures in English was identical in both repetitions. Looking at the two main categories in which the gestures were

divided, the composition of the total gesture count seems to be only slightly different (not statistically).

Looking at the percentage of gesture usage (tab.12) it can be noticed that the occurrence of discursive gestures is always above 60% in all datasets. In particular, Ita t1 shows the highest percentage of discursive gesture usage (70,25%). Representational gestures are a slightly more used in both English datasets.

% Gesture Occurrence		
	Discursive	Representational
Ita t1	70,25%	29,75%
Ita t2	66,34%	33,66%
Eng t1	63,05%	36,95%
Eng t2	63,59%	36,41%

Tab. 12. Percentage of gesture usage in each experimental condition

The overall Gesture Rate (fig.28) follows the same trend. When comparing their Italian performances, subjects seem to gesticulate more in their second repetition while, when telling the story in English, their gesture rate does not seem to change much, letting assume that

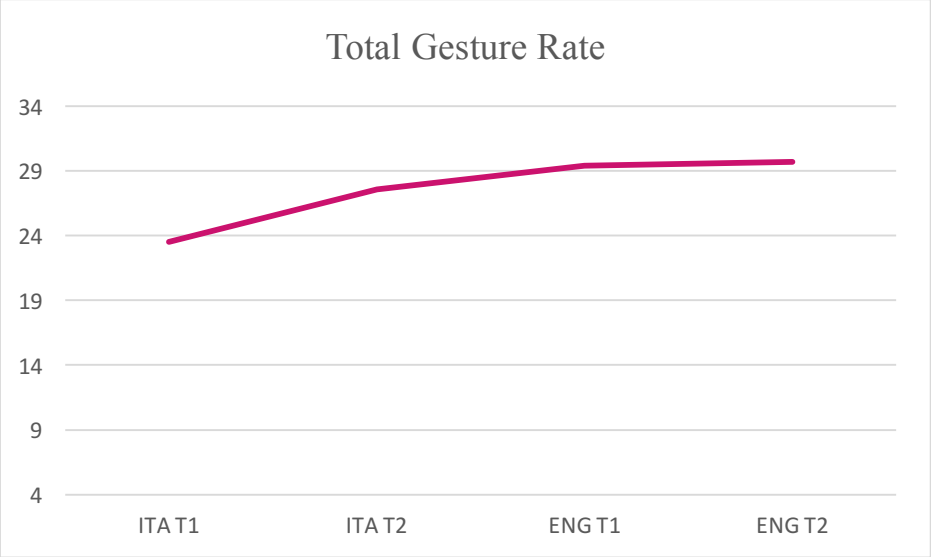


Fig. 28 Total Gesture rate

repetition alone does not influence gestural behavior when the subjects are speaking in their

second language in front of an unknown audience (in fact, if the audience had been the same of the first performance, the subjects probably would have gesticulated less, due to the common ground effect explained suggested by Holler & Wilkin, 2009).

The gesture rates (tab. 13) which is the relation between gestures and words (N of gestures/N of Words), show higher values in English. This means that the subjects gesticulated more when speaking their L2. In general, it seems that speakers increased their gesture rate when they repeated the experiment the second time. These values, however, show no statistical significance between the two repetitions in the two languages. The only statistical significance found is between Ita t1 and Eng t1 Representational gestures ($F(1,9) = 5.865$, $p < 0.05$), revealing that, in their first performance, the subjects performed less Representational gestures in English than they did when telling the story in Italian.

Gesture Rates			
	Discursive	*Representationals*	TOTAL
ITA T1	16,54	7,01	23.54
ITA T2	17,42	8,47	26.27
ENG T1	18,09	9,71	29.41
ENG T2	18,10	10,40	29.71

Tab. 13. Gesture rates

In the figure below (fig. 29) all the gesture rates, divided into main categories, can be compared:

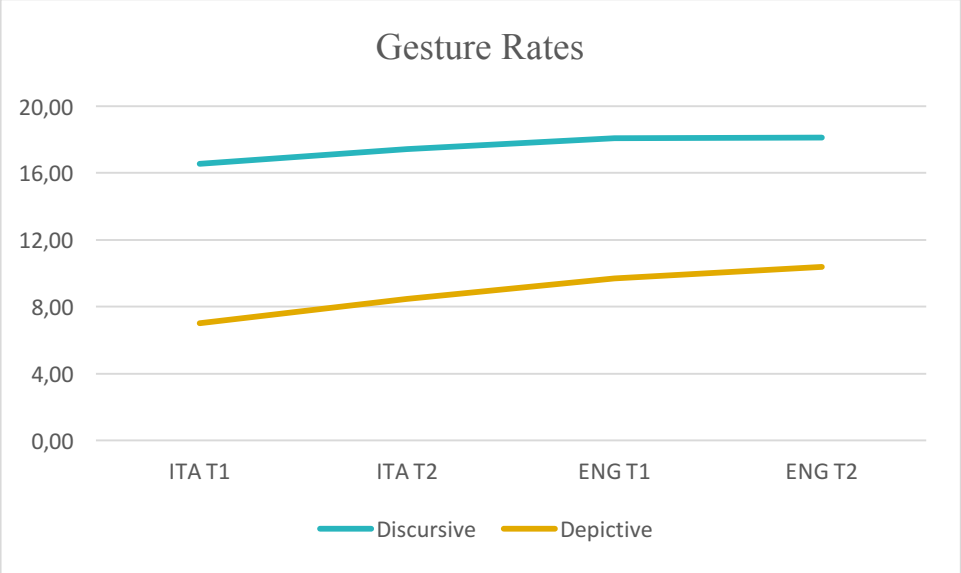


Fig. 29 Gesture rates of the four conditions for each gesture type

8. A Little Insight into the Vagueness Issue

Vagueness gets on stage whenever people are not sure of what they say or just want to deliver something with approximation. It is delivered in speech, for example through disfluencies and full pauses, but it is also communicated through gestures.

For this experiment the subjects were asked to recall a cartoon they just saw and to tell it in front of a small audience while being video-recorded. These factors created the right context for vagueness gestures to appear since the students had to speak in front of a camera trying to recall a cartoon they did not know before. Even if vagueness gestures were not the main focus of the current experiment, they were counted as a category and then included in the *discursive* gestures.

The table below (tab.14) reports the *vagueness* gesture rate which occurred in all four datasets:

Gesture Rate -Vagueness	
Ita t1	1,46
Ita t2	1.59
Eng t1	3.86
Eng t2	2.68

Tab. 14. Gesture rate for the vagueness category

The *vagueness* gesture rate shows how speakers' greater effort and greater difficulty occurred in their first telling of the story in English. Statistically, their performances differ from a cross-linguistic point of view, in fact, the difference between Ita t1 and eng t1 are significant (at the ANOVA within subjects test $F(1, 9) = 7.451$, $p < 0.05$). These data suggest that the lack of confidence and word retrieval processes seem to appear with a greater frequency when subjects are asked to tell the story for the first time in English. The significance, though, disappears when comparing the two languages in their second repetitions (t2). It seems that repetition brings *vagueness* values closer to those shown in the subjects' native language.

Speech Rate	
Ita t1	2.48
Ita t2	2.85
Eng t1	1.86
Eng t2	2.16

Tab. 15. Speech rate in the four conditions

Comparing the vagueness gesture rate with the speech rate reported above (tab. 15), it seems that at higher speech rate values correspond lower vagueness gesture rate values. Repetition allows speakers to reach a greater ease with the task and this seems to both help speech fluency (higher speech rates) and gestural fluency (through the decrease of vagueness gestures).

It can be speculated that vagueness in speech and gestures also parallel speakers' overall pitch values too. The Pitch Variation Quotient data (reported below, tab. 16) showed that the speakers were able to convey more pitch-varied speech in their second repetition in English. The subjects' patterns of vagueness gestures in Italian seem to be similar to their PVQ patterns. In fact, neither their vagueness gestures nor their pitch change significantly in the second repetition.

	PVQ
Ita t1	0.21
Ita t2	0.21
Eng t1	0.23
Eng t2	0.25

Tab. 16. Pitch Variation Quotient for the four conditions

There is not enough data to conclude that repetition has a positive effect on the decrease of vagueness gestures and that this can be related also to higher values of speech rate. Certainly, these results lead to further research on the correlation between vagueness gestures with speech and prosody. Though this interrelation could be expected, it would still be interesting to define how vagueness in gestures is linked to speech and prosody.

9. Discussion

This study examined the way in which Italian speakers told the same story twice in their native language and twice in English at two different times, at the distance of one week. The short time distance did not allow the subject to reach higher proficiency in their L2 while, at the same time, allowed them to keep the story in their memory. The object of investigation were the differences

the subjects could display in their global pitch range, as well as in their speech characteristics and gesture behavior. While in the previous experiment subjects were asked to be communicative in the second repetition, this time they were simply asked to repeat the story a week later. In order to avoid a possible common ground effect, the only thing that differed was the audience the subjects told the story to.

When speaking Italian, the subjects did not change their pitch values in their second repetition. This result suggests that repetition did not have any significant effect in speech liveliness (Hincks, 2005) and that a greater ease with their task (i.e due to repetition) had no effect on their pitch variation. In English, though, the subjects did change their pitch and actually varied it significantly in the second repetition. This result can be interpreted to mean that the first time the speakers narrate the story in English, their minor confidence with the language and with the task itself shows up in a less varied pitch range; however, when they repeat the task one week after, they are able to better imitate an English prosody through a greater pitch variation. Thus, it is possible to conclude that repetition has a positive effect on L2 prosody (and maybe on L2 liveliness) even if the subjects are not asked to increase their pitch variation or to be more communicative. So, with repetition, speakers gain more confidence than when they are asked to tell the story the first time. Also, since the word retrieval process becomes easier and their cognitive effort decreases, they are more able to deliver a more native-like prosody in their speech. These results are in line with Zimmerer et al. (2014), who suggest that when L2 speakers are less confident in their speech they concentrate more on segments and words while refraining from realizing pitch in a more native-like range.

The speech data showed that fluency does improve significantly only in speakers' L2. In fact, the results show a quite unexpected lack of significant difference between the first and the second repetition in Italian. In other words, even if it is true that the speech rates increase in the speakers' second re-telling of the fable, it is not possible to claim that repetition has a strong impact on the speakers' native language fluency. On the other side, the difference between the first and the second narration of speakers' L2 speech is statistically significant. In this case, the results suggest that repetition helps L2 speakers improve their fluency. The data confirm that, in their second repetition, speakers show less disfluencies (filled pauses, corrections and word repetitions). Disfluencies decrease in number and duration in both languages but their decrease is statistically significant only in English. Once again, repetition seems to play a large role in helping L2 speakers to reach a greater fluency.

