ISSN 2499-4553

IJCoL

Italian Journal of Computational Linguistics

Rivista Italiana di Linguistica Computazionale

Volume 7, Number 1-2
june-december 2021
Special Issue

Computational Dialogue Modelling: The Role of Pragmatics and Common Ground in Interaction

aA Accademia university press

Registrazione presso il Tribunale di Trento n. 14/16 del 6 luglio 2016

Rivista Semestrale dell'Associazione Italiana di Linguistica Computazionale (AILC)

Associazione Italiana di Linguistica Computazionale

direttore responsabile
Michele Arnese

isbn 9791280136770

Accademia University Press
via Carlo Alberto 55
I-10123 Torino
info@aAccademia.it
www.aAccademia.it/IJCoL_7_1-2

IJCoL

Volume 7, Number 1-2
june-december 2021
Special Issue

Computational Dialogue Modelling: The Role of Pragmatics and Common Ground in Interaction

Invited editors: *Hendrik Buschmeier and Francesco Cutugno*
co-editors: *Maria Di Maro and Antonio Origlia*

CONTENTS

How are gestures used by politicians? A multimodal co-gesture analysis*

Daniela Trotta**
Università degli Studi di Salerno

Raffaele Guarasci†
ICAR-CNR

Gestures are an inseparable part of the language system (McNeill 2005; Kendon 2004), they are semantically co-expressive with speech serving different semantic functions to accompany oral modality (Lin 2017; McNeill 2016). To study these phenomena, we analyse the co-gesture behavior of several Italian politicians during face-to-face interviews. We add a new annotation layer to the PoliModal corpus (Trotta et al. 2020) focused on semantic function of hand movements (Lin 2017; Colletta et al. 2015; Kendon 2004). Then, we explore the patterns of co-occurrence of speech and gestures for the single politicians and from a party perspective. In particular, we address following research questions: i) Are there categories of verbs that systematically accompany hand movements in political interviews? ii) Since the corpus used presents an annotation of "speech constants" (Voghera 2001), is the Lexical Retrieval hypothesis confirmed or are gestures used in correlation with other and different constants of speech? The Lexical Retrieval hypothesis assumes that (a) gesturing occurs during hesitation pauses or in pauses before words indicating problems with lexical retrieval (Dittmann and Llewellyn 1969; Butterworth and Beattie 1978), and (b) that the inability to gesture can cause verbal disfluencies. Finally, we analyse semantic patterns of gesture-speech relationship.

1. Introduction

Messages can be encoded through verbal or non-verbal signals (Wagner, Malisz, and Kopp 2014). Although communication research has traditionally focused on speech – demonstrated by the fact that in recent decades a huge quantity of work, tools and approaches have been developed in the field of Spoken Corpus Linguistics (Voghera 2020; O'Keeffe and McCarthy 2010) – interest has shifted mainly towards multimodality in recent years. This is evidenced by the numerous occasions of discussions in the scientific community on this topic, focused on: technical modeling of manual gestures in human-machine interaction (i.e. the GESPIN conferences 2009 and 2011; Gesture Workshop Series), technical aspects of multimodal facial communication (i.e. The Audio-Visual Speech Processing Workshops - AVSP) and on research approaches to gesture analysis (i.e. LREC Workshops on Multimodal Corpora; the International Society for Gesture Studies). At the same time, there has been a strong increase of multimodal corpora

* Although the authors have cooperated in the research work and in writing the paper, they have individually devoted specific attention to the following sections: Daniela Trotta: 1, 2, 3 and 8; Raffaele Guarasci 4, 5, 6 and 7

** Dept. of Political Science and Communication – Via Giovanni Paolo II 132, 84084 Fisciano, Italy. E-mail: dtrotta@unisa.it

† Institute for high performance computing and networking – Via Pietro Castellino 111, 80131 Napoli, Italy. E-mail: raffaele.guarasci@icar.cnr.it

that stimulates sophisticated investigations into the relationship between the verbal and nonverbal components of spoken communication (Knight 2011).

This growing interest is strictly related to the fact that it is not possible to get a complete picture of human communication excluding some of the information provided during speech. As best pointed out by (Allwood 2008): "The basic reason for collecting multimodal corpora is that they provide material for more complete studies of interactive face-to-face sharing and construction of meaning and understanding which is what language and communication are all about". In fact, every spontaneous spoken communication is accompanied by gestures (i.e. facial expressions, hand movements, postures and body movements) (Voghera 2020). Indeed – as we will explain better in the section 2 – gestures accompanying speech take on multiple functions, ranging from complete the utterance, to substitute part of the utterance and to contradict the verbal sequence (Kendon 2004; McNeill 2008; Poggi 2007).

However, developing multimodal resources is extremely time-consuming (Lin 2017), because of the difficulty of transcribing and keeping track of all the non-verbal elements. Therefore, multimodal resources currently developed for all the languages are few and of different domains. The vast majority of these resources are monolingual relying on English language only.

Concerning Italian, the recent research on multimodal corpora is limited to the experience of the IMAGACT project (Moneglia et al. 2014) which aims at setting up a cross-linguistic Ontology of Action for grounding disambiguation tasks and it makes use of the universal language of images to identify action types, avoiding the underdeterminacy of semantic definitions. There are currently no resources for the Italian language that simultaneously account for verbal and non-verbal dimensions, this lack has affected the development of lines of research focused particularly on the relationships between the co-occurrence of speech and gesture.

Given that the television interview is inherently a multimodal and multisemiotic text, in which meaning is created through the intersection of visual elements, verbal language, gestures, and other semiotic cues (Vignozzi 2019), this study focuses on the co-gesture behavior of several Italian politicians during TV face-to-face interviews.

Starting from PoliModal corpus (Trotta et al. 2019, 2020), an Italian multimodal corpus of political domain, we add a new annotation layer focused on semantic function (i.e. reinforcing, integrating, supplementary, complementary, contradictory) of hand movements (Lin 2017; Colletta et al. 2015; Kendon 2004) in order to explore the patterns of co-occurrence of speech and gestures for the single politicians and from a party perspective.

1.1 Research Objectives

This work investigates political non-verbal communication. To date, in the literature Multimodal corpora have been used to analyse how gestures are used in different contexts such as narratives (Gregersen, Olivares-Cuhat, and Storm 2009; Holler and Wilkin 2011; Parrill, Bullen, and Hoburg 2010), academic domain (Knight 2011; Ovendale 2012), child language development (Colletta et al. 2015) and in relation to Italian action verbs (Moneglia et al. 2014). This study aims to explore the patterns of co-occurrence of speech and gestures in the specific case of Italian political interviews from a multimodal corpus linguistics perspective, addressing the following research questions:

1. Are there categories of verbs that systematically accompany hand movements in political interviews? This research question is inspired by

the study presented in (Vignozzi 2019) in which the analysis of the representation of some peculiar indicators of speech (i.e. idiomatic expressions and phrasal verbs) in a corpus of English television interviews of different domain, revealed that phrasal verbs are more recurrent in political interviews, while hand movements are more often associated with business and economic interviews.

2. Since the corpus used as a case study presents an annotation of so-called "speech constant" (Voghera 2001) (i.e. pauses, interjections, false starts, repetitions, truncations), is the *Lexical Retrieval hypothesis* confirmed or are gestures used in correlation with other and different constants of speech? Note that the *Lexical Retrieval hypothesis* assumes that (a) gesturing occurs during hesitation pauses or in pauses before words indicating problems with lexical retrieval (Dittmann and Llewellyn 1969; Butterworth and Beattie 1978), and (b) that the inability to gesture can cause verbal disfluencies (Dobrogaev 1929).
3. In the case of political interviews, what are the semantic patterns of gesture-speech relationship?

Our examination of the co-occurrence of speech and gesture will shed light into how the two communication models interact.

