

Are Gestures Worth a Thousand Words? An Analysis of Interviews in the Political Domain

Daniela Trotta

Università degli Studi di Salerno
Via Giovanni Paolo II 132,
Fisciano, Italy
dtrotta@unisa.it

Sara Tonelli

Fondazione Bruno Kessler
Via Sommarive 18
Trento, Italy
satonelli@fbk.eu

Abstract

Speaker gestures are semantically co-expressive with speech and serve different pragmatic functions to accompany oral modality. Therefore, gestures are an inseparable part of the language system: they may add clarity to discourse, can be employed to facilitate lexical retrieval and retain a turn in conversations, assist in verbalizing semantic content and facilitate speakers in coming up with the words they intend to say. This aspect is particularly relevant in political discourse, where speakers try to apply communication strategies that are both clear and persuasive using verbal and non-verbal cues.

In this paper we investigate the co-speech gestures of several Italian politicians during face-to-face interviews using a multimodal linguistic approach. We first enrich an existing corpus with a novel annotation layer capturing the function of hand movements. Then, we perform an analysis of the corpus, focusing in particular on the relationship between hand movements and other information layers such as the political party or non-lexical and semi-lexical tags. We observe that the recorded differences pertain more to single politicians than to the party they belong to, and that hand movements tend to occur frequently with semi-lexical phenomena, supporting the lexical retrieval hypothesis.

1 Introduction

A bodily 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 et al., 2014).

Some authors (McNeill, 2005; Kendon, 2004) have considered gestures as an integrative, inseparable part of the language system. Indeed gestures

may provide important information or significance to the accompanying speech and add clarity to the children’s narrative (Colletta et al., 2015); they can be employed to facilitate lexical retrieval and retain a turn in conversations (Stam, 2008) and assist in verbalizing semantic content (Hostetter et al., 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 feature, long enough for the process of word production to take place (Morsella and Krauss, 2004).

Gestures can also convey semantic meanings. For example, Müller et al. (2013) discuss the principles of meaning creation and the simultaneous and linear structures of gesture forms. In this framework, they propose individual aspects of a “grammar” of gestures and conclude that in gestures we can find the seeds of language or the embodied potential of hand-movements for developing linguistic structures. As pointed out by Lin (2017) the link between speech and gesture can be explained by two gesture-speech characteristics: semantic coherence, i.e. combining gesture with meaningful and related speech, and temporal synchrony, i.e. producing gesture in synchrony with speech (Butcher, 2000). The role of synchronization is particularly relevant for the creation of multimodal resources (Allwood, 2008), because it allows researchers to overcome one of the historical limits of traditional corpora that are in one modality (either written or spoken): presenting data in a single format offers limited opportunities for exploring non-verbal, gestural features of discourse, while they are important aspects to understand intercultural face-to-face interaction (Adolphs and Carter, 2013; Knight, 2011).

Nevertheless, Beattie and Shovelton (1999) have shown that most of the time gestures are produced before the linguistic item to which they are related, defining this phenomenon “temporal asynchrony”.

Also [Butterworth and Beattie \(1978\)](#) presented some empirical evidence to prove that temporal asynchrony between gestures and speech was more common in spontaneous speech and that hand gestures were associated with low-frequency unpredictable lexical items, i.e., lexical items that were more difficult for speakers to reach in the course of language production ([Goldman-Eisler, 1958](#); [Beattie and Butterworth, 1979](#)). Their conclusion was the following: “Gestures are products of lexical preplanning processes, and seem to indicate that the speaker knows in advance the semantic specification of the words he will utter, and in some cases has to delay if he has to search for a relatively unavailable item” ([Butterworth and Beattie, 1978](#), p. 358).

Research on spoken interaction has suggested that non-verbal communication is currently the least understood and analyzed aspect of communication, despite recognizing its equal importance ([Knight, 2011](#); [McNeill, 2016](#)). For this reason, we believe it is very important to carry out studies on gesture-talk interaction and develop multimodal corpora.