Even if the differences are not statistically significant, the number of words used in the second repetition are slightly higher in both languages. The same happened in the first experiment, when speakers were asked to be more communicative the second time they told the story (Chapter 5). Yet, while in the previous experiment the difference between the first and the second repetition was highly significant [English t1-t2: ($F(1, 9) = 12.699$, $p < 0.01$)], this difference was not significant in the present experiment. The reason why the subjects used significantly more words in the first than in the second experiment needs to be further investigated. A possible speculation is that the fact that the number of words of all four datasets in the second experiment is similar could indicate that a *common ground* effect (Kuhlen, Galati & Brennan, 2012). Gesturing integrates top-down and bottom-up information: Joint effects of speakers' expectations and addressees' feedback. Language and Cognition, 4(1), 17-41.) effect was successfully avoided.

In short, as expected, the subjects are more fluent in Italian than in English. The pitch appeared to be identical in the two repetitions in Italian but to significantly increase in the English repetitions. The second time the subjects told the story, their cognitive load was reduced and this could have allowed them to increase their pitch variation (PVQ), in order to sound more "native" (as suggested by Zimmerer et al. 2014). As has been shown in the literature, English and Italian differ greatly in their pitch and intonation patterns (Busà, 2008; Busà & Stella, 2014; Polyanskaya & Busà, to appear). Although their pitch varied more in English, the subjects still produced speech with more and longer pauses and disfluencies than they did in Italian, even if these significantly decreased in the second repetition (Eng t2).

The speech analysis showed that subjects increase their fluency in the second repetition even without any instruction by the experimenter. While the better fluency seems to be barely noticeable in the subjects' native language, the subjects improve significantly when they repeat in their L2. Thus, by decreasing the speakers' cognitive effort, repetition not only allows speakers to deliver a more native-like pitch in the L2, but it also allows them to be more fluent, with values that are closer to the ones they have in their native language.

The subjects gesture more in English than in Italian. However, the total amount of gestures occurring when they tell the story in English in the 2 repetitions is identical. In Italian, instead, even if the total amount of gestures is smaller than in English, the number of gestures subjects made actually differs in the two tasks, showing an increased number of overall gestures in the second repetition (Ita t2). The fact that subjects perform the same amount of gestures both times

when they tell the story in English is quite unexpected. Since gestures are largely connected to the cognitive complexity of the task, it was expected to find a greater gesture rate in the first than in the second repetition, similarly to what happens with speech rate. In fact, speech rates in English actually show a statistically significant increase (from 1.86 in the first repetition to 2.16 in the second repetition) and become more similar to the values of the native language. Gesture rates show higher values in English, even if there the difference is not statistically significant. Thus, in this experiment speech rates, PVQ and gesture rates do not grow together as they did when, in the previous experiment, subjects were asked to deliver a more communicative fable.

In all the datasets, the subjects used more discursive gestures than representational gestures. The percentage of representational gestures occurring in English was higher than in Italian, and this could mean that, as suggested by Alibali et al. (2017) and Kita et al. (2017) representational gestures actually have an important cognitive and self-orientation function.

7 / Integrating the results

In gesture studies, when speaking about speech, researchers usually consider the semantic and pragmatic aspects of the utterance. In this work, even if the pragmatic and semantic aspects of gestures were not excluded from the annotations and analyses, the main focus of analysis about speech was on prosodic aspects. Speech was analyzed in order to investigate the role of pauses, disfluencies (which mark rhythmic aspects of the utterances), and speech and articulation rates in relation to gesture behavior (with a specific focus on gesture "representationality") in a situation in which subjects were instructed to repeat the task being more communicative (experiment 1 – Chapter 5) and in a case in which subject were only asked to repeat the task but were given no specific instruction (Experiment 2 – Chapter 6).

In both studies, the task was to (re)tell Aesop's fable "The Fox and the Crow" in L1 and L2 in front of a small public. The results of these studies can contribute to understanding the role of speech and gestures in both second language acquisition (SLA) and Public Speaking. In SLA, the results highlight the role of instruction in L2. In fact, the results show that, on the one side, repeating a task can significantly lead to a change in students' habits both at the verbal and non-verbal levels. On the other hand, providing students with instructions to be communicative can be effective in the L2 production. In the Public Speaking field, there is a lack of studies showing the effectiveness of communication instructions even though students are usually told that to be more communicative to enhance listeners' attention they need to use a more varied intonation and meaningful gestures. Public speaking classes also focus on the importance of repetition and rehearsal, which are claimed to be essential for bringing about changes in speakers' speaking and delivery styles. The first of the two studies presented in this work investigates also how repetition of a task can enhance substantial changes in the speakers' verbal and non-verbal behavior.

During the development and the conclusion of the experimental work, I found some limits that could be avoided in future research. One limit of this work is that the two investigations cannot give a complete portrait of the effects that communicative instruction and repetition have in the native language. The lack of comparison between a first and a second repetition in exp.1 still leaves the question open as to whether communicative intent might affect positively fluency and pitch variation in Italian. Since in exp.2 the native language pitch variation and fluency did not

result to be significantly changed by repetition, it would have been interesting to verify the pitch and fluency patterns of speakers' native language in a condition in which the subjects were instructed to increase their communicativeness. In the first investigation presented in this work the subcategorization of gestures was not registered. In other words, gestures were analyzed and compared as macrocategories, this means that it is not possible to understand if all the gesture types that are included in the representational category (iconic, metaphoric, abstract deictic) changed equally when the speaker told the story the second time. It would be interesting to repeat the two experiments with a wider number of subjects, controlling these limitations and, moreover, showing all the subjects the cartoon stimuli, in order to uniform the kind of speech elicitation and to avoid possible differences due to the different kind of stimuli the subjects received.

Despite these limitations, the results of the two investigations can be compared and all the main results are summarized in the table below (Tab. 17).

	Experiment 1: Instruction to be Communicative	**Experiment 2: Repetition with no instruction**
Pitch	**PVQ** shows a greater pitch variation in English t2.	**PVQ** does not change in Italian but it varies when subjects speak in English, showing they improve their pitch variation in the second repetition
Fluency	The number of **words** is significantly different between English t1 and t2, showing that the second time speakers increase the number of words they use.	The number of **words** does not change significantly
	Pauses decrease in duration and number in Italian t1 than in English t1 and English t2. Instead, there is no significant difference between the two English repetitions.	**Pauses** decrease in duration from t1 to t2 in both languages. In English they always last significantly more than in Italian. Pauses are also quantitatively more in English. Their number decreases at t2 in both languages.

		At the same time, **disfluencies** decrease significantly in number and duration in English, while their decrease is not significant in Italian.
	Articulated speech does not show statistical difference but it shows that in English there is a decreasing trend in time from the first to the second repetition. **Articulation rate** shows lower values in English t1 than in Italian t1. English t1 also differs statistically from English t2 condition, in which subjects increase their articulation rate.	**Articulated speech** is longer in English t1. In general, though, there is a general decrease in the total duration of the articulated speech in all conditions. The difference is between languages in both repetitions
	Speech rate The subjects' speech rate show a greater fluency in the second repetition. Instead, it shows the lowest values in English t1, showing a great cognitive effort by the speakers. Generally, the speakers' values get closer to those showed in Italian.	**Speech rate** increases significantly only in L2, showing repetition helps students reach a better fluency on their second repetition. There also is a cross-linguistic relevance, showing greater values in Italian –that is, speakers speak faster in Italian than in English.
Gestures	In general, speakers seem to gesture more in their second repetition. **Discursive gestures** are used less in Italian than in English. English t2 is the condition in which subjects used more discursive gestures. Though, there is no statistical significance in these differences. **Representational gestures** increase significantly in English t2. This might be due to the instruction to be more communicative. The **overall gesture rate** shows no significant difference. Thus,	In general, there is a greater **occurrence** of all gestures in English. The total count of gestures increases in Italian (t2), they do not increase in English (t2). **Representational** gestures are statistically more used in English t1 than in Italian t1. The **overall gesture rate** shows a growing trend in Italian, while it does not change in English.

	the only macro-category that significantly changes is the representational one.

Tab. 17. Comparison of the two investigations results

Conclusions

When speaking, delivering a speech, or narrating a tale, speakers have to employ a number of different processes that allow them to convey certain meanings and pragmatic contents to their interlocutors. Specifically, the aim of this work was to investigate the relationship between speech and gestures in communication. The study was carried out to contribute to filling a gap in the studies about gestures in relation with speech, and particularly prosody.

In both investigations, the pitch variation quotient resulted to be higher in the second repetition in English. Whereas in the first experiment (from now on, exp.1) there was no possibility of comparison with the PVQ in speakers' native language, in the second one (from now on, exp.2) the results show that subject did not increase their pitch variation in the second repetition when speaking Italian. The greater variation in L2 is probably due, as suggested by Jenkins, 2002 and by Zimmerer et al., 2014, to the fact that native and non-native speakers process prosodic information differently because of their different levels of competence in the L1/L2. In fact, non-native speakers, more than that on prosody, rely on segmental information to get their meanings across, since they lack the amount of extra-linguistic knowledge that native speakers can rely on when communicating. In exp.1, then, that differences between the L1 and the L2 PVQ may be more evident when non-native speakers are focusing on getting their meanings across because of the communicative instruction in the second task. Thus, the second time the subjects were called to tell the fable in English may have caused them to use a significantly more varied, wider, pitch. In exp.2, where no instruction was given though, the pitched resulted to be identical in the second narration in Italian, while there has been a significant increase in the English task. These results seem to confirm the previous results by Jenkins (2002) and Zimmerer et al. (2014) and corroborate the idea that, in the second repetition of English, students gain confidence with the task, reduce their cognitive load and, thus, are able to concentrate more on segments and words and subsequently improve the variation in the L2. In fact, English is commonly described as a more rhythmic or tone-varied language than Italian (Busà, 2008; Busà & Stella, 2014; Stella & Busà, 2014) and this could have led the students to vary more their pitch. In other words, in both exp.1 and in exp.2 the speakers could have got a benefit from repetition, that would have helped them to better imitate the English prosody. Unfortunately, in exp.1 the students were not asked

to repeat the task in their native language. Since in exp.2 the PVQ in Italian did not show any significant change (the subjects already had high proficiency in their mother language and did not need to change their own pitch), it would be interesting to integrate the results of exp.1 with a comparison of a first and second repetition in the subjects' mother language in order to verify whether the communicative instruction can elicit a change in PVQ also in Italian. Below (tab. 19) the comparison of the PVQ in both experiments:

PVQ	Italian t1	Italian t2	English t1	English t2
Exp.1	0.19	---	0.20	0.23
Exp. 2	0.21	0.21	0.23	0.25

Tab. 18. Comparison of the PVQ results

Beyond prosody, a second aspect that was investigated in this work was how repetition and communicative instruction can affect speech features such as word production, pauses, disfluencies and speech rate.

In a study by Hoetjes et al. (2014) the speech characteristics of subjects that were asked to describe a tie knot in two different contexts were compared in two conditions in which they were able or not able to gesture about it. The authors also looked at the number of speakers' attempts (or repetitions), under the assumption that the first attempt is cognitively more difficult than the following ones. The results of their study did not show any effect of the ability to gesture on the number of words produced, speech duration, speech rate, production of pauses, and pitch range. However, the number of attempts had an effect on the decrease of the number of total words, speech duration, and production of filled pauses. Eventually the results led to the conclusion that, with repetitions, speakers become more fluent and use fewer words; at the same time, though, pitch and speech rate did not change significantly. The results of the investigations ran in this work strongly agree with those by Hoetjes el al. (2014) mainly for what concerns exp.2, in which the subjects were just asked to repeat the task one week after their first narration.