2. Background

2.1 Co-Gesture Analysis: a new perspective of linguistic analysis

A gesture is a visible action of any body part, when it is used as an utterance, or as part of an utterance (Kendon 2004). If such actions are produced while speaking, we can talk about co-speech gestures. Their occurrence, simultaneous or concomitant to speech, has led to different views regarding their role in communication (Wagner, Malisz, and Kopp 2014). First of all – as pointed pout by (Voghera 2020) – when we think about the relationship between verbal sequence and gestures, we should not imagine that the latter have a merely subordinate function to the word, but rather that there is a relationship of semiotic cooperation between them. The presence of gestures is useful to both the addressee and the speaker to maintain the rhythm of the speech rhythm of the speech and to mark the progression of information.

Some authors (McNeill 2005; Kendon 2004) have considered gestures as an integrative, inseparable part of the language system, since speaking itself is regarded as a variably multimodal phenomenon (Cienki and Müller 2008). Indeed gestures may provide important information or significance to the accompanying speech and add clarity to discourse (Colletta et al. 2015); they can be employed to facilitate lexical retrieval and retain a turn in conversations (Stam and McCafferty 2008) and assist in verbalizing semantic content (Hostetter, Alibali, and Kita 2007). From this point of view gestures facilitate speakers in coming up with the words they intend to say by sustaining the activation of a target word's semantic features long enough for the process of word production to take place (Morsella and Krauss 2004). Co-gesture speech can also refer to the spoken words or phrases that are co-produced with hand gestures in face-to-face spoken conversation (Lin 2017). According to (Krauss 1998) these co-occurring words or entire lexical phrases were identified to reflect the meaning of the co-occurring gesture; they are also known as "lexical affiliates" of the gesture, especially if they play a

particular role in the lexical retrieval. Indeed if gestures play a role in a lexical retrieval, they must stand in a particular temporal relationship to the speech they are supposed to facilitate.

Over the years, studies have shown that the production of gestures is influenced by the syntax of the language itself and by the socio-cultural context of the language. As explained in a 2015 study by (Colletta et al. 2015) – focused on co-speech gesture production in children's narratives – language syntax influences gesture production. For example – as known – some languages require an explicit subject (i.e. English, French, etc.), whereas others (i.e. Italian, Spanish, etc.) are null-subject languages. This characteristic requires distinct marking of referential continuity in the textual use of language, with less need to repeat anaphora in the latter case (Hickmann 2002). Another key factor influencing the communication is culture as a set of values and norms that helps shape the social behavior of individuals who belong to a cultural group as well as social interaction between them. Very well known is the study in (Kendon 2004), showing that Italians use a great number of gestures when communicating.

2.2 Gesturing with hands

The gestural movements of the hands and arms are probably the most studied co-speech gestures (Wagner, Malisz, and Kopp 2014). Based on the seminal works by (Kendon 1972) about the relationship between body motion and speech and by (Kendon 2011) about gesticulation and speech in the process of utterance, they are usually separated into several *gestural phases*: rest position, preparation phase, gesture stroke, holds and retraction or recovery phase (Bressem and Ladewig 2011). More generally, gestures can be described in terms of their form, semantic and pragmatic functions, their temporal relation with other modalities, and their relationship to discourse and dialogue context.

Since hand movements serve multiple functions in communication, it is often useful to define their semantic function. One of the best known classifications in this respect is that of (McNeill 1992) which attributes five semantic functions to hand movements:

- *emblematic gestures* bear a conventionalized meaning ("thumbs up");
- *iconic gestures* resemble a certain physical aspect of the conveyed information, e.g. they may convey the shape of a described object or the direction of a movement;
- *metaphoric gestures* are iconic gestures that resemble abstract content rather than concrete entities (McNeill 1992; Cienki and Müller 2008);
- *beat gestures* are simple and fast movements of the hands (also called batons (Ekman and Friesen 1972)).

This classification should not be understood as defining distinct categories. (McNeill 2005) argued that a simple functional classification of gestures is usually misleading. As (Wagner, Malisz, and Kopp 2014) pointed out due to the multifaceted nature of most gestures, he preferred a dimensional characterization of gestures, with dimensions including iconicity, metaphoricity, deixis, temporal highlighting (beats), and social interactivity. This acknowledges the fact that the majority of gestures can be characterized along several of these dimensions, e.g. when a pointing gesture also depicts the direction of a movement, or when a beat is superimposed onto the stroke onset of an emblematic gesture (Tuite 1993).

As we will explain more fully in Section 3.1 a further classification is proposed by (Lin 2017) adapting (Colletta et al. 2015; Kendon 2004), according to which gesture-speech relationship can assumes five possible semantic functions (i.e. reinforcing, integrating, supplementary, complementary, contradictory). Since this classification can be effectively used to capture the semantic contribution of gestures the utterances, we adopt it in our study and include such classes in our classification scheme.

2.3 Using multimodal corpora for analyzing gesture and speech in interaction

The concept of a multimodal corpus has been defined by (Allwood 2008) in terms of an annotated collection of "language and communication-related material drawing on more than one modality". Multimodal corpora (or multimedia corpora as they are often defined in the Italian literature) are used especially for pragmatic research purposes (i.e. in studies on proxemic correlates of spoken language or on the bodily manifestation of emotions), in which the starting sessions consist of videos that are transcribed and annotated (Cresti and Panunzi 2013). About what can be analyzed through the use of multimodal corpora, according to (Allwood 2001), although there are many research questions that could be answered through the use of these resources, they can be divided primarily into three major areas: *human-human face-to-face communication* (e.g. the nature of communicative gestures, multimodal communication in different national/ethnic cultures, communication and consciousness/awareness, etc.), *media of communication* (e.g. multimodality in writing, multimodality in songs and music, etc.), *applications* (e.g. better modes of multimodal human-computer communication, better modes of multimodal distance teaching/instruction, etc.).

In addition, multimodal corpora can be useful resources in the development of various computer-based applications, supporting or extending our ability to communicate, with regard to: modes of multimodal human-computer communication, better computer support for multimodal human-human communication, modes of multimodal communication for persons who are physically challenged (handicapped), modes of multimodal presentation of information from databases (for example for information extraction or for summarization), better multimodal modes of translation and interpretation, modes of multimodal distance language teaching (including gestures), better multimodal modes of buying and selling (over the internet, object presentation in shops, etc.), computerized multimodal corpora can, of course, also be useful outside of the areas of computer-based applications. In general, they can provide a basis for studying any type of communicative behavior in order to fine-tune and improve that behavior.

However these resources – probably due to the difficulty of construction – in Italy are difficult to find and consult, in fact between the 286 multimodal resources certified for all the languages by the LRE map[1] only one is in Italian, IMAGACT, a corpus-based ontology of action concepts, derived from English and Italian spontaneous speech (Moneglia et al. 2014; Bartolini et al. 2014). So this language is not well represented.

As specified in the section 1 – given that the television interview is inherently a multimodal and multisemiotic text, in which meaning is created through the intersection of visual elements, verbal language, gestures, and other semiotic cues (Vignozzi 2019) –

1 LRE map (Language Resources and Evaluation) is a freely accessible large database on resources dedicated to Natural language processing. The original feature of LRE Map is that the records are collected during the submission of different major Natural language processing conferences. The records are then cleaned and gathered into a global database called "LRE Map" (Calzolari et al. 2010). The map is freely available from the site https://lremap.elra.info/

this study focuses on the co-gesture behavior of several Italian politicians during TV face-to-face interviews, this therefore requires not only the presence of a multimodal resource but also of political domain.

In particular, non-verbal aspects acquire considerable importance especially in debates and interviews in the political domain, which is the area that is most suitable for this type of analysis (Seiter and Harry Jr. 2020). One of the particularly successful lines of research in recent years in the political domain is the analysis of gestures used by the speaker with the function of discrediting the opponent. These aspects have been the subject of various studies even in Italian language (D'Errico, Poggi, and Vincze 2013, 2012).