Our study focuses in particular on the relationships between the co-occurrence of speech and gesture in Italian in the specific case of political interviews since: *i*) television interviews are inherently multimodal and multisemiotic texts, in which meaning is created through the co-presence of visual elements, verbal language, gestures, and other semiotic cues ([Vignozzi, 2019](#)); *ii*) linguistic studies in the political domain can be of interest also beyond the NLP community, for example in political science and communication studies, and *iii*) the Italian political scene has been little studied from this perspective.

In particular, in the following sections we address research questions such as:

1. Are there semantic patterns of gesture-speech relationship?
2. Does political party affiliation influence this relationship?
3. Does the presence of gesturing indicate problems with the retrieval of words during speech?

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

communication models interact. We also release the corpus of political interviews with the new annotation layer encoding the functions of hand movements at <https://github.com/dhfbk/InMezzoraDataset>.

2 Political and multimodal corpora in Italian

In recent years, political language has received increasing attention, especially in English, since it is possible to have free access to speech transcriptions from UK and US government portals and personal foundation websites such as the White House portal, William J. Clinton Foundation, Margaret Thatcher Foundation. This has fostered research on political and media communication and persuasion strategies ([Guerini et al., 2010](#); [Esposito et al., 2015](#)). As for Italian, which is the language of interest for this study, only few corpora in the political domain are available.

One of the first experiments was the CorpusB ([Bolasco et al., 2006](#)) composed by 111 speeches by the former Prime Minister Silvio Berlusconi – for a total of 325,000 tokens – created to study the evolution of Berlusconi’s political language from the moment when he started his political career in 1994 until the last programmatic speech of his third government in 2006. Subsequently, the work of [Salvati and Pettorino \(2010\)](#) analyzed diachronically some of the suprasegmental aspects of Berlusconi’s speeches from 1994 to 2010, including an analysis of the length of logical chains, the number of syllables per chain, the maximum and minimum pitch and frequency of speech, average duration of empty pauses, fluency and tonal range.

Among the most recent corpora made available in Italian, the largest one includes around 3,000 public documents by Alcide De Gasperi ([Tonelli et al., 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 \(2017, p. 157\)](#) “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”.

This is also confirmed by the fact that – to date –

exist 286 multimodal resources certified for all languages by the LRE map¹ but only one is in Italian, i.e. IMAGACT a corpus-based ontology of action concepts, derived from English and Italian spontaneous speech (Moneglia et al., 2014; Bartolini et al., 2014). So both from the political and the multimodal point of view, this language is not well represented.

In an attempt to fill this gap, we first developed the PoliModal corpus (Trotta et al., 2019, 2020), containing the 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 annotation has been done using XML as markup language and following the TEI standard for Speech Transcripts in terms of utterances to keep track of so-called “speech constants” (Voghera, 2001). In particular, the corpus contains the annotation of the following hesitation phenomena:

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

(b) **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;

(c) **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;

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

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

The corpus includes also the annotation of facial displays, hand gesture and body posture, which we carried out using the MUMIN coding scheme annotation (Allwood et al., 2007) and ANVIL (Kipp, 2001) a tool for the automatic annotation of audiovisual material containing multimodal dialogue. This corpus is considered as a starting point for our study.

¹A mechanism to monitor the use and creation of language resources by collecting information on both existing and newly-created resources, freely available at <http://lremap.elra.info/>

3 Annotation of Hand Movements

Starting from PoliModal corpus described in Section 2, 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 et al., 2014). Based on the seminal works by Kendon (1972, 1980) about the relationship between body motion and speech on the one hand, and about gesticulation and speech in the process of utterance on the other, they are usually separated into several gestural phases: initial/rest position, preparation, (pre-stroke) hold, stroke, (post-stroke) hold, retraction/recovery and rest position (Kita, 1990; Kendon, 2004; Bressemer and Ladewig, 2011; Ladewig and Bressemer, 2013). Note that all gestures are not necessarily constituted by all these phases and that some phases may also be duplicated.