As for the number of words, whereas in exp.1 there is a significant difference between the first and the second repetition in English, with a relevant increase in the number of words used to tell the story, in exp.2 the total number of words did not show any particular difference between

repetitions in any language. In fact, even if the speakers used slightly less words when they told the story in Italian, the average number of words was not too different in any of the 4 repetitions. This different result points to the result of how the subjects reacted to the instruction to be communicative in exp.1.

When comparing the results about pauses, both experiments (exp.1 and exp.2) show that speakers made fewer and shorter pauses in Italian, also in the second repetition in exp.2. However, the pauses show a decrease in number and duration in both languages when, after a week, the subjects repeated the task. Disfluencies show a similar pattern, with a significant reduction in number and duration in English. This trend shows that repetition actually helps L2 fluency even one week later and with no rehearsal between the two repetitions. In Italian, though, repetition did not bring any significant improvement, probably due to the already high proficiency of the speakers.

As for the articulated speech, there seems to be a decreasing trend in L2 in both experiments. Exp.2 shows a general decrease of the total duration of the articulated speech also in Italian. These data suggest that in both cases fluency improves. In general, when subjects were asked to speak more communicatively (in exp.1) they increased their fluency in a way that is comparable to the one they show in exp.2, when they only have to repeat the task. This means that the instruction to be communicative does not affect fluency as much as repetition does.

The same pattern is shown for speech rates. In both exp.1 and exp.2, in fact, the subjects showed a significant increase in speech rates in their second repetition in L2. In exp.2, the improvement of speech rate did not result to be relevant in Italian, since the speakers' fluency is already high.

In conclusion, when analyzing speech features and fluency, in both experiments the subjects showed an increase in the values in L2, showing that the instruction to be more communicative that was given in exp.1 does not affect or increase speakers' fluency more than the lack of any instruction at all. This suggests that the increase in fluency is given mainly by the repetition effect. These results lead to the assumption that helping subjects decrease the cognitive effort with repetition allows speakers to convey a more native-like prosody in L2.

One of the main focuses of this work was to better understand the role of gestures in their communicative function. Specifically, this work contributes to the existing debate on the role of gestures in communication (see, for example, Alibali & Kita, 2010; by Alibali et al. 2017; Kita et al., 2017; Bavelas, 2008; Kita, 2010, 2017; Krauss, Morrels-Samuels, & Colasante, 1991) with

a focus on the role of representational gestures. In particular, the main hypothesis was that when subjects are asked to increase their communicative effort in a way that can attract their recipients' attention, they will increase the representational gestures. To do so, gestures were analyzed and divided into two macro-categories: the first one collected all the representational gestures (that somehow physically represent features of what the person is saying, both iconically or metaphorically) and was called *Representational.* The other macro-category included all the gestures that bring no representationality but that rather focus on pragmatic or rhythmical aspects of the utterance, this category was called *discursive*.

In exp.1 the results show a trend of a higher number of gestures and higher gesture rate the second time speakers tell the story, although these differences are not statistically significant. What is highly significant in exp.1, instead is the difference between *representational* gestures in the two repetitions. In exp.2, instead, gestures do not seem to behave too differently in the two repetitions. In both experiments there is a generally greater occurrence of gestures in English (both the first and second time) than in Italian. However, there are no differences between the two repetitions in either language.

These results lead to the conclusion that, in exp.1, the students reacted to the instruction to be more communicative by increasing their overall gestures, probably in order to help their recipients better understand and visualize the content of the story. But, even if both categories show an increasing number of gestures, the one that increases the most is the *representational* one. This result suggests that, when asked to increase their communicativeness, subjects use a more "visual" strategy and increment the "representationality" of their gestures.

In exp.2, only the cross-linguistic difference in representational gestures was significant, meaning that, in their first performance, the subjects used more representational gestures in English than they did when telling the story in Italian. This result actually supports the idea that representational gestures help speaking (Kita, 2010). Students, in fact, may have used these kinds of gestures to help themselves retrieving the words they intended to say. In their second repetition this difference is no longer significant and there is no difference in the use of representational gestures between the second repetitions in Italian and English.

The way in which people employ representationality through gestures in co-occurring speech is not clear. Some, (as, for example Alibali & Kita; 2010; Kita 2010, 2017) claim that representational gestures, and consequently iconic gestures, help speakers in the conceptual

planning of speaking, or that gestures have just an occasional and incidental role in communication (see Freyersen et al., 1988; Krauss et al., 1990). Others, (such as Holler et al., 2009; McNeil, 1985, Melinger & Levelt, 2004) claim that iconic gestures do carry an important communicative function and can be intentionally used to convey additional semantics to the recipients.

So, gesture and speech are part of the same process (Kendon, 2004; McNeill, 1992) and provide a support for some cognitive processes that lead to speak but also do convey information to the addressees, helping the communicative function of an utterance. Yet, the measure in which gestures and speech are related when speakers intend to convey a communicative message is still not deeply investigated. In particular, speech features like pauses, speech rate and pitch variation were investigated in relation to gesture production only in few studies (Hoetjes et al., 2014; Krahmer & Swerts, 2005; McClave, 1998).

The approach followed in this work was to understand whether the speaker's communicative intention can modify the way they use speech features and gestures. In particular, what was expected was that, when speakers intend to convey a more communicative speech, their fluency would change, their pitch variation would increase and they would employ a greater iconic gesture rate. This study was originally thought to investigate the role of gestures in L2 and the possible implications or correlations with prosodic and rhythmic features in public speaking.

The results of exp.1 lead to the conclusion that if a person is asked to be communicative, they will probably increase both their pitch variation and the total number of gestures produced, with a significant increase in iconic and representational gestures. Nevertheless, the relation between pitch and gestures was not clear. Therefore, it was necessary to get a clearer view on the effect of task repetition on pitch and gestures without the communicative instruction. Exp.2 was then designed to verify if task repetition can cause an increased PVQ and increased gesturing by itself: with the facilitation of the task and a subsequently facilitation of the lexical retrieval processes, speakers may use a more varied pitch range and increased gesturing.

The integrated results of exp.1 and exp.2 seem to confirm that repetition and communicative task do have a positive effect in speech and gestures characteristics. When people are asked to tell a story in a more communicative way, their response is to increase the representationality carried in their gestures and consequently use more representational gestures. This result is confirmed by the results of exp.2, in which speakers, did not increase their representational gestures in the

second repetition, since they did not receive any different instruction besides the one to tell the story they had just seen in the cartoon. On the other side, speech features and fluency improve also in the exp.2, when the instruction to be communicative is absent.

There still is too little evidence in exp1to claim that prosodic variation and fluency are positively affected by the communicative instruction. Instead, there is enough evidence to affirm that repetition plays a role in producing a more varied pitch in L2. The task repetition, in both exp.1 and exp.2, could have helped the speakers with their second retelling: they learned the story better and thus had no need to make a consistent cognitive effort, this could have helped them gain confidence and hence, they could both reduce their disfluencies and deliver a more native-like prosody in L2 (as also suggested by Zimmerer et al. 2014). The instruction to be communicative did indeed produce an effect on the way speakers told their story. Thus gestures seemed to be more affected by the instruction, showing a significant increase in the representational category. Representational gestures, therefore, besides being used as a compensation strategy for speech organization and lexical retrieval (Kita, 2010) are also seen, by the speakers themselves, as a communication strategy to help the recipients visualize the content of their speech.

References

Adams, C., & Munro, R. R. (1978). In search of the acoustic correlates of stress: fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English. *Phonetica*, *35*(3), 125-156.

Ahrens, B. (2005). Analysing prosody in simultaneous interpreting: difficulties and possible solutions. *The Interpreters' Newsletter*, *13*, 1-14.

Argyriou, P., Mohr, C., & Kita, S. (2017). Hand Matters: Left-Hand Gestures Enhance Metaphor Explanation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* no pagination specified

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*(2), 169-188.

Alibali, M. W., & Kita, S. (2010). Gesture highlights perceptually present information for speakers. *Gesture, 10*(1), 3-28.

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15*(6), 593-613.

Alibali, M. W., Yeo, A., Hostetter, A. B., & Kita, S. (2017). Representational gestures help speakers package information for speaking. *Why Gesture?: How the Hands Function in Speaking, Thinking and Communicating, 7*, 15.

Aoyama, K., & Guion, S. G. (2007). Prosody in second language acquisition. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege.Amsterdam,*

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37*(5), 715.

Axtell, R.E. (1991). *Gestures – The Dos and Taboos of body language around the world*. New York: John Wiley & Sons.

Backman, N. (1979). Intonation errors in second-language pronunciation of eight Spanish-speaking adults learning English. *Interlanguage studies bulletin*, 239-265.

Beckman, M., Díaz-Campos, M., McGory, J. T., & Morgan, T. A. (2002). Intonation across Spanish, in the Tones and Break Indices framework. *Probus*, *14*(1), 9-36.

Bavelas, J. B. (1994). Gestures as part of speech: Methodological implications. *Research on language and social interaction*, *27*(3), 201-221.

Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin, 21*(4), 394-405.

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse processes*, *15*(4), 469-489.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*(2), 495-520.

Bavelas, J., Kenwood, C., Johnson, T., & Phillips, B. (2002). An experimental study of when and how speakers use gestures to communicate. *Gesture*, *2*(1), 1-17.

Beattie, G., & Coughlan, J. (1999). An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. *British Journal of Psychology*, *90*(1), 35-56.

Beattie, G., & Shovelton, H. (1999a). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? an experimental investigation. *Semiotica, 123*(1-2), 1-30.

Beattie, G., & Shovelton, H. (1999b). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology, 18*(4), 438-462.

Beattie, G., & Shovelton, H. (2002). What properties of talk are associated with the generation of spontaneous iconic hand gestures?. *British Journal of Social Psychology*, *41*(3), 403-417.

Beattie, G., Webster, K. A., & Ross, J. A. (2014). Do speakers really unconsciously and imagistically gesture about what is important when they are telling a story? *Semiotica, 2014*(202), 41-79.

Berger, K. W., & Popelka, G. R. (1971). Extra-facial gestures in relation to speechreading. *Journal of Communication Disorders, 3*(4), 302-308.

Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia, 44*(2), 178-190.

Birdwhistell, R. L. (1955). Background to kinesics. *ETC: A review of general semantics*, 10-18.

Birdwhistell, R. L. (1970). Kinesics and context: Essays on body motion communication.

Bolinger, D. (1983). Intonation and gesture. *American speech*, *58*(2), 156-174.

Bolinger, D., & Bolinger, D. L. M. (1986). Intonation and its parts: Melody in spoken English. Stanford University Press.

Bonifacio, G. 1616. *L'Arte de'Cenni*.