Concerning Italian language, some corpora have been made available recently, the largest one includes around 3,000 public documents by Alcide De Gasperi (Tonelli, Sprugnoli, and Moretti 2019) that has been mainly used to study the evolution of political language over time (Menini et al. 2020). All the corpora cited above are monomodal and none of them takes into account gestural traits. Indeed, corpora that include only one modality have a long tradition in the history of linguistics. According to (Lin 2017) "the construction and use of multimodal corpora is still in its relative infancy. Despite this, work using multimodal corpora has already proven invaluable for answering a variety of linguistic research questions that are otherwise difficult to consider".

Furthermore, none of the multimodal resources currently available in Italian present a systematic annotation of gestures, since is not possible to construct a state of the art on the presence and behavior of co-gestural patterns for this language.

2.4 PoliModal corpus: description and new layer of annotation

PoliModal corpus (Trotta et al. 2019, 2020) contains transcripts of 56 TV face-to-face interviews of 14 hours, taken from the Italian political talk show "In mezz'ora in più" broadcast between 2017 and 2018, for a total of 100,870 tokens. The corpus has a double level of annotation using XML as markup language. The first one was done manually following the TEI standard for Speech Transcripts in terms of utterances and takes into account the "speech constant" (Voghera 2001). In particular:

(a) **Metadata**: these include useful information for a quick identification of transcriptions, for example the tools used for the transcription, a link to the interview, the owner account, the title of the talk show, the date of airing, the guests, etc.

(b) **Pause**: this tag is used to mark a pause either between or within utterances;

(c) **Semi-Lexical**: this tag is used to label interjections (i.e. 'eh', 'ehm' etc.), or more generally words that convey the meaning of an entire sentence, constituting a complete linguistic act demonstrated by their paraphrasability;

(d) **FalseStart**: this tag shows the speaker's abandonment of an already produced word or sequence of words, with or without repetition of previously used linguistic material;

(e) **Repetition**: with this tag are marked cases of repetition of utterances in order to give coherence and cohesion to the speech or self-repetition as a control mechanism of the speech programming;

(f) **Truncation**: truncation indicates the deletion of a phoneme or a syllable in the final part of a word.

This annotation task addressed so far falls – from a qualitative point of view – in the first of the general types identified by (Mathet, Widlöcher, and Métivier 2015), in which the subjective interpretation is limited. Indeed, it deals with the "identification of units" (Krippendorff 2018), in which the annotator, given a written or spoken text, must iden-

tify the position and boundary of linguistic elements (e.g. identification of prosodic or gestural units, topic segmentation). In order to evaluate the reliability of our annotation scheme, we computed inter-annotator agreement by performing a double annotation of verbal and non-verbal traits of the first ten minutes of Renzi's, Di Maio's and Salvini's interview. Both annotators were expert linguists. Macro-averaged F1 computed on exact matches amounts to 0.82, which corresponds to a good agreement, given that by exact match we consider the correct choice of the trait, the position of the tag and the exact extension of the marked string, if any. This result confirms the reliability of the task and the corresponding annotation guidelines.

The second annotation level was performed automatically using ANVIL (Kipp 2001) – a tool for the annotation of audiovisual material containing multimodal dialogue – following the MUMIN (Allwood et al. 2007) annotation scheme that takes into account ten types of gestures divided into three categories:

(a) **Facial displays**: they refer to timed changes in eyebrow position, expressions of the mouth, movement of the head and of the eyes (Cassell and others 2000). The coding scheme includes features describing gestures and movements of the various parts of the face, with values that are either semantic categories such as Smile or Scowl or direction indications such as Up or Down.

(b) **Hand gesture**: we follow a simplification of the scheme from the McNeill Lab[2] The features, 7 in total, concern Handedness and Trajectory, so that we distinguish between single-handed and double-handed gestures, and among a number of different simple trajectories analogous to what is done for gaze movement. The value Complex is intended to capture movements where several trajectories are combined.

(c) **Body posture**: this tag comprises trajectory indications for the movement of the trunk. The categories are mutually exclusive to facilitate the annotation work.

The annotation – made at the moment by a single expert annotator – follows the criterion highlighted by (Allwood et al. 2007), claiming that annotators are expected to select gestures to be annotated only if they have a communicative function. In other words, gestures are annotated if they are either intended as communicative by the communicator (displayed or signalled) (Allwood 2001), or judged to have a noticeable effect on the recipient.

However, this last level of annotation does not take into account the semantic functions covered by these gestures and therefore would not allow to develop an in-depth analysis of the semantic contribution they could make to the discourse. So – as we will explain in depth in the Section 3.1 – we manually add a new level of annotation that takes into account the semantic functions covered by one of the gestures already tagged in the corpus: hand movements.

3. Methodology

3.1 Coding co-speech gesture in PoliModal corpus

In the paper by (Allwood 2001), the authors highlight that synchronization of information in different modalities is a crucial issue in assembling a multimodal corpus. Therefore the authors suggest to adopt the general principle of spatio-temporal contiguity. This means that a text occurs at the same point in time as the event it describes or represents. When temporal contiguity concerns the relation between transcribed speech

2 Duncan, S. (2004). Coding manual. Technical Report availale from http://www.mcneilllab.uchicago.edu.

(or gesture) and recorded speech (or gesture), it is often referred to as "synchronized alignment" of recording and transcription. What synchronization means is that for every part of the transcription (given a particular granularity), it is possible to hear and view the part of the interaction it is based on and that for every part of the interaction, it is possible to see the transcription of that part. The form of connection between the transcriptions and the material in the recordings can vary from just being a pairing of a transcription and video or audio recording, where both recording and transcription exist but they have not yet been synchronized, to being a complete temporal synchronization of recordings and transcription. In our case, audio and video signals as well as the annotations have been temporally synchronized by hand. Although the most convenient solution for synchronization is to carry it out using a computer program already while making the recording (see for example the AMI project and CHIL project), we did it manually since the recording and transcription of the corpus were done before knowing what layers would be exactly annotated.

Starting from PoliModal corpus described in 2.3, we manually add a new level of annotation that takes into account the semantic functions covered by one of the gestures already tagged in the corpus: hand movements. This is because the gestural movements of the hands and arms, i.e. spontaneous communicative movements that accompany speech (McNeill 2005), are probably the most studied co-speech gestures (Wagner, Malisz, and Kopp 2014). Based on the seminal works by (Kendon 1972) about the relationship between body motion and speech and by (Kendon 2011) about gesticulation and speech in the process of utterance, they are usually separated into several *gestural phases*: rest position, preparation phase, gesture stroke, holds and retraction or recovery phase (Bressem and Ladewig 2011). Additionally, the point of maximal gestural excursion is often regarded as a gestural *apex*.

In PoliModal the **hand movement trajectory** tag indicates only the start and end of the movement in terms of time and the trajectory of the gesture, in particular *up, down, sideways, complex*. In order to keep track also of the semantic function covered by the tag, we added an additional information layer to those already present – following the classification proposed by (Lin 2017) adapting (Colletta et al. 2015) and (Kendon 1972) – which attributes five functions to hand movements:

- *Reinforcing*: the information brought by the gesture is equal to the linguistic information it is in relation with. For example, one of the interviewees emphasizes the sacrifices to which Italians have been subjected in the last fifteen years, including "il 3% del rapporto deficit/PIL" (*en.* "the 3% deficit/PIL ratio"). In saying this he makes the sign of the number three with the fingers of his right hand.
- *Integrating*: the information provided by the gesture does not add supplementary information to the verbal message, but makes the abstract concepts more precise. A frequent example in our annotation is when a politician, in order to contrast two items such as left and right parties, points one of his hands toward the right and the other toward the left.
- *Supplementary*: the information brought by gestures adds new information not coded in the linguistic content. For example, in one of the interviews, the interviewee comments on the amount of members of Parliament elected from another party saying "…non so quanti parlamentari porterà in Parlamento" (*en.* "…I don't know how many MPs they will bring to

Parliamen") and in the meantime he opens his arms as if to imply a large number.