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 manually added an additional information layer to those already present – following the classification proposed by Lin (2017) adapting Colletta et al. (2015) and Kendon (1972)² – 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

²We specify that one of the first classifications of gestures was proposed by Ekman and Friesen (1969) that classified kinesic behavior into four broad categories: (1) emblems (“are those nonverbal acts which have a direct verbal translation, or dictionary definition, usually consisting of a word or two, or perhaps a phrase” (Ekman and Friesen, 1969, p. 63)), (2) illustrators (“they are movements which are directly tied to speech, serving to illustrate what is being said verbally” (Ekman and Friesen, 1969, p. 68)), (3) affect displays (“can be related to verbal behaviour in a number of ways. They can repeat, qualify or contradict a verbally stated affect, or be a separate, unrelated channel of communication” (Ekman and Friesen, 1969, p. 77)), and (4) regulators (“These are acts which maintain and regulate the back-and-forth nature of speaking and listening between two or more interactants” (Ekman and Friesen, 1969, p. 82)).

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 Parliament”) 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. Following this principle, each annotator looked at the portion of the video in which the hand movements were occurring and depending on the meaning that he/she thought the gesture had in

that particular context of utterance, attributed the corresponding semantic 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 these 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 they consider prevalent in the context of use. When the gesture-speech relationship appears too vague, it is good practice to conduct interviews with speakers to confirm the interpretation of gesture meanings. In fact, as suggested by Kochman et al. (2014), through multiple methods of data analysis, such as triangulation, we can test whether interpretations of the results were consistent and internally coherent.

In our case, since such checking is not possible, we try to ensure a high-quality and consistent annotation by computing inter-annotator agreement. Specifically we perform 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 with the tag ‘semantic_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.

4 Description of gesture-speech annotation

A summary of the hand gesture annotations in the corpus is reported in Table 1 and 2. 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), Five-Star Movement (center-populist), Lega (right-populist), Casa Pound (right), Popolo della Libertà (center-right). The ‘Contradic-

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

Table 1: Frequency of the type of gestures annotated for each interviewee.

Political party	Integrat.	Reinforc.	Supplement.	Complement.	Other
PD	117	17	2	91	51
M5S	35	1	1	29	1
Lega	36	17	4	20	6
Casa Pound	5	0	2	3	0
Il Popolo delle Libertà	21	11	1	16	8
Total	214	46	10	159	66

Table 2: Frequency of the type of gestures for each political party

```
<u gender="m" length="725" role="
"politician" who="Walter Veltroni">
Vede la cosa che più mi ha colpito
in queste ore è<vocal desc="ehm"
type="semi-lexical" />l'immagine
del corteo che c'è stato in Polonia
ieri. Ieri c'è stato.<del type="
>falseStart" /><movement start="
"158.8" end="162.67999" attribute="
"Hand movement trajectory"
attribute_text="complex"
semantic_function="complementary">
Nessuno se ne è occupato,
telegiornali, giornali, pochissimi.
</movement></u>
```

Figure 1: Example of the new annotation level in xml

tory' 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 move-

ments, 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'.

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. So – since some socio-cultural factors seem to influence the production of gestures – we wonder whether political party affiliation is among them. Then as a next step, we investigate whether there is a significant relationship between the political party of affiliation of interviewees in the dataset and the type of gestures used, and thus whether the political party of affiliation affects the choice of gesture categories. The political parties considered are the ones reported in Table 2. We compute one-way ANOVA with independent samples. The results obtained suggest that the null hypothesis cannot be rejected since the significance value obtained is $p = 0.11$. Therefore we can conclude that the party affiliation does not play any role in the combination of gestures and speech.

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. A 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 per altro** [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 idiot.”)

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.

5 Lexical Retrieval hypothesis

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 dysfluent 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 (Guiasu, 1977) between hand movements and each of the speech disfluencies reported in Table 3. 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 3 the tag incidence per 100 turns for each interview considered.

Among the politicians included in this dataset, the one that most accompanies his speech with hand gestures 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 3. The values obtained are shown in the heatmap reported in Figure 2, with lighter colors corresponding to higher WMI values.