Borràs-Comes, J., Roseano, P., Vanrell, M. d. M., Chen, A., & Prieto, P. (2011). Perceiving uncertainty: Facial gestures, intonation, and lexical choice. *Proceedings of GESPIN*

Brown, A. (2007). Crosslinguistic influence in first and second languages: Convergence in speech and gesture. PhD Thesis, Boston University, Boston.

Brown, A., & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1-L2 encoding of manner in speech and gesture: A study of japanese speakers of english. *Studies in Second Language Acquisition, 30*(02), 225-251.

Bull, P. (1983). *Body movement and interpersonal communication* John Wiley & Sons Inc.

Bull, P., & Connelly, G. (1985). Body movement and emphasis in speech. *Journal of Nonverbal Behavior, 9*(3), 169-187.

Busà, M. G. (2008). Teaching posody to italian learners of english: Working towards a new approach.

Busà, M. G., & Stella, A. (2014). The acquisition of english l2 prosody by italian native speakers: Experimental data and pedagogical implications. *Pronunciation in Second Language Learning and Teaching Conference (Issn 2380-9566),* pp. 15.

Busà, M. G., & Urbani, M. (2011). A cross linguistic analysis of pitch range in English L1 and L2. In *Proc. 17th International Congress of Phonetic Sciences (ICPhS XVII), Hong Kong* (pp. 380-383).

Butterworth, G. (2003). Pointing is the royal road to language for babies. *Pointing: Where language, culture, and cognition meet*, 9-33.

Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. *Recent advances in the psychology of language* (pp. 347-360) Springer.

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill.

Calbris, G. (1990). *The semiotics of french gestures* Indiana Univ Pr.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics, 47*(1), 14-27.

Casey, S., & Emmorey, K. (2009). Co-speech gesture in bimodal bilinguals. *Language and Cognitive Processes, 24*(2), 290-312.

Cassell, J. (1988). Metapragmatics in language development: Evidence from speech and gesture. *Acta Linguistica Hungarica*, *38*(1-4), 3-18.

Cassell, J., McNeill, D., & McCullough, K. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & Cognition, 7*(1), 1-34.

Castelfranchi, C., & Poggi, I. (1998). *Bugie, finzioni, sotterfugi*. Roma: Carocci.

Chomsky, C. (1972). Stages in language development and reading exposure. *Harvard Educational Review*, *42*(1), 1-33.

Clark, H. H. (1996). *Using language* Cambridge university press.

Clark, R. A. (2003). *Generating synthetic pitch contours using prosodic structure* (Doctoral dissertation, Department of Linguistics, University of Edinburgh).

Clark, H. H., & Tree, J. E. F. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73-111.

Clark, R. A. (2003). *Generating synthetic pitch contours using prosodic structure* (Doctoral dissertation, Department of Linguistics, University of Edinburgh).

Cohen, A. A., & Harrison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology, 28*(2), 276.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106*(2), 1047-1058.

Cowley, S. J. (1998). Of timing, turn-taking, and conversations. *Journal of Psycholinguistic Research, 27*(5), 541-571.

Craib, I. (1978). Erving goffman: Frame analysis. *Philosophy of the Social Sciences, 8*(1), 79.

Crystal, D. (1976). *Prosodic systems and intonation in english.* CUP Archive.

Curl, T. S., Local, J., & Walker, G. (2006). Repetition and the prosody–pragmatics interface. *Journal of Pragmatics, 38*(10), 1721-1751.

Creider, C. A. (1986). Interlanguage comparisons in the study of the interactional use of gesture: Progress and prospects. *Semiotica*, *62*(1-2), 147-164.

Cruttenden, A. (1997). *Intonation*. Cambridge University Press: CUP

Davis, M., & Skupien, J. (1982). Body movement and nonverbal communication 1971–1981: An annotated bibliography.

De Fornel, M. (1992). The return gesture: Some remarks on context, inference, and iconic gesture.

De Geer, B. (1992). Internationally adopted children in communication. *Interaction, 2*, 2.

De Jorio, A. (1832). *La mimica degli antichi investigata nel gestire napoletano*. Dalla stamperia e cartiera del Fibreno.

De Ruiter, J. P. (1998). Gesture and speech production (Unpublished doctoral dissertation). Katholieke Universiteit, Nijmegen, the Netherlands.

De Ruiter, J. P. (2003) The function of hand gesture in spoken conversation. In M. Bickencach, A. Klappert & H. Pompe (Eds)*, Manus loquens*. Cologne: Dummont.

Dittmann, A. T., & Llewellyn, L. G. (1969). Body movement and speech rhythm in social conversation. *Journal of Personality and Social Psychology, 11*(2), 98.

Eckert, H., & Laver, J. (1994). *Menschen und ihre Stimmen: Aspekte der vokalen Kommunikation*. Weinheim: Psychologie Verlags Union.

Efron, D. (1941). *Gesture and environment*. King's crown Press: New York.

Efron, D., & van Veen, S. (1972). *Gesture, race and culture* Mouton The Hague.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*(4), 384.

Ekman, P. (2004). Emotional and conversational nonverbal signals. In *Language, knowledge, and representation* (pp. 39-50). Springer Netherlands.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*(1), 49-98.

Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology, 29*(3), 288.

Ellis, K. (1995). Apprehension, self-perceived competency, and teacher immediacy in the laboratory-supported public speaking course: Trends and relationships. *Communication Education, 44*(1), 64-78.

Emmorey, K., & Casey, S. (2002). Gesture, thought, and spatial language. *Spatial language* (pp. 87-101) Springer.

Essig, W. (2005). Storytelling: Effects of planning, repetition and context. *Teachers exploring tasks in english language teaching* (pp. 201-213) Springer.

Esteve-Gibert, N., Borràs-Comes, J., Swerts, M. G. J., & Prieto, P. (2014). Head gesture timing is constrained by prosodic structure.

Esteve-Gibert, N., Pons, F., Bosch, L., & Prieto, P. (2014). Are gesture and prosodic prominences always coordinated? evidence from perception and production. *Proceedings of the Speech Prosody Conference Dublin, Ireland,* pp. 222-226.

Esteve-Gibert, N., & Prieto, P. (2013). Intonational and gestural structures are temporally coordinated. Laboratory Phonology (LabPhon 13). Stuttgart, July 27-29.

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes, 22*(8), 1151-1177.

Feyereisen, P. (1987). Gestures and speech, interactions and separations: A reply to McNeill (1985).

Feyereisen, P., Van de Wiele, M., & Dubois, F. (1988). The meaning of gestures: What can be understood without speech. *European Bulletin of Cognitive Psychology, 8*(3-25), 27.

Feyereisen, P. (1997). The competition between gesture and speech production in dual-task paradigms. *Journal of memory and language*, *36*(1), 13-33.

Freedman, N. (1977). Hands, words, and mind: On the structuralization of body movements during discourse and the capacity for verbal representation. *Communicative structures and psychic structures* (pp. 109-132) Springer.

Freedman, N., & Hoffman, S. P. (1967). Kinetic behavior in altered clinical states: Approach to objective analysis of motor behavior during clinical interviews. *Perceptual and motor skills*, *24*(2), 527-539.

Freedman, N., van Meel, J. M., Barroso, F., & Bucci, W. (1986). On the development of communicative competence. *Semiotica*, *62*(1-2), 77-106.

Friesen, W. V., Ekman, P., & Wallbott, H. (1979). Measuring hand movements. *Journal of Nonverbal Behavior, 4*(2), 97-112.

Fromont, L., Biau, E., & Soto-Faraco, S. (2010). Taking prosody by the hand: The effect of pauses and gestures on sentence disambiguating. In *29th Conference on Generative Grammar*.

Galati, A., & Brennan, S. E. (2014). Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture. *Language, Cognition and Neuroscience*, *29*(4), 435-451.

Gass, S., Mackey, A., Alvarez-Torres, M. J., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning, 49*(4), 549-581.

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture*, *4*(2), 157-195.

Gerwing, J., & Dalby, A. M. L. (2014). Gestures convey content: An exploration of the semantic functions of physicians' gestures. *Patient education and counseling*, *96*(3), 308-314.

Gibbon, D. (1998). German Intonation. In D. J. Hirst & A. Di Christo (eds.), *Intonation systems: A survey of twenty languages* (pp. 78–95). Cambridge, MA: Cambridge University Press.

Gibbon, D. (2009). Gesture theory is linguistics: On modelling multimodality as prosody. *Paclic,* pp. 9-18.

Goffman, E. (1963). *Behavior in public places: Notes on the social organization of gatherings* Free Press New York.

Goffman, E. (1981). *Forms of talk* University of Pennsylvania Press.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences, 3*(11), 419-429.

Goldin-Meadow, S. (2000). Beyond words: The importance of gesture to researchers and learners. *Child development*, *71*(1), 231-239.

Goldin-Meadow, S., & Alibali, M. W. (1995). Learning with a helping hand. *Psychology of Learning and Motivation: Advances in Research and Theory, 33*, 115.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology, 64*, 257-283.

Goldin-Meadow, S., & Butcher, C. (2003). Pointing toward two-word speech in young children. *Pointing: Where Language, Culture, and Cognition Meet,* 85-107.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, *12*(6), 516-522.

Goldsmith, J. A. (1990). *Autosegmental and metrical phonology*, England: Basil Blackwell.

Goodrich Smith, W., & Hudson Kam, C. L. (2012). Knowing 'who she is' based on 'where she is': The effect of co-speech gesture on pronoun comprehension.

Goodwin, C. (1981). Conversational organization: Interaction between hearers and speakers. New York: Academic Press

Goodwin, C. (1986). Gestures as a resource for the organization of mutual orientation. *Semiotica*, *62*(1-2), 29-50.

Goodwin, M. H. (1980). Processes of mutual monitoring implicated in the production of description sequences. *Sociological Inquiry, 50*(3-4), 303-317.

Goodwin, C., & Goodwin, M. H. (1992). Context, activity and participation. *The Contextualization of Language,* 77-99.

Goodwin, C., & Heritage, J. (1990). Conversation analysis. *Annual Review of Anthropology, 19*, 283-307.

Goodwin, M.H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, *62*(1-2), 51-76.

Graham, C. (2013). Revisiting f0 range production in Japanese-English simultaneous bilinguals. Annual Report of UC Berkeley Phonology Lab, 110–125.

Graham, C. (2014). Fundamental frequency range in Japanese and English: The case of simultaneous bilinguals. *Phonetica*, *71*(4), 271-295.

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures (1). *International Journal of Psychology, 10*(1), 57-67.

Graham, J. A., & Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology, 5*(2), 189-195.

Geenen, J. G., & Koops, A. (2016). Co-speech gesture in spontaneous and planned speech. Congruence between production fluency and structural integrity. In *Proceedings of ICOM 8, 8th International Conference on Multimodality* (pp. 52-52). Cape Town, SA:[sn].

Graf, H. P., Cosatto, E., Strom, V., & Huang, F. J. (2002). Visual prosody: Facial movements accompanying speech. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on* (pp. 396-401). IEEE.

Gregersen, T., Olivares-Cuchat, G., & Storm, J. (2009). An examination of L1 and L2 gesture use: What role does proficiency play? *The Modern Language Journal, 93*(2), 195-208.