- *Complementary:* the information provided by the gesture brings a necessary complement to the incomplete linguistic information provided by the verbal message. The gesture usually disambiguates the message, for example, in our annotation it is common to find cases where deictic adverbs such as *qui* (*en.* here) are accompanied by the corresponding pointing gesture.
- *Contradictory*: the information provided by the gesture contradicts the linguistic information provided by the verbal message. This kind of gesture was not found in our annotation.
- *Other*: within this category we include all the gestures that annotators were not able to classify with the above mentioned semantic labels.

Our annotation follows the selection criterion highlighted by (Allwood et al. 2007), claiming that annotators are expected to select gestures to be annotated only if they have a communicative function. However, as (Yoshioka 2008) points out gestures can be functionally ambiguous and thus have multiple semantic functions simultaneously. According to (Tsui 1994), the source of this multiple functions often lies in the sequential environment of the conversation in which the utterance occurs. To simplify the task, annotators are therefore asked to assign a single semantic function to the gestures under investigation, choosing the function that s/he considers prevalent in the context of use.

In order to evaluate the reliability of our annotation scheme, we compute inter-annotator agreement by performing a double annotation of the semantic functions listed above on three of the interviews considered (Matteo Renzi, Luigi Di Maio, Matteo Salvini) for a total of about 2 hours of interviews. Both annotators (one male and one female) are expert linguists. Macro-averaged F1 computed on exact matches amounts to 0.83, which corresponds to an almost perfect agreement. This result confirms that the task is well-defined and that the corresponding annotation guidelines are clear.

Figure 1 shows an example annotation with the new information layer specified the semantic function (tag 'function'). For each observed gesture, the PoliModal corpus already contained: i) the start and end point in the video in terms of milliseconds; ii) the type of gesture observed; iii) the movement trajectory. We add to this the semantic function covered by the gesture in the context.

```
<u gender="m" length="928" role="Minister of the foreign
business and of the international cooperation" time=
"452.28" who="Angelino Alfano">C'è qualcosa di più grave
e di più profondo di cui mi sono occupato da Ministro
dell'Interno. Perché io ho gestito l'immigrazione
<movement start="470.2" end="471.2" attribute="Hand
movement trajectory" attribute_text="sideways" function=
"integrating">e ho gestito l'ordine pubblico.
</movement></u>
```

Figure 1
Annotation sample in xml

4. Systematical co-occurrence of hand-movement and specific categories of verbs

The study presented in (Vignozzi 2019) aimed to analysing the representation of some peculiar indicators of spokenness (i.e. idiomatic expressions and phrasal verbs) across TV interviews featuring different interviewees (politicians, business people and personalities from showbiz). The analysis pointed out that phrasal verbs are more recurrent in political interviews than in business and economic discussions, and that the specialized domain with which hand or arm movements are more often associated is again business and economics (60.86%). In political interviews, instead, gestures appear in 58.02% of cases, while in showbiz interviews the lowest frequency is observed, since gestures occur only in 40.42% of the cases. Besides, the study shows that beats gestures are the most frequent kind of gestures co-occurring with phrasal verbs, especially in political interviews, where they account for more than half of the total of gestures. The study was conducted on "The ESP Video Clip Corpus" in English.

In order to understand whether hand gestures (identified by the tag *hand movement trajectory*) is related in a systematic way to particular types of verbs (e.g. predicative, phrasal verbs etc.), we created a subcorpus containing only the sentences of the interviews co-occurring with the tag under investigation were extracted (for a total of 495 sentences).

The qualitative approach has been preferred in this phase for two main reasons: first of all, because the amount of data to be analyzed is controllable; moreover because existing resources for Italian such as LexIt (Lenci, Lapesa, and Bonansinga 2012), MultiWordNet (Pianta, Bentivogli, and Girardi 2002) and T-PAS (Jezek et al. 2014) do not make explicit the function that the verbs assume in the context (e.g. no tool will tell us if the verb is servile, appellative, estimative, elective, etc.).

Through a qualitative analysis, we then manually classified verbs according to their function in the text (Jezek 2003). The verbal classes identified are as follows (with the total number of occurrences in parenthesis):

- Predicative verbs: they have full lexical meaning and can independently give rise to a verbal predicate of full meaning. The class of predicative verbs encompasses the vast majority of verbs in a language, and is descriptively opposed to the class of copulative verbs that need to rely on a predicative complement to fulfill the predicate function: *sembrare [to appear] (13), parere [to seem] (5), risultare [to result] (4),* ***stare (131)****, restare (7), rimanere (2) [to stay, to remain], diventare (5) [to become]*

- Predicative verbs which can carry a predicative complement of the subject, but only if conjugated in the passive form: *chiamare [to call] (2)*, *eleggere [to elect] (2)*, *giudicare [to judge] (1)* and ***fare*** *[to do]* ***(12)***.

- Phrasal verbs are verbs that, when combined with another non-finite mode verb with the interposition of a preposition (to, of, for, from), specify a particular time-expectant mode. They are divided into 5 groups:

  – the imminence of an action: *stare per (3)* + infinitive

  – the beginning of an action: *cominciare a [begin to] (7)* + infinitive

  – the development of an action: ***stare*** *[stay] (****38****)* and *venire (15)* [come] + *gerund*

- the duration and continuity of an action: *continuare a [continue to] (6)* + infinitive
- the conclusion of an action: *finire di (1)* and *smettere di (1)* [stop to] + infinitive

- Causative verbs: indicate that the action is caused by the subject, but that he does not perform it directly. The only causative of the Italian language that occurs in the corpus is the verb ***fare*** *[to do]* ***(20)*** + infinitive

- Performative verbs: they exist only in the first person singular of the present indicative and are so defined because pronouncing them is equivalent to performing the action they describe, i.e. to perform the action they describe one must pronounce them. The only verb belonging to this class present in the corpus is *negare [to deny](1)*.[3]
  The other verb taking a performative function in the first person of the present indicative is ***dire*** *[to say]* ***(26)***.

Most function verbs are predicative, that is, they have an independent meaning, forming what in syntax is called a verbal predicate. Among them we notice a more frequent use of the verb **stare** [to stay] with 131 occurrences.

(4) Salvini:*"Ci possono essere altre sfumature, a qualcuno* ***sta*** *simpatico Macron, a qualcuno* ***sta*** *simpatica la Le Pen, è il rapporto con l'Europa che per me è determinante al di là delle simpatie."* (*en.* "There may be other nuances, someone **likes** Macron, someone **likes** Le Pen, it is the relationship with Europe that for me is decisive beyond sympathies.")

Among verbs with a predicative function of the subject (only when used in the passive form), the most commonly used are effective verbs, i.e. copulative verbs indicating a state, semblance, or transformation. In this case the most frequent is **fare** [to do] with 12 occurences.

(5) Padoan:*"Secondo te questa campagna elettorale sta dividendo il paese in due. Tra chi vuole continuare e rafforzare quello che* ***è stato fatto*** *e ha portato i risultati che lei ricordava, piuttosto che chi vuole eliminare."* (*en.* "According to you, this election campaign is dividing the country in two. Between those who want to continue and strengthen what **has been done** and has brought the results that you recalled, rather than those who want to eliminate.")

On the other hand, with respect to phrasal verbs, the results obtained do not confirm what emerged in (Vignozzi 2019), in which a predominance of servile verbs was noted in political domain interviews, because in our case there is a slight but not clear prevalence of verbs that indicate the performance of an action, in particular of the verb **stare** [to stay] + gerund with **38** occurrences.