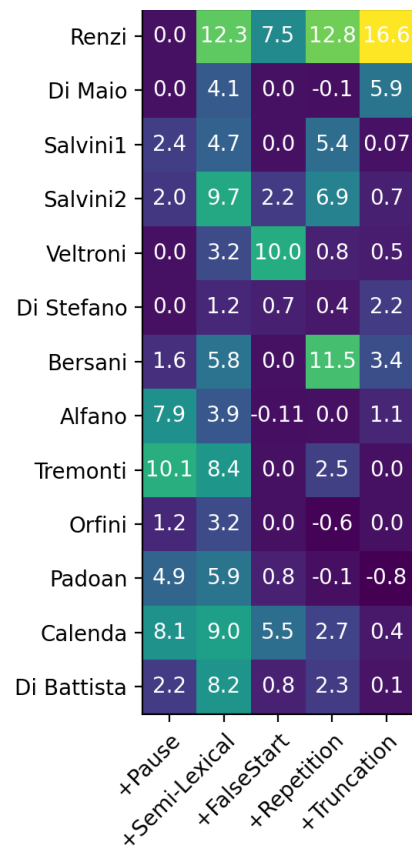


Figure 2: WMI values between hand movements and tags reported on the x-axis for each interviewee on the y-axis

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).

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

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

Table 3: Tag incidence per 100 turns for each interview

evident differences in gesturing behaviour 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 behaviour that is unique compared to all the other interviewees, with hand gestures that are almost always used in association with other speech phenomena.

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.

6 Conclusions

In this work, we investigate co-speech gestures of several Italian politicians during face-to-face interviews. To this purpose, we enrich an existing corpus with labels describing the semantic type of the different hand movements. 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 spoken information. We also show that party affiliation does not significantly influence the gesture-speech relationship. Finally we test the *Lexical Retrieval Hypothesis* by computing the association between hand movements produced by each interviewee and speech disfluen-

cies 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 the future we plan to conduct further analyses aimed at understanding whether such gestures co-occur with specific types of words (e.g. copulative verbs, predicative verbs, etc.) and whether other linguistic or socio-linguistic variables such as language complexity or age influence the use of hand movements and their semantic functions.

References

- Svenja Adolphs and Ronald Carter. 2013. *Spoken corpus linguistics: From monomodal to multimodal*. Routledge.
- Jens Allwood. 2008. *Multimodal corpora*. Mouton de Gruyter.
- Jens Allwood, 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.
- Roberto Bartolini, Valeria Quochi, Irene De Felice, Irene Russo, and Monica Monachini. 2014. From synsets to videos: Enriching italwordnet multimodally. In *Proceedings of LREC*, pages 3110–3117.
- Geoffrey Beattie and Heather Shovelton. 1999. Do iconic hand gestures really contribute anything to the