Greimas, A., Courtès, J., & Rengstorf, M. (1976). The cognitive dimension of narrative discourse. *New Literary History, 7*(3), 433-447.

Gullberg, M. (1998). Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish *Linguistics and Phonetics.* (Vol. 35).

Gullberg, M. (2003). Gestures, referents, and anaphoric linkage in learner varieties. *Information Structure and the Dynamics of Language Acquisition,*311-328.

Gullberg, M. (2006). Handling discourse: Gestures, reference tracking, and communication strategies in early L2. *Language Learning, 56*(1), 155-196.

Gullberg, M. (2008a). Gestures and second language acquisition. *Handbook of Cognitive Linguistics and Second Language Acquisition,*276-305.

Gullberg, M. (2008b). A helping hand? gestures, L2 learners, and grammar. *Gesture, Second Language Acquisition and Classroom Research. Applied Linguistics Series,*185-210.

Gunter, T. C. (2013). Gesture, speech, and communication: Exploring gesture style and covert gesture activation. In *Max Planck Institute for Psycholinguistics*.

Gunter, T. C., Weinbrenner, J. E., & Holle, H. (2015). Inconsistent use of gesture space during abstract pointing impairs language comprehension. *Frontiers in Psychology, 6*, 80.

Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, *115*(1-2), 147-172.

Hall, E. T. (1966). *The hidden dimension*. New York: Doubleday.

Hall, E. T., Birdwhistell, R. L., Bock, B., Bohannan, P., Diebold Jr, A. R., Durbin, M. & La Barre, W. (1968). Proxemics [and comments and replies]. *Current anthropology*, *9*(2/3), 83-108.

Halliday, M. A. K., & Greaves, W. S. (2008). *Intonation in the Grammar of English*. Equinox Pub.

Haviland, J. B. (1993). Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. Journal of linguistic anthropology, 3(1), 3-45.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies* University of Chicago Press.

Hayes, F. & Green, J. (1971). A Gesture Inventory for the Teaching of Spanish. *The Modern Language Journal*, 55(6), p.411.

Heath, C. (1992). Gesture's discreet tasks: Multiple relevancies in visual conduct and in the contextualisation of language. in *The Contextualization of Language*, Edited by Auer and Di Luzio p.101.

Hincks, R. (2004). Standard deviation of F0 in student monologue. *Proceedings of FONETIK 2004*, 132-135.

Hincks, R. (2005). Measuring liveliness in presentation speech. In *Interspeech,* pp. 765-768.

Hirschberg, J. B., & Rosenberg, A. (2005). Acoustic/prosodic and lexical correlates of charismatic speech. In *INTERSPEECH* (pp. 513-516).

Hoetjes, M., Krahmer, E., & Swerts, M. (2009, August). Untying the knot between gestures and speech. In *AVSP* (pp. 90-95).

Hoetjes, M., Krahmer, E., & Swerts, M. (2014). Does our speech change when we cannot gesture?. *Speech Communication*, *57*, 257-267.

Hoetjes, M., Krahmer, E., & Swerts, M. (2015). On what happens in gesture when communication is unsuccessful. *Speech Communication*, *72*, 160-175.

Hoetjes, M. (2016). Gestures become more informative when communication is unsuccessful. In Proceedings of MMSYM 2016, European Symposium Series on Multimodal Communication (p. electr).

Holler, J., & Beattie, G. (2002). A micro-analytic investigation of how iconic gestures and speech represent core semantic features in talk. *Semiotica-La Haye then Berlin-, 142*(1/4), 31-70.

Holler, J., & Beattie, G. (2003). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener? *Gesture, 3*(2), 127-154.

Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior, 33*(2), 73-88.

Holler, J., Tutton, M., & Wilkin, K. (2011). Co-speech gestures in the process of meaning coordination. In *2nd GESPIN–Gesture & Speech in Interaction Conference, Bielefeld*.

Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and cognitive processes*, *24*(2), 267-289.

Holler, J., & Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior*, *35*(2), 133-153.

Hopper, P. J. (1979). Aspect and foregrounding in discourse in discourse and syntax. *Syntax and Semantics Ann Arbor, Mich., 12*, 213-241.

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, pp. 251-299.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*(1), 91-117.

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, *22*(3), 313-336.

Hubbard, A. L., Wilson, S. M., Callan, D. E., & Dapretto, M. (2009). Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapping, 30*(3), 1028-1037.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*(5), 367-371.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*(2), 291-303.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied linguistics*, *23*(1), 83-103.

Jenner, B. R. (1976). Interlanguage and foreign accent. *Interlanguage studies bulletin*, 166-195.

Johns-Lewis, C. (1986). Prosodic differentiation of discourse modes. In C. Johns-Lewis, (Ed.), *Intonation in discourse*, Breckenham, Kent: Croom Helm, 199-220.

Jun, S. (2006). *Prosodic typology: The phonology of intonation and phrasing* Oxford University Press on Demand.

Jügler, J., Zimmerer, F., Trouvain, J., & Möbius, B. (2016). The Perceptual Effect of L1 Prosody Transplantation on L2 Speech: The Case of French Accented German. In INTERSPEECH (pp. 67-71).

Kaulfers, W. V. (1931). Curiosities of colloquial gesture. *Hispania, 14*(4), 249-264.

Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language, 40*(4), 577-592.

Kendon, A. (1972). Some relationships between body motion and speech. *Studies in Dyadic Communication, 7*(177), 90.

Kendon, A. (1973). The role of visible behaviour in the organization of social interaction. In M. Von Cranach and I. Vine, editors, *Social Communication and Movement: Studies of Interaction and Expression in Man and Chimpanzee*. London: Academic Press, pp.29-74.

Kendon, A. (1978). Differential perception and attentional frame in face-to-face interaction: Two problems for investigation. *Semiotica*, *24*(3-4), 305-316.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. *The Relationship of Verbal and Nonverbal Communication, 25*, 207-227.

Kendon, A. (1985). Some uses of gesture. *Perspectives on Silence,* , 215-234.

Kendon, A. (1988). How gestures can become like words. *Cross-Cultural Perspectives in Nonverbal Communication, 1*, 131-141.

Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction, 27*(3), 175-200.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in southern italian conversation. *Journal of Pragmatics, 23*(3), 247-279.

Kendon, A. (1997). Gesture. *Annual Review of Anthropology,* 109-128.

Kendon, A. (2000). Language and gesture: Unity or duality. *Language and Gesture*

Kendon, A. (2004). *Gesture: Visible action as utterance* Cambridge University Press.

Kendon, A. (2011). Vocalisation, speech, gesture, and the language origins debate: An essay review on recent contributions. *Gesture*, *11*(3), 349-370.

Kendon, Adam (2013). History of the Study of Gesture. Chapter 3 in Keith Allan, ed. *Oxford Handbook for the History of Linguistics*. Oxford: Oxford University Press.

Key, M. R. (1980). *The relationship of verbal and nonverbal communication* Walter de Gruyter.

Kita, S. (1990). The temporal relationship between gesture and speech: A study of japanese-english bilinguals. *MS, Department of Psychology, University of Chicago, 90*, 91-94.

Kita, S. (2000). How representational gestures help speaking. *Language and Gesture,* 162-185.

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes, 24*(2), 145-167.

Kita, S., & Özyürek, A. (2007). *How does spoken language shape iconic gestures* John Benjamins Publishing Company, Amsterdam and Philadelphia, PA.

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*.

Komatsu, M. (2002). What constitutes acoustic evidence of prosody? the use of linear predictive coding residual signal in perceptual language identification. *The LACUS Forum, , 28.* pp. 277-286.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*(3), 396-414.

Krahmer, E., & Swerts, M. (2009). Audiovisual prosody—introduction to the special issue.

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in Experimental Social Psychology, 28*, 389-450.

Krauss, R.M., & Chi-Yue C. (1998). Language and social psychology. in *Handbook of social psychology*. Edited by D. Gilbert, S. Fiske-Emory, and G. Lindzey, vol. 2, pp. 41–88. New York: Guilford. [cc]

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gesture. *Journal of Experimental Social Psychology*, *31*(6), 533-552.

Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. *Gesture, Speech, and Sign, 93*

Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology, 61*(5), 743.

Kuhlen, A. K., Galati, A., & Brennan, S. E. (2012). Gesturing integrates top-down and bottom-up information: Joint effects of speakers' expectations and addressees' feedback. *Language and Cognition*, *4*(1), 17-41.

Ladefoged, P. (2003). *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques/Peter Ladefoged*. Oxford: Blackwell.

Ladewig, S. H. (2011). Putting the cyclic gesture on a cognitive basis. *CogniTextes.Revue De l'Association Française De Linguistique Cognitive,* (Volume 6)

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

Larsen-Freeman, D. (1991). Second language acquisition research: Staking out the territory. *TESOL Quarterly, 25*(2), 315-350.

Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö., Faria, V., et al. (2008). In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *Journal of Nonverbal Behavior, 32*(4), 195-214.

Lešić, B. P. (2010). The metaphorization of practical action and everyday life experience in the words, emblems, and coverbal gestures of spoken language. *Studia Romanica Et Anglica Zagrabiensia, 55*, 223-233.

Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System, 32*(4), 505-524.

Levy, E. T., & McNeill, D. (1992). Speech, gesture, and discourse. *Discourse processes*, *15*(3), 277-301.

Levy, E. T., & McNeill, D. (2015). *Narrative development in young children: Gesture, imagery, and cohesion*. Cambridge University Press.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic inquiry*, *8*(2), 249-336.

Lock, A. E. (1978). *Action, gesture and symbol: The emergence of language.* Academic Press.

Lock, A., & Young, A. Service, V., & chandler, P.(1990). some observations on the origins of the pointing gesture. *From Gesture to Language in Hearing and Deaf Children,* 42-55.

Loehr, D. P. (2004). *Gesture and intonation* (Doctoral dissertation, Georgetown University).

Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture, 7*(2), 179-214.

Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology, 3*(1), 71-89.

Lucero, C., Zaharchuk, H., & Casasanto, D. (2014). Beat gestures facilitate speech production. *Proceedings of the 36th Annual Conference of the Cognitive Science Society,* pp. 898-903.

Lyons, A., & Kashima, Y. (2006). Maintaining stereotypes in communication: Investigating memory biases and coherence-seeking in storytelling. *Asian Journal of Social Psychology, 9*(1), 59-71.

Mahl, G. F. (1968). Gestures and body movements in interviews. *Research in Psychotherapy Conference, 3rd, May-Jun, 1966, Chicago, IL, US,*

Mandel, M. (1977). Iconic devices in american sign language. *On the Other Hand: New Perspectives on American Sign Language*, 57-107.

Marcos, L. R. (1979). Nonverbal behavior and thought processing. *Archives of general psychiatry*, *36*(9), 940-943.

Marslen-Wilson, W., Levy, E., & Tyler, L. K. (1982). Producing interpretable discourse: The establishment and maintenance of reference. *Speech, place, and action*, 339-378.

Masur, E. F. (1990). Gestural development, dual-directional signaling, and the transition to words. *From gesture to language in hearing and deaf children* (pp. 18-30) Springer.