3 See for example the utterance by Di Battista: *Io ho avuto credo 84 giorni di espulsione dalla Camera dei Deputati e non ho mai picchiato nessuno, mai.* ***Anche se non le nego…*** (*en.* I've had I think 84 days of expulsion from the House of Representatives and I've never hit anybody, ever. Although I don't deny them…).

(6) Veltroni:*"E quello che **sta succedendo** in Italia, l'affermazione non delle forze tradizionali…"* (*en.* "And what **is happening** in Italy, the assertion not of traditional forces…")

Among causative verbs, the most present is the verb **fare** [to do] (20 occurrences), while the among performative ones it is **dire** [to say] (26).

(7) Tremonti:*"E quando comincio a vedere che perfino Prodi parla di un colpo di quel tipo, avremmo dovuto andare a votare e non **ci hanno fatto** andare a votare. Perché dovevano mandarci il Governo tecnico che tecnicamente ci ha buttato giù."* (*en.* "And when I start to see that even Prodi is talking about that kind of hit, we should have gone to vote and they **didn't make us to go to vote**. Because they had to send us the technical government that technically brought us down.")

(8) Di Maio:*"Guardi io le **dico** noi parleremo con tutti coloro che aderiranno però…"* (*en.* "Look **I'll tell you** we're going to talk to everyone who joins though….")

Causative verbs are verbs that express an action not performed by the subject, but made to be performed by others. In this case, we notice a prevalence of the verb **fare** [to do], mainly used with a negative valence and referred to the political opposition; in fact, this verb mainly describes actions that the subjects were forced to carry out because of the determined political circumstance of the moment.

The concept of performative act was introduced by the theory of linguistic acts elaborated in (Austin 1975). Verbs that take on this function are so defined because pronouncing them is equivalent to performing the action they describe. In other words, in order to perform the action they describe, one must pronounce them. Probably the performative verb **dire** is more present in these interviews because - being in the middle of an electoral campaign - politicians want to give an impression of being concrete and aim at emphasising their statements.

5. Is the Lexical Retrieval hypothesis confirmed?

Many studies have suggested that gestures, especially representational gestures (Krauss and Hadar 1999) play a direct role in speech production by priming the lexical retrieval of words. This view has been termed the *Lexical Retrieval hypothesis*.

The hypothesis is based on research arguing that (1) gesturing occurs during hesitation pauses or in pauses before words indicating problems with lexical retrieval (Dittmann and Llewellyn 1969; Butterworth and Beattie 1978), and (2) that the inability to gesture can cause verbal disfluencies (Dobrogaev 1929). In addition – as (Krauss 1998) pointed out – speakers were more disfluent overall in constrained-speech conditions than in natural conditions. Since the corpus used as the object of study presents a level of annotation that takes into account some hesitation pauses and verbal disfluencies, we decided to verify this hypothesis in the political domain, where speakers usually have to control well their communication and be persuasive.

We compute weighted mutual information between hand movements and each of the speech disfluencies reported in Table 1. This measure is calculated to show existing mutual dependencies between co-occurring tags. We consider only the interviews in the PoliModal corpus that have a minimal length of 50 turns, so to have a good amount of

annotations to consider. We report in Table 1 the tag incidence per 100 turns for each interview considered.

Table 1
Tag incidence per 100 turns for each interview

Interviewee	*Hand mov.*	*Pause*	*Semi-Lexical*	*FalseStart*	*Repetit.*	*Truncat.*
Matteo Renzi	35.82	0	8.50	10.16	22.45	36.89
Luigi Di Maio	22.97	0	14.86	0	18.91	18.91
Matteo Salvini1	54.38	5.20	24.56	0	24.56	19.29
Matteo Salvini2	52.87	14.94	21.83	3.44	21.83	3.44
Walter Veltroni	41.81	0	14.54	21.81	29.09	18.18
Simone Di Stefano	10.98	0	4.39	5.49	21.97	16.48
Pierluigi Bersani	32.29	1.04	26.04	0	31.25	20.83
Angelino Alfano	57.00	9.00	33.00	3.00	17.00	3.00
Giulio Tremonti	10.71	16.07	10.71	0	14.28	0
Matteo Orfini	29.85	1.49	11.94	0	14.92	0
Pier Carlo Padoan	49.27	11.94	30.43	1.44	7.24	13.5
Carlo Calenda	74.63	32.60	24.63	9.42	7.24	0.72
Alessandro Di Battista	39.02	9.26	32.19	6.82	11.70	10.58
Average	39.35	7.81	18.89	4.74	17.74	12.45

Among the politicians included in this dataset, the one that most accompanies his speech with the movements of the hands is Matteo Salvini (Lega) considering both interviews, followed by Carlo Calenda (PD) and Angelino Alfano (Il Popolo della Libertà). Their belonging to different political parties suggests that the use of hand movements is more an individual trait than a feature characterising specific political positions.

Weighted mutual information (WMI) is computed between hand movements and tags reported in Table 1. The values obtained are shown in the heatmap reported in Figure 2, with lighter colors corresponding to higher WMI values.

Overall, hand movements tend to have a higher association with semi-lexical traits and pauses, which would confirm the assumptions of *Lexical Retrieval hypothesis* according to which gesturing occurs during hesitation pauses or in pauses before words indicating problems with lexical retrieval (Dittmann and Llewellyn 1969; Butterworth and Beattie 1978). Indeed, semi-lexical expressions, such as 'ah', 'eh', 'ehm', have been associated with the fact that linguistic planning is very cognitively demanding, and it is difficult to plan an entire utterance at once. This effect is however not present for some politicians, such as Di Battista and Alfano, while it is evident for some others such as Bersani and Salvini. Therefore, our findings are not generally applicable to all interviewees in our corpus. Fig. 2 shows also evident differences in gesturing behavior among the considered politicians. For instance, although Carlo Calenda and Angelino Alfano present a high incidence of hand movements, they do not seem to be associated with specific tags. Matteo Renzi, instead, shows a gesturing behavior that is unique compared to all the other interviewees, with hand gestures that are almost always used in association with other speech phenomena.

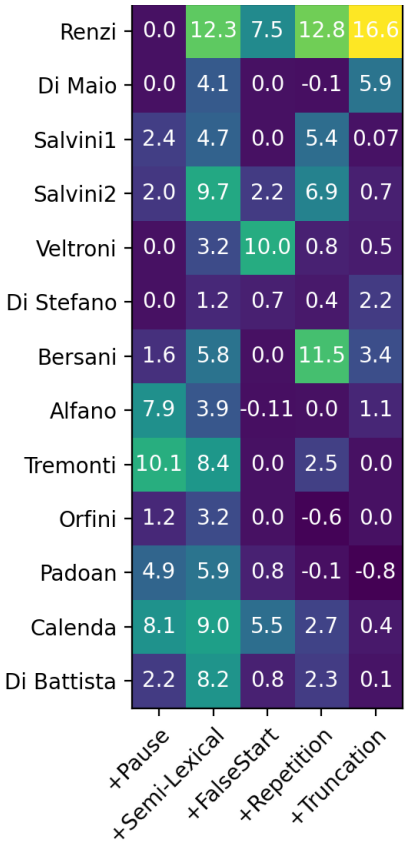

Figure 2
WMI values between hand movements and tags reported on the x-axis for each interviewee on the y-axis

In the interviews, we observe also the presence of negative values for WMI obtained in relation to false-starts (-0.11), repetitions (-0.1 and -0.6) and truncations (-0.8), suggesting that hand movements are less likely to be accompanied by such linguistic phenomena.