- semantic information conveyed by speech? An experimental investigation. *Semiotica*, 123(1-2):1–30.
- Geoffrey W Beattie and Brian L Butterworth. 1979. Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and speech*, 22(3):201–211.
- Sergio Bolasco, Nora Galli de'Paratesi, and Luca Giuliano. 2006. *Parole in libertà: un'analisi statistica e linguistica dei discorsi di Berlusconi*. Manifestolibri.
- Jana Bressemer and Silva H Ladewig. 2011. Rethinking gesture phases: Articulatory features of gestural movement? *Semiotica*, 2011(184):53–91.
- Cynthia Butcher. 2000. Two-word speech: When hand and mouth come together. *Language and gesture*, 2:235.
- Brian Butterworth and Geoffrey Beattie. 1978. Gesture and silence as indicators of planning in speech. In *Recent advances in the psychology of language*, pages 347–360. Springer.
- Jean-Marc Colletta, 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.
- Allen T Dittmann and Lynn G Llewellyn. 1969. Body movement and speech rhythm in social conversation. *Journal of personality and social psychology*, 11(2):98.
- SM Dobrogaev. 1929. Uchenie o refleksie v problemakh iazykovedeniia [observations on reflexes and issues in language study]. *Iazykovedenie i materializm*, pages 105–173.
- Paul Ekman and Wallace V. Friesen. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1):49–98.
- Fabrizio Esposito, Pierpaolo Basile, Francesco Cutugno, and Marco Venuti. 2015. The CompWHoB Corpus: Computational construction, annotation and linguistic analysis of the white house press briefings corpus. *Proceedings of CLiC-it*.
- Frieda Goldman-Eisler. 1958. The predictability of words in context and the length of pauses in speech. *Language and speech*, 1(3):226–231.
- Marco Guerini, Danilo Giampiccolo, Giovanni Moretti, Rachele Sprugnoli, and Carlo Strapparava. 2010. The new release of Corps: A corpus of political speeches annotated with audience reactions. In *International Workshop on Political Speech*, pages 86–98. Springer.
- Silviu Guiasu. 1977. *Information Theory with Applications*. McGraw-Hill, New York.
- Maya Hickmann. 2002. *Children's discourse: person, space and time across languages*, volume 98. Cambridge University Press.
- Autumn B Hostetter, 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.
- Adam Kendon. 1972. Some relationships between body motion and speech. *Studies in dyadic communication*, 7(177):90.
- Adam Kendon. 1980. Gesticulation and speech: Two aspects of the Process of Utterance. *The relationship of verbal and nonverbal communication*, (25):207.
- Adam Kendon. 2004. *Gesture: Visible action as utterance*. Cambridge University Press.
- Michael Kipp. 2001. Anvil – A generic annotation tool for multimodal dialogue. In *Seventh European Conference on Speech Communication and Technology*.
- Sotaro Kita. 1990. The temporal relationship between gesture and speech: A study of Japanese-English bilinguals. *Masters Thesis, Department of Psychology, University of Chicago*.
- Dawn Knight. 2011. *Multimodality and active listener-ship: A corpus approach*. A&C Black.
- Katty Kochman, Dirk Moelants, Marc Leman, Nick Bailey, and Jane Ginsborg. 2014. Gesture as a communicative tool in vocal pedagogy. *Journal of Interdisciplinary music studies*.
- Robert M Krauss. 1998. Why do we gesture when we speak? *Current directions in psychological science*, 7(2):54–54.
- Robert M Krauss and Uri Hadar. 1999. The role of speech-related arm/hand gestures in word retrieval. *Gesture, speech, and sign*, 93.
- Silva H Ladewig and Jana Bressemer. 2013. 69. a linguistic perspective on the notation of gesture phases. In *Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (HSK) 38/1*, pages 1060–1079. De Gruyter Mouton.
- Yen-Liang Lin. 2017. Co-occurrence of speech and gestures: A multimodal corpus linguistic approach to intercultural interaction. *Journal of Pragmatics*, 117:155–167.
- D. McNeill. 2005. *Gesture and thought*. University of Chicago Press.
- David McNeill. 2016. *Why we gesture: The surprising role of hand movements in communication*. Cambridge University Press.

- Stefano Menini, 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. Academia University Press.
- Massimo Moneglia, 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 (LREC'14)*, pages 3425–3432, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Ezequiel Morsella and Robert M Krauss. 2004. The role of gestures in spatial working memory and speech. *The American journal of psychology*, pages 411–424.
- Cornelia Müller, Jana Bressemer, and Silva H Ladewig. 2013. 45. towards a grammar of gestures: A form-based view. In *Volume 1*, pages 707–733. De Gruyter Mouton.
- Luisa Salvati and Massimo Pettorino. 2010. A diachronic analysis of face-to-face discussions: Berlusconi, fifteen years later. In *International Workshop on Political Speech*, pages 65–74. Springer.
- Sara Tonelli, 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 CLIC-it*.
- Daniela Trotta, Alessio Palmero Aprosio, Sara Tonelli, and Elia Annibale. 2020. Adding gesture, posture and facial displays to the polimodal corpus of political interviews. In *12th Language Resources and Evaluation Conference (LREC 2020)*, pages 4320–4326. European Language Resources Association.
- Daniela Trotta, 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)*.
- Amy BM Tsui. 1994. *English conversation*. Oxford University Press.
- Gianmarco Vignozzi. 2019. How gestures contribute to the meanings of idiomatic expressions and phrasal verbs in tv broadcast interviews: A multimodal analysis. *Lingue e Linguaggi*, 29.
- Miriam Voghera. 2001. Teorie linguistiche e dati di parlato. In *Dati empirici e teorie linguistiche*, Congresso Internazionale di Studi della Società di linguistica italiana, pages 75–95, Roma. Bulzoni.
- Petra Wagner, Zofia Malisz, and Stefan Kopp. 2014. Gesture and speech in interaction: An overview.
- K Yoshioka. 2008. Linguistic and gestural introduction of inanimate referents in I1 and I2 narrative. *ESL & applied linguistics professional series*, pages 211–230.