McCafferty, S. G. (1998). Nonverbal expression and L2 private speech. *Applied Linguistics, 19*(1), 73-96.

McCafferty, S. G. (2004). Space for cognition: Gesture and second language learning. *International Journal of Applied Linguistics, 14*(1), 148-165.

McClave, E. Z. (1991). Intonation and gesture (Doctoral dissertation, Georgetown University).

McClave, E. Z. (1998). Pitch and manual gestures. *Journal of Psycholinguistic Research, 27*(1), 69-89.

McKinnon, S., & Prieto, P. (2014). The role of Prosody and Gesture in the Perception of Mock Impoliteness. *Journal of Politeness Research*, *10*(2), 185-219.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92*(3), 350.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press. Chicago London

McNeill, D. (2000). 15 catchments and contexts: Non-modular factors in speech and gesture production. *Language and Gesture, 2*, 312.

McNeill, D. (2000). *Language and gesture*. Cambridge University Press.

McNeill, D. (2005). Gesture, gaze and Ground. In *International Workshop on Machine Learning for Multimodal Interaction* (pp.1-14). Springer Berlin Heidelberg

McNeill, D. (2015). *Why we gesture: The surprising role of hand movements in communication*. Cambridge University Press.

McNeill, D., & Levy, E. (1980). *Conceptual representations in language activity and gesture*. ERIC Clearinghouse.

McNeill, D., & Levy, E. T. (1993). Cohesion and gesture. *Discourse processes*, *16*(4), 363-386.

McNeill, D., Pedelty, L. L., & Levy, E. T. (1990). Speech and gesture. *Advances in Psychology, 70*, 203-256.

McNeill, D., Quek, F., McCullough, K., Duncan, S., Furuyama, N., Bryll, R. (2001). Catchments, prosody and discourse. *Gesture, 1*(1), 9-33.

Mehrabian, A. (1972). *Nonverbal communication*. Chicago, IL: Aldine-Atherton.

Melinger, A., & Levelt, W. J. (2004). Gesture and the communicative intention of the speaker. *Gesture, 4*(2), 119-141.

Mennen, I., Schaeffler, F., & Docherty, G. (2007). Pitching it differently: A comparison of the pitch ranges of german and english speakers. *16th International Congress of Phonetic Sciences,* pp. 1769-1772.

Mennen, I., Schaeffler, F., & Docherty, G. (2008). A methodological study into the linguistic dimensions of pitch range differences between German and English. In *Proceedings of the 4th Conference on Speech Prosody, Campinas*.

Mennen, I., Chen, A., & Karlsson, F. (2010). Characterising the internal structure of learner intonation and its development over time. *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds,* pp. 319-324.

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America, 131*(3), 2249-2260.

Moerman, M. (1990). Exploring talk and interaction. *Research on Language & Social Interaction, 24*(1-4), 173-187.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). Communicative gestures and memory load. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society. Austin, TX*.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2011). Seeing and being seen: the effects on gesture production. *Journal of Computer-Mediated Communication*, *17*(1), 77-100.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 615.

Mozziconacci, S. (2002) Prosody and emotions. *Speech Prosody 2002, International Conference*

Nicoladis, E. (1999). Gesture and early bilingual development. *Developmental Psychology, 35*(2), 514-26.

Nicoladis, E., Pika, S., Yin, H., & Marentette, P. (2007). Gesture use in story recall by Chinese–English bilinguals. *Applied Psycholinguistics, 28*(04), 721-735.

Nobe, S. (1993). Cognitive process of speaking and gesturing: A comparison between first language speakers and foreign language speakers. *Unpublished master's thesis, University of Chicago*.

Nobe, S. (1996). *Representational gestures, cognitive rhythms, and acoustic aspects of speech: a network/threshold model of gesture production* (Doctoral dissertation, University of Chicago, Department of Psychology).

Norman, S. J., Blackwell, A. F., Warren, L., & Woolford, K. (2010). Gesture and embodied interaction: Capturing motion / data / value. *Leonardo, 43*(2), 198-199.

Okada, T., & Brosnahan, L. (1990). *Japanese and English gesture: Contrastive nonverbal communication*. Tokyo: Taishukan Press.

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? the effects of addressee location on representational gestures. *Journal of Memory and Language, 46*(4), 688-704.

Özyürek, A., & Kita, S. (1999). Expressing manner and path in english and turkish: Differences in speech, gesture, and conceptualization. *Twenty-First Annual Conference of the Cognitive Science Society,* pp. 507-512.

Özyürek, A., Kita, S., Allen, S. E., Furman, R., & Brown, A. (2005). How does linguistic framing of events influence co-speech gestures? Insights from crosslinguistic variations and similarities. *Gesture, 5*(1-2), 219-240.

Parrill, F.,Cabot, J., Kent, H., Chen, K., and Payneau, A. (2016). Do people gesture more when instructed to? *Gesture*, 15, 3, 357-371.

Patterson, D. J. (2000). *Linguistic approach to pitch range modelling* (Doctoral dissertation, Edinburgh University).

Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes*, *23*(1), 19-43.

Pika, S., Nicoladis, E., & Marentette, P. F. (2006). A cross-cultural study on the use of gestures: Evidence for cross-linguistic transfer?. *Bilingualism: Language and Cognition*, *9*(03), 319-327.

Perniss, P., & Özyürek, A. (2015). Visible Cohesion: A Comparison of Reference Tracking in Sign, Speech, and Co-Speech Gesture. *Topics in Cognitive Science*, *7*(1), 36-60.

Pettorino, M. (2003). La velocità di articolazione. *Atti delle XIV Giornate di Studio del Gruppo di Fonetica Sperimentale-GFS.*

Pizzuto, E. A., & Capobianco, M. (2005). The link (and differences) between deixis and symbols in children's early gestural-vocal system. *Gesture, 5*(1), 179-199.

Polyanskaya, L., Ordin, M., & Busa, M. G. (2016). Relative Salience of Speech Rhythm and Speech Rate on Perceived Foreign Accent in a Second Language. *Language and Speech*, 0023830916648720.

Poggi, I. (1987). *Le parole nella testa*. Bologna: Il Mulino.

Poggi, I., & Magno Caldognetto, E. (1997). *Mani che parlano*. Padova: Unipress.

Prieto, P., Mücke, D., Becker, J., & Grice, M. (2007). Coordination patterns between pitch movements and oral gestures in catalan. *Proceedings of the XVIth International Congress of Phonetic Sciences, Pirrot GmbH: Dudweiler,* pp. 989-992.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265-292.

Rangarajan, V., & Narayanan, S. (2006). Analysis of disfluent repetitions in spontaneous speech recognition. *Signal Processing Conference, 2006 14th European,* pp. 1-5.

Raphael, L. J., Borden, G. J., & Harris, K. S. (2007). *Speech science primer: Physiology, acoustics, and perception of speech* Lippincott Williams & Wilkins.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*(4), 226-231.

Ravizza, S. (2003). Movement and lexical access: Do noniconic gestures aid in retrieval? *Psychonomic Bulletin & Review, 10*(3), 610-615.

Ricci Bitti, P. E., & Poggi, I. (1991). Symbolic nonverbal behavior: Talking through gestures. *Fundamentals of Nonverbal Behavior*, 433-457.

Rimé, B. (1982). The elimination of visible behaviour from social interactions: Effects on verbal, nonverbal and interpersonal variables. *European journal of social psychology*, *12*(2), 113-129.

Rimé, B. and Schiaratura, L. (1991). Gesture and speech. In Feldman, R. S. and Rim, B., editors, Fundamentals of Nonverbal Behaviour, 239–281. New York: Press Syndacate of the University of Cambridge.

Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior, 5*(3), 172-183.

Rochet-Capellan, A., Laboissière, R., Galván, A., & Schwartz, J. L. (2008). The speech focus position effect on jaw–finger coordination in a pointing task. *Journal of Speech, Language, and Hearing Research*, *51*(6), 1507-1521.

Rogers, W. T. (1978). The contribution of kinesic illustrators toward the comprehension of verbal behavior within utterances. *Human Communication Research*, *5*(1), 54-62.

Roth, W. M. (2001). Gestures: Their role in teaching and learning. *Review of educational research*, *71*(3), 365-392.

Roth, W. (2002). From action to discourse: The bridging function of gestures. *Cognitive Systems Research, 3*(3), 535-554.

Roustan, B., & Dohen, M. (2010). Co-production of contrastive prosodic focus and manual gestures: Temporal coordination and effects on the acoustic and articulatory correlates of focus. In *Speech Prosody 2010-5th International Conference on Speech Prosody* (pp. 100-110).

Ruthrof, H. (2016). Speculations on the origins of language. *Linguistic and Philosophical Investigations*, (15), 90-107.

Sacks, H., & Schegloff, E. A. (2002). Home position. *Gesture*, *2*(2), 133-146.

Sainsbury, P., & Wood, E. (1977). Measuring gesture: Its cultural and clinical correlates. *Psychological medicine*.

Saitz, R. L. (1966). Gestures in the language classroom. *English Language Teaching, 21*, 33-37.

Selting, M. (1995). *Prosodie im Gespräch: Aspekte einer interaktionalen Phonologie der Konversation* (Vol. 329). Walter de Gruyter.

Sharma, R., Yeasin, M., & Kettebekov, S. (2008). *Prosody Based audio/visual Co-Analysis for Co-Verbal Gesture Recognition,*

Sherman, J., & Nicoladis, E. (2004). Gestures by advanced Spanish-English second-language learners. *Gesture, 4*(2), 143-156.

Silverman, M. J. (2010). The effect of pitch, rhythm, and familiarity on working memory and anxiety as measured by digit recall performance. *Journal of Music Therapy, 47*(1), 70-83.

Singer, M. A., & Coldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science, 16*(2), 85-89.

Si-Qing, C. (1990). A study of communication strategies in interlanguage production by chinese EFL learners. *Language Learning, 40*(2), 155-187.

So, W. C. (2010). Cross-cultural transfer in gesture frequency in Chinese–English bilinguals. *Language and Cognitive Processes, 25*(10), 1335-1353.

Stam, G. (2006). Thinking for speaking about motion: L1 and L2 speech and gesture. *IRAL-International Review of Applied Linguistics in Language Teaching, 44*(2), 145-171.

Stam, G. (2008). What gestures reveal about second language acquisition. *Gesture: Second Language Acquistion and Classroom Research,* 231.

Stam, G. (2016). Gesture as a window onto conceptualization in multiple tasks: Implications for second language teaching. *Yearbook of the German Cognitive Linguistics Association*, *4*(1), 289.

Stella, A., and Busà, M.G., Transfer intonativo in inglese L2: analisi fonetico-fonologica delle produzioni di parlanti padovani. In: Vincenzo Galatà (a cura di), *Multimodalità e Multilingualità. La sfida più avanzata della comunicazione orale. Saggi dal IX Convegno Nazionale AISV - Associazione Italiana Scienza della Voce*. Roma: Bulzoni, 395-410.

Stone, G. R. (1974). Fora americana: Speech in public places. *The Supreme Court Review,* 233-280.