Notice that the results are consistent with the Tradeoff Hypothesis (De Ruiter, Bangerter, and Dings 2012). Qualitative analysis shows that when respondents are more disfluent in speech, they gesticulate more. This behavior reflects what is stated in the hypothesis "when gesturing gets harder, speakers will rely relatively more on speech, and when speaking gets harder, speakers will rely relatively more on gestures".

6. Is the gesture-speech relationship influenced by linguistic variables?

The third analysis carried out was aimed to understand if the hand movements produced by the interviewees have significant correlations with language complexity. As in the previous analysis, a threshold was established, therefore only interviews with a minimal length of 50 turns was taken into account.

For complexity we consider the type-token ratio and the average lexical density, i.e. the number of content words divided by the total number of tokens. We do not take into account the Gulpease index (Lucisano and Piemontese 1988), despite it is considered the standard metric of readability in Italian. But its reliability is undermined by several limitations (Tonelli, Tran Manh, and Pianta 2012) like sentence length and polysyllabic words; in addition it has been specifically designed, not very suitable for transcripts.

We perform an analysis of the correlation between language complexity and hand movements, normalised by the number of tokens uttered by each politician multiplied by one thousand. Since the variables under examination are both cardinal or quantitative, the Person's correlation index had been used for each interviewee and for each political party they belong to.

Table 2
Normalized values of hand movements, TTR, and lexical density for each interviewee

Interviewee	*Hand movement*	*TTR*	*Lexical Density*
Matteo Renzi	35.82	0.71	0.563
Luigi Di Maio	22.97	0.8	0.562
Matteo Salvini1	54.38	0.73	0.567
Matteo Salvini2	52.87	0.82	0.569
Walter Veltroni	41.81	0.7	0.569
Simone Di Stefano	10.98	0.75	0.583
Pierluigi Bersani	32.29	0.73	0.547
Angelino Alfano	57	0.61	0.564
Giulio Tremonti	10.71	0.75	0.585
Matteo Orfini	29.85	0.72	0.566
Pier Carlo Padoan	49.27	0.75	0.570
Carlo Calenda	74.63	0.73	0.580
Alessandro Di Battista	39.02	0.8	0.568

Individual interviewee computations reveal that both the TTR and the conceptual density show a moderate negative correlation with hand movements, respectively $r = -0.3$ and $r = -0.12$. Since in all cases considered the correlation is negative it could deduce that the Information Retrieval hypothesis is confirmed. The value of the TTR could mean that the more you gesticulate the more the lexical richness decreases and therefore there are more hesitations.

Instead in the case of conceptual density, the negative value $r = -0.12$ could mean that the more you gesticulate the more the speech tends to be simple and understandable (this could find even more justification in the format of the interview that being televised and being broadcast at a time when the audience is quite varied, it could tend to be easier to be understood by all).

Also political parties computations reveal that both the TTR and the conceptual density show a moderate negative correlation with hand movements, respectively $r = -0.7$ and $r = -0.71$ even if slightly higher than the correlation per single respondent with a deviation of 0.5 for TTR and 0.6 for conceptual density. The correlation values obtained by political party of belonging show a slight negative correlation, which could mean that the party of belonging does not significantly influence the use of the semantic communication plan and consequently the use of language.

Table 3
Values of hand movements, TTR, and lexical density for each political party

Political Party	*Avg. Hand movement*	*Avg. TTR*	*Avg. Lexical Density*
PD	43.94	0.73	0.566
M5S	30.99	0.80	0.565
Lega	39.32	0.76	0.574
CasaPound	10.98	0.75	0.583
Il Popolo della Libertà	57	0.61	0.564

Therefore, the first correlation values obtained allow us to state that the gesture-speech relationship is not influenced either by the political party or by the linguistic variables considered.

7. What are the semantic patterns of gesture-speech relationship?

A summary of the hand movement annotations in the corpus is reported in Table 4 and 3. In the first one, the number of annotated tags is reported for each politician, while in the second table the values are aggregated by political party. The parties include PD (left-center), Movimento 5 Stelle (center-populist), Lega (right-populist), Casa Pound (right), Popolo delle Libertà (center-right). The "Contradictory" category is not reported in the tables because it was never found in the interviews. This is probably due to the fact that in political interviews broadcast on TV, politicians try to be as clear as possible, avoiding statements and behaviour that may be misunderstood. Therefore, gestures and speech that are in contradiction are generally avoided. Probably for the same reason, supplementary movements, adding new information that is lacking in the linguistic content, are not frequent. 'Integrating' movements, instead, can be seen as an attempt to emphasise the speech content without adding supplementary information. This type of movement is the most frequent one, followed by "Complementary".

A qualitative analysis of the single interviews shows interesting differences in attitude and communication style, which pertain to single politicians rather than to party positions. Matteo Renzi, for example, uses gestures very frequently to accompany his speech. We report an example of 'Integration' below:

Matteo Renzi:"*Quello che sta accadendo invece in queste settimane, in questi mesi,* ***conferma che c'è una grande distanza tra la politica dei palazzi e la politica della quotidianità*** *[integrating]."*

(Eng. *"Instead what is happening in these weeks, in these months, confirms that there is a great distance between the politics of the Palaces and the politics of everyday life."*)

Renzi underlines that the distance between politics made by elites, detached from the real problems of the country ("politics of the Palaces"), and politics of everyday life, that is, attentive to reality and to citizens, is increasingly evident. Gesture is used to stress this difference: the speaker's open right hand points away from his torso in correspondence with the metaphorical expression politics of the Palaces, almost as if to indicate that it is something in which he does not recognize himself. His right hand then immediately rejoins his left hand and points downwards at the moment in which the expression politics of everyday life is pronounced, as if to indicate a politics that is instead attentive to relevant and concrete things.

Concerning the *Reinforcing* type of gesture-speech relationship, it is mainly used to reiterate a concept already expressed linguistically, and it is not very used, probably because it may seem redundant. Angelino Alfano turns out to be the interviewee who makes most use of this type of gesture. In this example, Alfano, talking about the consensus obtained by one of his political opponent Matteo Salvini, claims that this consensus was obtained at his expense. So, in saying "contro di me" (against me), the open hands are close to his bust.

Angelino Alfano:*"Quindi la sfida di Salvini, avendo aggregato consenso –* ***contro di me peraltro*** *[reinforcing] – sull'immigrazione, è incanalarlo su un regime di legislazione democratica."*

(Eng. *"So Salvini's challenge, by aggregating consensus – against me by the way – on immigration, is to channel it on a regime of democratic legislation."*)

As mentioned above, *Supplementary* gestures are used with a very low frequency. One of the few examples in the corpus is present in Simone di Stefano's interview, where he is asked to clarify the alleged relations of the party with a convicted member of the Mafia. The interviewee tries to provide an explanation, but the interviewer continues to put him under pressure. At this point the interviewee lowers his gaze and moves his open right hand away from his torso while saying "*but I don't want to avoid [your question]*", as if to implicitly ask the journalist to stop her suppositions and let him explain his position.

Complementary gestures bring a necessary complement to the incomplete linguistic information provided by the verbal message. They are frequently used by the respondents in the corpus under analysis, in most cases to disambiguate the message or simply some linguistic elements. This indicates the speaker's intention to be as clear as possible. For example, at the beginning of the interview with Carlo Calenda, he is shown a photo that portrays him wearing a worker's helmet. The interviewee refers to the photo by pointing with his left hand away from his torso to the screen where the photo is displayed, making it easier for viewers to understand what he was referring to:

Carlo Calenda:*"Benché gli operai non si sentiranno, come posso dire, contenti dopo aver visto* ***la mia foto con quel caschetto*** *[complementary] in cui sembravo un totale ebete."*

(Eng. *"Although the workers won't feel, how can I say, happy after seeing the picture of me in that helmet where I looked like a total stupid."*)

As noted above, a residual category has been added to the tags. The *Other* category includes all the gestures that annotators were not able to classify with the above mentioned semantic labels. This problem was found most frequently in the interviews with Pier Carlo Padoan and Carlo Calenda. These gestures are different from the others because they show a *batonic* value, that is, they are used to mark the rhythm of the enunciation, for example by tapping a finger on the table.

8. Conclusion

This paper investigate co-gesture speech of several Italian politicians during face-to-face interviews. To this purpose, we enrich PoliModal – a multimodal Italian political domain corpus – with a new layer of annotation, describing the semantic function of the different hand movements.

Concerning the type of verbs used – which in Italian cab be broadly distinguished in predicative, copulative, auxiliary, phrasal, performative and causative (Jezek 2003) – it was noticed that: among the verbs with a predicative function of the subject the most commonly used are effective verbs, i.e. copulative verbs indicating a state, semblance, or transformation; with respect to phrasal verbs, the results obtained do not confirm what

Table 4
Frequency of the type of gestures annotated for each interviewee

Interviewee	*Integrat.*	*Reinforc.*	*Supplement.*	*Complement.*	*Other*
Matteo Renzi	32	9	2	23	1
Luigi Di Maio	6	0	1	9	1
Matteo Salvini1	16	6	3	5	1
Matteo Salvini2	17	10	0	14	5
Walter Veltroni	8	3	0	8	4
Simone Di Stefano	5	0	2	3	0
Pierluigi Bersani	13	4	0	12	2
Angelino Alfano	21	11	1	16	8
Giulio Tremonti	3	1	1	1	0
Matteo Orfini	7	0	0	10	3
Pier Carlo Padoan	16	0	0	3	15
Carlo Calenda	41	1	0	35	26
Alessandro Di Battista	29	1	0	20	0
Total	214	46	10	159	66

emerged in (Vignozzi 2019), in which a predominance of servile verbs was noted in political domain interviews, because in our case there is a slight but not clear prevalence of verbs that indicate the performance of an action, in particular of the verb **stare** + gerund with **38** occurrences. Among causative verbs, the verb **fare** (20 occurrences) is the one that occurs most frequently, while the among performative ones it is **dire** (26). Causative verbs has been detected a prevalence of the verb **fare**, mainly used with a negative valence and referred to the political opposition; in fact, this verb mainly describes actions that the subjects were forced to carry out because of the determined political circumstance of the moment. Other evidence is in favor of performative verb **dire** probably more present in these interviews because – being in the middle of an electoral campaign – politicians want to give an impression of being concrete and aim at emphasising their statements.

Furthermore, we test the *Lexical Retrieval Hypothesis* by computing the association between hand movements produced by each interviewee and speech disfluencies using *weighted mutual information*. Results show that hand movements tend to co-occur with full pauses (i.e. repetition) and empty pauses (i.e. pause) and more frequently with interjections (i.e. semi-lexical), suggesting that gesticulating may represent an attempt at lexical retrieval. In future developments we plan to extend the analysis taking into account more recent theories, e.g. the Tradeoff Hypothesis(De Ruiter, Bangerter, and Dings 2012), more general and empirically better supported.

Concerning gesture-speech relationship, the results obtained suggest that hand movements are mainly used with an integrative and complementary functions. So, the information provided by such gestures adds precision and emphasis to linguistic information.

Finally we perform an analysis of the correlation between language complexity and hand movements. Individual interviewee computations revealed negative correlation values for both TTR and conceptual density, further confirming Information Retrieval and letting us assume that probably the more you gesticulate the more your lexical

richness decreases, leading to more hesitation in speech. At the same time, the negative correlation values obtained for lexical density might suggest that the more the speaker makes use of gestures in his speech, the simpler and more comprehensible it tends to be. However, concerning the correlation by political party, again negative correlation values were obtained for both TTR and conceptual density, suggesting that party affiliation would not influence the use of gestures.

In the future we plan to make this new level of annotation freely accessible in order to make possible both comparative studies in other languages and other fields of knowledge such as political science. In addition, we will initiate a predictive study aimed at understanding which of the variables under investigation may be effective predictors of the occurrence of hand movements. A further future development could be to use a comparison between sentences with hand movements and those ones in which no movements are present – through the creation of two different subcorpora – in order to understand if the increase in complexity of language is accompanied by a parallel growth of gestures with the aim of increasing clarity of speech.

A further aspect that we propose to investigate concerns the function of gestures to discredit the opponent in political debates. This topic has been much discussed in the literature, both with regard to rhetorical and persuasive aspects, and with particular focus on multimodal communication (D'Errico, Poggi, and Vincze 2013, 2012; D'Errico and Poggi 2012; D'Errico 2019). Currently these aspects have not been considered because they are not present in the sample used as the object of analysis. However, given the nature of the interviews composing the corpus, may be a promising line of research.

References

Allwood, Jens. 2001. Dialog coding - Function and grammar: Göteborg coding schemas. *Gothenburg Papers in Theoretical Linguistics*.

Allwood, Jens. 2008. Multimodal corpora. In A. Lüdeling and M. Kytö, editors, *Corpus Linguistics. An International Handbook*. Mouton de Gruyter.

Allwood, Jens, Loredana Cerrato, Kristiina Jokinen, Costanza Navarretta, and Patrizia Paggio. 2007. The mumin coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation*, 41(3-4):273–287.

Austin, John Langshaw. 1975. *How to do things with words*, volume 88. Oxford university press.

Bartolini, Roberto, Valeria Quochi, Irene De Felice, Irene Russo, and Monica Monachini. 2014. From synsets to videos: Enriching italwordnet multimodally. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3110–3117, Reykjavik, Iceland, May. European Language Resources Association (ELRA).

Bressem, Jana and Silva H. Ladewig. 2011. Rethinking gesture phases: Articulatory features of gestural movement? *Semiotica*, 2011(184):53–91.

Butterworth, Brian and Geoffrey Beattie. 1978. Gesture and silence as indicators of planning in speech. In *Recent advances in the psychology of language*. Springer, pages 347–360.

Calzolari, Nicoletta, Claudia Soria, Riccardo Del Gratta, Sara Goggi, Valeria Quochi, Irene Russo, Khalid Choukri, Joseph Mariani, and Stelios Piperidis. 2010. The LREC map of language resources and technologies. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May. European Language Resources Association (ELRA).

Cassell, Justine et al. 2000. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. *Embodied conversational agents*, 1.

Cienki, Alan and Cornelia Müller. 2008. *Metaphor and gesture*, volume 3. John Benjamins Publishing.

Colletta, Jean-Marc, Michele Guidetti, Olga Capirci, Carla Cristilli, Ozlem Ece Demir, Ramona N. Kunene-Nicolas, and Susan Levine. 2015. Effects of age and language on co-speech gesture