Streeck, J. (1993). Gesture as communication I: Its coordination with gaze and speech. *Communications Monographs*, *60*(4), 275-299.

Streeck, J., & Hartge, U. (1992). Previews: Gestures at the transition place. *The Contextualization of Language,* , 135-157.

Streeck, J., & Knapp, M. L. (1992). The interaction of visual and verbal features in human communication. *Advances in Nonverbal Communication, 10*, 3-23.

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture, 8*(2), 219-235.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78*(3), 705-722.

Traunmüller, H., & Eriksson, A. (1995). The perceptual evaluation of F 0 excursions in speech as evidenced in liveliness estimations. *The Journal of the Acoustical Society of America*, *97*(3), 1905-1915.

Tree, J. E. F. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34*(6), 709-738.

Trouvain, J. (2014). Laughing, breathing clicking—The prosody of nonverbal vocalisations. *Proceedings of Speech Prosody (SP7), Dublin,* 598-602.

Tsao, Y., Weismer, G., & Iqbal, K. (2006). Interspeaker variation in habitual speaking rate: Additional evidence. *Journal of Speech, Language, and Hearing Research, 49*(5), 1156-1164.

Tuite, K., (1993). The production of gesture. *Semiotica 1,2*, 83-105.

Urbani, M. (2013). The pitch range of Italians and Americans. a comparative study. (*Doctoral dissertation)*

Vaissière, J. (1997). Langues, prosodies et syntaxe: Prosodie et syntaxe. *TAL. Traitement automatique des langues*, *38*(1), 53-82.

Valbonesi, L., Ansari, R., McNeill, D., Quek, F., Duncan, S., McCullough, K., et al. (2002). Multimodal signal analysis of prosody and hand motion: Temporal correlation of speech and gestures. *Signal Processing Conference, 2002 11th European,* pp. 1-4.

Valokorpi, P. (1981). Non-verbal elements in foreign language communication. *Paper presented AILA, Lund, Sweden*.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and speech*, *38*(3), 253-265.

Vihman, M. M. (1991). Ontogeny of phonetic gestures: Speech production. *Modularity and the Motor Theory of Speech Perception: Proceedings of a Conference to Honor Alvin M. Liberman,* pp. 69-84.

Vincze, L., & Poggi, I. (2013, December). Multimodal Markers of One's Own and the Other's Vagueness. In *Proceedings of the Workshop AI HCI, Turin*.

Vygotsky, L. S. (1962). Language and thought. *Massachusetts Institute of Technology Press, Ontario, Canada.*

Vygotsky, L. S. (1986). *Thought and language, revised edition.* (A. Kozulin, Ed.). Cambridge, MA: MIT Press

Ullakonoja, R. (2007). Comparison of pitch range in Finnish (L1) and Russian (L2). Proc. 16* ICPhS, Saarbrücken, 1701-1704

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, *57*, 209-232.

Willems, N. (1982). *English intonation from a Dutch point of view*. Walter de Gruyter. Berlin

Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex (New York, N.Y.: 1991), 17*(10), 2322-2333.

Wells, J. C. (2006). *English Intonation.* Cambridge University Press.

Wessel-Tolvig, B. (2016, September). Getting it right: Advanced Danish learners of Italian acquire speech and gesture L2 forms. In *Proceedings from the 3rd European Symposium on Multimodal Communication, Dublin, September 17-18, 2015* (No. 105, pp. 75-81). Linköping University Electronic Press.

Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language, 101*(3), 234-245.

Wundt, W. (1973). The language of gestures (JS thayer, CM greenleaf, & MD silberman, trans.). *The Hague: Mouton*

Zellner, B. (1994). Pauses and the temporal structure of speech. In *Zellner, B.(1994). Pauses and the temporal structure of speech, in E. Keller (Ed.) Fundamentals of speech synthesis and speech recognition.(pp. 41-62). Chichester: John Wiley.* (pp. 41-62). John Wiley.

Zellers, M., House, D., & Alexanderson, S. (2016). Prosody and hand gesture at turn boundaries in Swedish. In *8th Speech Prosody 2016, 31 May 2016 through 3 June 2016* (pp. 831-835). International Speech Communications Association.

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B., & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and german speakers. *Proc. Speech Prosody, 7.* pp. 1037-1041.

Appendix 1

The Fox and the Crow

Aesop

Once upon a time there was a crow, and the crow was hungry and the crew around and around and around, looking for food.

Suddenly, the crow spotted a piece of cheese on the shelf of a window. So, the crow flew down and picked up the cheese in its beak, and then flew a bit more until it found the highest possible tree. It went to the very top of the tree where it planned to eat its stolen catch.

But walking along the ground came a fox, and that fox looked up and saw the cheese and thought: 'I want that'.

And so, the fox stopped and looked deeply at the crow and said:

'Crow! Oh, your feathers are so beautiful! I bet that that your voice is as beautiful as your feathers!

The crow popped itself up a bit and looked and shook its head, and the fox asked:

'Will you sing for me?'

So flattered, the crow with the cheese in its beak began to sing:

'craa, craa, craa'

But the cheese did not fall, so the fox said:

'No, no, no, I need to hear that beautiful voice, the one that goes with those feathers. Won't you sing loud for me?'.

And with that, the crow popped itself up, opened its beak to sing loud, and when it did so, the cheese fell to the ground. When that cheese started to fall, that fox jumped up and caught it in its jaws, and walked away.

Appendix 2

Gesture occurrences – experiment 1

STUD A	Representational	Discursive	Emblems	TOT	%dep.	%disc.	%emblems
Ita t1	9	14	0	23	39,13	60,87	0,00
Eng t1	9	31	1	41	21,95	75,61	2,44
Eng t2	15	37	1	53	28,30	69,81	1,89
TOT	*33*	*82*	*2*				

STUD B	Representational	Discursive	Emblems	TOT	%dep.	%disc.	%emblems
Ita t1	5	28	1	34	14,71	82,35	2,94
Eng t1	10	27	1	38	26,32	71,05	2,63
Eng t2	17	22	2	41	41,46	53,66	4,88
TOT	*32*	*77*	*4*				

STUD D	Representational	Discursive	Emblems	TOT	%dep.	%disc.	%emblems
Ita t1	1	20	0	21	4,76	95,24	0,00
Eng t1	3	21	0	24	12,50	87,50	0,00
Eng t2	15	23	1	39	38,46	58,97	2,56
TOT	*19*	*64*	*1*				

STUD E	Representational	Discursive	Emblems	TOT	%dep.	%disc.	%emblems
Ita t1	7	36	0	43	16,28	83,72	0,00
Eng t1	1	40	0	41	2,44	97,56	0,00
Eng t2	10	40	0	50	20,00	80,00	0,00
TOT	*18*	*116*	*0*				

STUD F	Representational	Discursive	Emblems	TOT	%dep.	%disc.	%emblems
Ita t1	16	25	0	41	39,02	60,98	0,00
Eng t1	13	48	0	61	21,31	78,69	0,00
Eng t2	31	27	7	65	47,69	41,54	10,77
TOT	*60*	*100*	*7*				

STUD I	Representational	Discursive	Emblems	TOT	%dep.	%disc.	%emblems

Ita t1	28	32	1	61	45,90	52,46	1,64
Eng t1	1	12	0	13	7,69	92,31	0,00
Eng t2	21	33	0	54	38,89	61,11	0,00
TOT	*50*	*77*	*1*				

STUD J	**Represent ational**	**Discursive**	**Emblems**	**TOT**	**%dep.**	**%disc.**	**%emblems**
Ita t1	8	40	0	48	16,67	83,33	0,00
Eng t1	18	31	0	49	36,73	63,27	0,00
Eng t2	19	39	0	58	32,76	67,24	0,00
TOT	*45*	*110*	*0*				

STUD K	**Represent ational**	**Discursive**	**Emblems**	**TOT**	**%dep.**	**%disc.**	**%emblems**
Ita t1	0	9	0	9	0,00	100,00	0,00
Eng t1	1	8	0	9	11,11	88,89	0,00
Eng t2	6	28	0	34	17,65	82,35	0,00
TOT	*7*	*45*	*0*	*52*	13,46	86,54	0,00

Appendix 3

Summary of the speech data for each speaker – experiment 1 (Chapter 5)

Italian t1								
	stud A	**stud B**	**stud D**	**stud E**	**stud F**	**stud I**	**stud J**	**stud k**
Tot. time	73,639	74,999	82,556	78,642	90,917	105,877	85,801	87
Time pauses	6,376	7,033	11,976	7,253	8,077	15,13	5,634	11,646
Articulated time	67,263	67,966	70,58	71,389	82,84	90,747	80,167	75,354
words	154	239	168	211	234	256	202	205
N. pauses	16	19	31	19	26	32	16	27
mean pitch	185,95	219,45	175,16	211,37	197,43	228,1	188,41	204,47
min pitch	76,26	77,7	77,7	76,55	75,83	81,79	80,54	78,85
max pitch	491,51	491,81	436,25	357,04	493,68	438,92	397,27	457,52

English t1								
	stud A	**stud B**	**stud D**	**stud E**	**stud F**	**stud I**	**stud J**	**stud k**
Tot. time	124,919	75,401	85,918	80,043	126,164	110,279	75,919	91,202
Time pauses	14,036	10,557	14,108	10,894	15,407	18,104	8,996	10,841
Articulated time	110,883	64,844	71,81	69,149	110,757	92,175	66,923	80,361
words	175	177	129	197	233	242	204	161
N. pauses	38	25	42	31	44	46	30	30
mean pitch	188,52	218,8	182,27	212,77	149,23	233,99	197,27	213,19
min pitch	75,43	77,03	75,12	85,08	76,3	90,69	82,62	77,43
max pitch	370,97	490,33	357,03	382	499,99	489,85	449,48	421,22

English t2								
	stud A	**stud B**	**stud D**	**stud E**	**stud F**	**stud I**	**stud J**	**stud k**
Tot. time	100,038	76,398	94,721	103,6	102,161	85,198	96,16	80,84
Time pauses	14,161	9,957	14,601	15,846	12,634	13,661	13,532	12,269

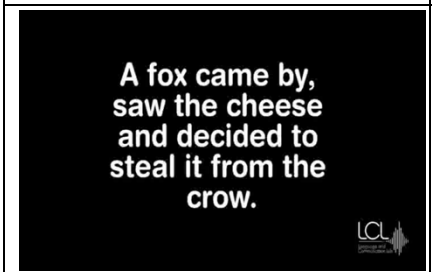
Articulated time	85,877	66,441	80,12	87,754	89,527	71,537	82,628	68,571
words	191	228	163	266	277	225	245	184
N. pauses	40	30	38	44	39	36	40	33
mean pitch	203,93	257,84	195,07	236,67	241,45	250,1	213,64	227,09
min pitch	75,5	77,37	75,85	77,01	77,77	75,02	75,86	79,18
max pitch	486,92	499,57	349,93	490,99	499,04	486,37	488,4	499,94

Appendix 4

The Cartoon Stimuli – 2nd experiment (Chap.6)

Screen captions of the written description

1	2	3
		

4	5	6
		

7	8	9
		

Screen captions of the cartoon

1	2
	

3	4
	

5	6
	

7	8
	

9	10
	

Appendix 5

Gesture data in experiment 2 (Chapter 6)

Eng t1	**A**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**L**	**M**
Beat	15	16	3	0	33	20	17	47	24	17
vagueness	6	6	2	0	18	5	2	1	7	7
Pointing	0	0	0	0	0	0	0	0	0	0
Metaphoric	6	0	1	0	0	5	0	3	8	2
Iconic	11	6	8	0	13	16	11	4	20	10
Abs. point.	1	0	0	0	0	0	0	2	3	1
Emblem	0	0	0	0	1	0	0	0	0	0

Eng t2	**A**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**L**	**M**
Beat	26	12	14	5	44	17	14	32	21	18
vagueness	5	4	3	0	7	3	3	2	7	3
Pointing	0	0	0	0	0	0	0	1	0	0
Metaphoric	9	0	2	0	0	0	0	5	8	0
Iconic	19	10	8	0	14	6	8	15	21	5
Abs. point.	2	0	0	0	0	0	0	0	3	2
Emblem	0	0	0	0	0	0	0	0	0	0

Ita t1	**A**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**L**	**M**
Beat	26	13	1	2	19	21	14	36	19	0
vagueness	2	2	0	0	3	4	2	3	4	1
Pointing	0	0	0	0	0	0	0	0	0	0
Metaphoric	9	0	2	0	1	3	1	3	1	1
Iconic	3	6	0	0	10	13	5	7	10	1
Abs. point.	0	0	0	0	0	0	0	0	0	0
Emblem	0	0	0	0	0	0	0	0	0	0

Ita t2	**A**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**L**	**M**
Beat	20	16	14	12	22	17	16	36	14	16
vagueness	0	2	3	0	2	3	2	1	2	5
Pointing	0	0	0	0	0	0	0	0	0	0
Metaphoric	7	0	0	0	0	1	5	4	3	0
Iconic	10	4	9	5	9	8	4	18	10	3
Abs. point.	0	0	0	0	0	0	0	3	0	0
Emblem	0	0	0	0	0	0	0	0	0	0

Appendix 6

Gesture data in experiment 2 (Chapter 6)

Number of occurrences of speech features in the four conditions for each speaker

Chunks: the number of speech flows between two pauses
Ps: Silent Pauses
Pp: Filled Pauses
Pr: Respiratory Pauses
Tot p: Total number of pauses
Corr: Corrections
Dis: Disfluencies
Rep: Repetitions
Tot: Total number of disfluencies

Occurrences - Italian t1								
	chunks	**ps**	**pp**	**pr**	**tot p.**	**corr**	**dis**	**rep**	**tot**
M_ita 1	20	9	6	7	22	1	15	6	22
L_ita1	14	2	2	10	14	0	13	1	14
I_ita1	22	10	3	10	23	0	8	3	11
H_ita1	10	2	0	7	9	1	6	2	9
G_ita1	22	9	4	12	25	0	12	1	13
F_ita1	15	1	4	11	16	0	13	6	19
E_ita1	4	0	0	3	3	1	2	0	3
D_ita1	30	10	12	16	38	0	15	0	15
C_ita1	10	3	3	4	10	1	2	0	3
A_ita1	16	2	6	11	19	0	12	0	12

Occurrences - Italian t2								
	chunks	**ps**	**pp**	**pr**	**tot p.**	**corr**	**dis**	**rep**	**tot**
M_ita 2	15	2	6	8	31	2	7	5	14
L_ita2	11	1	2	8	11	0	9	1	10
I_ita2	18	8	3	10	21	0	5	0	5
H_ita2	15	1	4	9	14	2	7	0	9
G_ita2	16	3	1	12	16	1	12	8	21
F_ita2	12	3	2	7	12	2	8	5	15
E_ita2	5	0	1	4	5	2	1	0	3
D_ita2	12	3	3	7	13	0	15	1	16
C_ita2	12	4	3	5	12	0	7	0	7
A_ita2	11	2	1	8	11	0	1	0	1

Occurrences - English t1								
	chunks	**ps**	**pp**	**pr**	**tot**	**corr**	**dis**	**rep**	**tot**
M_eng1	34	21	7	1	63	5	21	4	30
L_eng1	41	19	9	24	52	1	36	5	42

I_eng1	35	15	11	18	44	2	14	0	16
H_eng1	41	22	7	19	48	2	36	8	46
G_eng1	36	12	15	12	39	3	27	13	43
F_eng1	44	14	26	23	63	2	40	12	54
E_eng1	14	4	4	6	14	3	6	0	9
D_eng1	40	18	18	15	51	2	26	5	33
C_eng1	22	4	12	9	25	0	20	1	21
A_eng1	25	11	9	14	34	4	11	0	15

Occurrences - English t2

	chunks	**ps**	**pp**	**pr**	**tot**	**corr**	**dis**	**rep**	**tot**
M_eng2	28	14	18	9	69	1	22	8	31
L_eng2	37	11	18	14	43	1	29	2	32
I_eng2	35	20	7	15	42	0	15	0	15
H_eng2	19	3	3	13	19	2	12	1	15
G_eng2	21	4	7	14	25	1	19	6	26
F_eng2	34	7	18	19	44	2	28	13	43
E_eng2	9	2	0	5	7	1	2	1	4
D_eng2	35	19	13	13	45	2	28	7	37
C_eng2	14	1	7	9	17	2	9	3	14
A_eng2	26	10	5	13	28	5	11	1	17

Duration (in seconds) of speech features in the four conditions for each speaker

Art sp: articulated speech
Ps: Silent Pauses
Pp: Filled Pauses
Pr: Respiratory Pauses
Tot p: Total duration of pauses
Corr: Corrections
Dis: Disfluencies
Rep: Repetitions
Tot: Total duration of disfluencies

Duration Ita1 in seconds

	Art. sp.	**ps**	**pp**	**pr**	**Tot p.**	**T. story**	**corr**	**dis**	**rep**	**tot**
M_ita 1	45,78	3,97	2,85	4,88	11,70	57,48	0,24	8,42	1,22	9,88
L_ita1	49,41	0,63	1,57	4,61	6,80	56,21	0,00	5,23	0,15	5,38
I_ita1	45,16	1,39	1,60	4,53	7,51	52,67	0,00	3,21	0,75	3,95
H_ita1	29,27	0,32	0,00	3,27	3,59	32,86	0,43	1,89	0,33	2,64

G_ita1	55,54	3,48	1,85	7,07	12,40	67,95	0,00	4,25	0,32	4,57
F_ita1	47,27	0,55	1,98	4,86	7,40	54,67	0,00	4,78	0,86	5,63
E_ita1	23,35	0,00	0,00	1,43	1,43	24,77	0,70	0,39	0,00	1,09
D_ita1	63,20	2,25	3,61	7,33	13,18	76,38	0,00	3,93	0,00	3,93
C_ita1	30,03	0,87	1,30	2,58	4,76	34,79	0,66	0,57	0,00	1,24
A_ita1	39,39	0,21	3,98	5,40	9,59	48,98	0,00	5,10	0,00	5,10

Duration Ita2 in seconds

	Art. sp.	**ps**	**pp**	**pr**	**Tot p.**	**T. story**	**corr**	**dis**	**rep**	**tot**
M_ita 2	36,24	0,89	5,60	4,34	10,83	47,08	0,97	5,96	5,00	11,93
L_ita2	45,16	0,66	1,21	4,86	6,72	51,89	0,00	4,02	0,19	4,21
I_ita2	42,86	1,07	1,86	5,01	7,94	50,79	0,00	2,36	0,00	2,36
H_ita2	31,92	0,76	1,18	3,78	5,72	37,65	0,81	2,53	0,00	3,33
G_ita2	57,23	1,46	0,47	6,75	8,68	65,91	0,33	4,36	1,98	6,68
F_ita2	43,77	0,55	0,70	3,91	5,16	48,93	0,58	2,94	0,91	4,43
E_ita2	26,57	0,00	0,32	1,85	2,18	28,75	0,69	0,32	0,00	1,01
D_ita2	31,96	0,70	1,62	3,63	5,94	37,90	0,00	5,64	0,09	5,74
C_ita2	33,36	1,24	1,09	3,10	5,43	38,79	0,00	2,71	0,00	2,71
A_ita2	38,44	0,77	0,36	4,22	5,35	43,79	0,00	0,36	0,00	0,36

Duration Eng1 in seconds

	Art. sp.	**ps**	**pp**	**pr**	**Tot p.**	**T. story**	**corr**	**dis**	**rep**	**tot**
M_eng1	49,51	12,6	6,14	11,34	30,06	79,58	2,39	13,06	4,00	19,4
L_eng1	51,04	4,49	4,78	11,85	21,11	72,16	0,00	15,59	0,88	16,4
I_eng1	48,65	2,94	4,71	6,48	14,13	62,78	1,35	5,09	0,00	6,43
H_eng1	44,76	8,32	3,74	10,04	22,10	66,86	0,66	13,04	1,36	15,1
G_eng1	66,43	2,58	6,96	24,60	34,14	100,5	1,95	10,62	4,37	16,9
F_eng1	62,70	5,03	15,6	10,39	31,09	93,79	0,96	20,14	3,15	24,2
E_eng1	29,37	0,83	1,37	2,74	4,93	34,31	1,39	1,98	0,00	3,37
D_eng1	49,94	5,47	7,92	6,66	20,05	69,99	1,32	11,19	1,89	14,4
C_eng1	42,28	1,94	5,24	6,49	13,67	55,95	0,00	6,74	0,56	7,30
A_eng1	41,46	2,63	4,85	7,59	15,06	56,53	1,93	5,56	0,00	7,49

Duration English t2 (in seconds)

	Art. sp.	**ps**	**pp**	**pr**	**Tot p.**	**T. story**	**corr**	**dis**	**rep**	**tot**
M_eng2	40,51	7,35	10,5	5,04	63,41	63,41	0,51	12,20	1,51	14,22
L_eng2	55,17	3,83	10,6	8,04	22,53	77,70	0,35	14,59	0,22	15,16
I_eng2	56,66	5,38	3,24	9,53	18,16	74,82	0,00	7,77	0,00	7,77
H_eng2	36,61	0,97	0,58	5,95	7,51	44,12	0,81	3,49	0,21	4,51
G_eng2	57,68	1,18	2,79	6,65	10,61	68,29	0,94	7,18	1,46	9,58
F_eng2	62,03	2,64	12,3	8,41	23,34	85,37	1,23	15,09	3,29	19,61

E_eng2	23,30	0,44	0,00	1,96	2,40	25,70	1,22	0,45	0,06	1,73
D_eng2	53,63	6,59	5,71	5,64	17,94	71,57	1,17	11,75	1,22	14,14
C_eng2	39,84	0,27	2,82	3,36	6,45	46,29	1,19	3,63	0,59	5,41
A_eng2	45,95	2,13	1,75	5,67	9,55	55,50	2,68	2,76	0,20	5,63