production: an investigation of french, american, and italian children's narratives. *Journal of child language*, 42(1):122–145.
Cresti, Emanuela and Alessandro Panunzi. 2013. *Introduzione ai corpora dell'italiano*. Il mulino.
De Ruiter, Jan P., Adrian Bangerter, and Paula Dings. 2012. The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science*, 4(2):232–248.
Dittmann, Allen T. and Lynn G. Llewellyn. 1969. Body movement and speech rhythm in social conversation. *Journal of personality and social psychology*, 11(2):98.
Dobrogaev, Sergej M. 1929. Ucnenie o reflekse v problemakh iazykovedeniia [observations on reflexes and issues in language study]. *Iazykovedenie i materializm*, pages 105–173.
D'Errico, Francesca. 2019. 'Too humble and sad': The effect of humility and emotional display when a politician talks about a moral issue. *Social Science Information*, 58(4):660–680.
D'Errico, Francesca and Isabella Poggi. 2012. Blame the opponent! Effects of multimodal discrediting moves in public debates. *Cognitive Computation*, 4(4):460–476.
D'Errico, Francesca, Isabella Poggi, and Laura Vincze. 2012. Discrediting signals. A model of social evaluation to study discrediting moves in political debates. *Journal on Multimodal User Interfaces*, 6(3):163–178.
D'Errico, Francesca, Isabella Poggi, and Laura Vincze. 2013. Discrediting body. a multimodal strategy to spoil the other's image. In *Multimodal Communication in Political Speech Shaping Minds and Social Action: International Workshop, Political Speech*, volume 7688, pages 181–206. Springer, November.
Ekman, Paul and Wallace V. Friesen. 1972. Hand movements. *Journal of communication*, 22(4):353–374.
Gregersen, Tammy, Gabriela Olivares-Cuhat, and John Storm. 2009. An examination of l1 and l2 gesture use: What role does proficiency play? *The Modern Language Journal*, 93(2):195–208.
Hickmann, Maya. 2002. *Children's discourse: person, space and time across languages*, volume 98. Cambridge University Press.
Holler, Judith and Katie Wilkin. 2011. An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics*, 43(14):3522–3536.
Hostetter, Autumn B, Martha W Alibali, and Sotaro Kita. 2007. I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, 22(3):313–336.
Jezek, Elisabetta. 2003. *Classi di verbi italiani tra semantica e sintassi*. Bulzoni.
Jezek, Elisabetta, Bernardo Magnini, Anna Feltracco, Alessia Bianchini, and Octavian Popescu. 2014. T-PAS: A resource of corpus-derived typed predicate argument structures for linguistic analysis and semantic processing. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 890–895, Reykjavik, Iceland, May. European Language Resources Association (ELRA).
Kendon, Adam. 1972. Some relationships between body motion and speech. *Studies in dyadic communication*, 7(177):90.
Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge University Press.
Kendon, Adam, 2011. *Gesticulation and Speech: Two Aspects of the Process of Utterance*, pages 207–228. De Gruyter Mouton.
Kipp, Michael. 2001. Anvil - A generic annotation tool for multimodal dialogue. In *Seventh European Conference on Speech Communication and Technology*, pages 1367–1370, Aalborg, Denmark, September.
Knight, Dawn. 2011. *Multimodality and active listenership: A corpus approach*. A&C Black.
Krauss, Robert M. 1998. Why do we gesture when we speak? *Current directions in psychological science*, 7(2):54–54.
Krauss, Robert M. and Uri Hadar. 1999. The role of speech-related arm/hand gestures in word retrieval. *Gesture, speech, and sign*, 93.
Krippendorff, Klaus. 2018. *Content analysis: An introduction to its methodology*. Sage publications.
Lenci, Alessandro, Gabriella Lapesa, and Giulia Bonansinga. 2012. Lexit: A computational resource on italian argument structure. In *Proceedings of the eighth international conference on Language Resources and Evaluation (LREC2012)*, pages 3712–3718, Istanbul, Turkey, May.
Lin, Yen-Liang. 2017. Co-occurrence of speech and gestures: A multimodal corpus linguistic approach to intercultural interaction. *Journal of Pragmatics*, 117:155–167.

Lucisano, Pietro and Maria Emanuela Piemontese. 1988. Gulpease: una formula per la predizione della difficoltà dei testi in lingua italiana. *Scuola e città*, 3(31):110–124.
Mathet, Yann, Antoine Widlöcher, and Jean-Philippe Métivier. 2015. The unified and holistic method gamma (γ) for inter-annotator agreement measure and alignment. *Computational Linguistics*, 41(3):437–479.
McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago press.
McNeill, David. 2005. *Gesture and thought*. University of Chicago Press.
McNeill, David. 2008. *Gesture and thought*. University of Chicago press.
McNeill, David. 2016. *Why we gesture: The surprising role of hand movements in communication*. Cambridge University Press.
Menini, Stefano, Giovanni Moretti, Rachele Sprugnoli, and Sara Tonelli. 2020. Dadoeval@ evalita 2020: Same-genre and cross-genre dating of historical documents. In *7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. EVALITA 2020*, pages 391–397, Online, December. Accademia University Press.
Moneglia, Massimo, Susan Brown, Francesca Frontini, Gloria Gagliardi, Fahad Khan, Monica Monachini, and Alessandro Panunzi. 2014. The IMAGACT visual ontology. An extendable multilingual infrastructure for the representation of lexical encoding of action. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pages 3425–3432, Reykjavik, Iceland, May.
Morsella, Ezequiel and Robert M. Krauss. 2004. The role of gestures in spatial working memory and speech. *The American journal of psychology*, pages 411–424.
O'Keeffe, Anne and Michael McCarthy. 2010. *The Routledge handbook of corpus linguistics*. Routledge.
Ovendale, Alice. 2012. *The Role of Gesture in Cross-cultural and Cross-linguistic Learning Contexts: The Effect of Gesture on the Learning of Mathematics*. Ph.D. thesis, University of Johannesburg.
Parrill, Fey, Jennifer Bullen, and Huston Hoburg. 2010. Effects of input modality on speech–gesture integration. *Journal of Pragmatics*, 42(11):3130–3137.
Pianta, Emanuele, Luisa Bentivogli, and Christian Girardi. 2002. Multiwordnet: developing an aligned multilingual database. In *First international conference on global WordNet*, pages 293–302, Mysore, India, January.
Poggi, Isabella. 2007. *Mind, hands, face and body: a goal and belief view of multimodal communication*. Weidler.
Seiter, John S. and Weger Harry Jr. 2020. *Nonverbal communication in political debates*. Lexington Books.
Stam, Gale and Steven G. McCafferty. 2008. Gesture studies and second language acquisition: A review. *Gesture: second language acquistion and classroom research*, pages 3–24.
Tonelli, Sara, Rachele Sprugnoli, and Giovanni Moretti. 2019. Prendo la parola in questo consesso mondiale: A multi-genre 20th century corpus in the political domain. In *Proceedings of the Sixth Italian Conference on Computational Linguistics (CLIC-it 2019)*, Bari, Italy, November.
Tonelli, Sara, Ke Tran Manh, and Emanuele Pianta. 2012. Making readability indices readable. In *Proceedings of the First Workshop on Predicting and Improving Text Readability for target reader populations*, pages 40–48, Montréal, Canada, June. Association for Computational Linguistics.
Trotta, Daniela, Alessio Palmero Aprosio, Sara Tonelli, and Annibale Elia. 2020. Adding gesture, posture and facial displays to the PoliModal corpus of political interviews. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4320–4326, Marseille, France, May. European Language Resources Association.
Trotta, Daniela, Sara Tonelli, Alessio Palmero Aprosio, and Elia Annibale. 2019. Annotation and analysis of the polimodal corpus of political interviews. In *Sixth Italian Conference on Computational Linguistics (CLiC-it 2019)*, Bari, November.
Tsui, Amy B.M. 1994. *English conversation*. Oxford University Press.
Tuite, Kevin. 1993. The production of gesture. *Semiotica*, 93(1-2):83–105.
Vignozzi, Gianmarco. 2019. How gestures contribute to the meanings of idiomatic expressions and phrasal verbs in tv broadcast interviews: A multimodal analysis. *Lingue e Linguaggi*, 29.
Voghera, Miriam, 2001. *Teorie linguistiche e dati di parlato*, pages 75–96. Bulzoni, Roma.
Voghera, Miriam. 2020. What we learn about language from spoken corpus linguistics? *Caplletra. Revista Internacional de Filologia*, (69):125–154.
Wagner, Petra, Zofia Malisz, and Stefan Kopp. 2014. Gesture and speech in interaction: An overview. *Speech Communication*, 57:209–232.

Yoshioka, K. 2008. Linguistic and gestural introduction of inanimate referents in l1 and l2 narrative. *ESL & applied linguistics professional series*, pages 211–230.