Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 4320–4326
Marseille, 11–16 May 2020

Adding Gesture, Posture and Facial Displays to the PoliModal Corpus of Political Interviews

Daniela Trotta[1], Alessio Palmero Aprosio[2], Sara Tonelli[2] and Annibale Elia[1]
[1] Università di Salerno [2] Fondazione Bruno Kessler
{dtrotta,elia}@unisa.it {aprosio,satonelli}@fbk.eu

Abstract

This paper introduces a multimodal corpus in the political domain, which on top of transcribed face-to-face interviews presents the annotation of facial displays, hand gestures and body posture. While the fully annotated corpus consists of 3 interviews for a total of 120 minutes, it is extracted from a larger available corpus of 56 face-to-face interviews (14 hours) that has been manually annotated with information about metadata (i.e. tools used for the transcription, link to the interview etc.), pauses (used to mark a pause either between or within utterances), vocal expressions (marking non-lexical expressions such as burp and semi-lexical expressions such as primary interjections), deletions (false starts, repetitions and truncated words) and overlaps. In this work, we describe the additional level of annotation relating to non-verbal elements used by three Italian politicians belonging to three different political parties and who at the time of the talk-show were all candidates for the presidency of the Council of Ministers. We also present the results of some analyses aimed at identifying existing relations between the proxemic phenomena and the linguistic structures in which they occur, in order to capture recurring patterns and differences in the communication strategy.

Keywords: multimodal corpora, political communication, multi-layered annotation

1. Introduction

In the context of a political interview, the host, typically a journalist, acts as a mediator, a representative of the audience (Koutsombogera and Papageorgiou, 2010). This means that, if a politician manages to convince or deal with the criticism that the host addresses, then her/his trustworthiness, reliability and credibility will be easily established. In this situation, a politician is judged not only based on one's arguments and rhetorical choices, but also on the attitude, self-confidence, and in general on an overall convincing behaviour. For example, if a politician seems to be conversationally dominant and manages interruptions to a satisfactory degree, it is more likely that the host, and therefore the audience, will be convinced by the arguments put forward by the interviewee.

In televised political interviews, politicians struggle to establish an image for themselves as competent personalities, a goal which is considered as important as the topic under discussion. For this reason, analysing the combination of verbal and non-verbal elements (such as their image, including their non-verbal interactional behavior) is crucial and could be very interesting for scholars in political science and communication science, and in general to study consensus mechanisms. Indeed politicians make use of non-verbal means to express positive or negative evaluations towards persons or facts and thus raise emotions to the public through means that are absent in speech. At the same time, they have to confront the interviewers' behavior, challenges and comments and, in a way, survive the turn competition, i.e. strive to have the floor and thus be able to support their opinions and arguments. In this perspective, a common phenomenon in political interviews is the issue of conversational dominance, i.e. a speaker's tendency to control the other speaker's conversational actions over the course of an interaction (Itakura, 2001). In order to better understand this aspect, as well as other linguistic phenomena related to persuasion in political speeches, we are developing a multimodal corpus in the political domain, which on top of transcribed face-to-face interviews presents the annotation of facial displays, hand gestures and body posture. By 'multimodal' we mean that the corpus is composed of manual transcriptions of interviews broadcast on TV and annotated with information not only about the linguistic structure of the utterances but also about non-verbal expressions[1]. The corpus, which we call PoliModal, addresses the need to make up for the lack of Italian linguistic resources for political-institutional communication and is annotated in XML following the standard for the transcriptions of speech TEI Guidelines for Electronic Text Encoding and Interchange[2] and the MUMIN coding scheme for the annotation of facial displays, hand gesture and body posture (Allwood et al., 2007).

At the moment, the PoliModal corpus includes 56 face-to-face interviews (14 hours), which have been manually annotated with information about metadata (i.e. tools used for the transcription, a link to the interview etc.), pauses (used to mark a pause either between or within utterances), vocal expressions (marking non-lexical expressions such as burp and semi-lexical expressions such as primary interjections), deletions (false starts, repetitions and truncated words) and overlaps. Details on this annotation have been reported in (Trotta et al., 2019). Three of these interviews have been enriched with an additional level of annotation related to the

[1] According to (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".

[2] P5: Guidelines for Electronic Text Encoding and Interchange. See more `https://tei-c.org/release/doc/tei-p5-doc/en/html/TS.html#TSSAPA`

gestures and facial expressions used by three Italian politicians belonging to three different political parties during the interviews, and who at the time of the talk-show were candidates for the presidency of the Council of Ministers. The addition of this novel annotation layer is aimed at investigating existing relations between the proxemic phenomena noted and the linguistic structures in which they occur, in order to capture recurring patterns and significant differences in the gestural strategy of each interviewee. The main contributions of this paper are: *i)* the presentation of this additional annotation layer with gestures, facial expressions and posture, *ii)* the release of the three newly annotated interviews at the link `https://github.com/dhfbk/InMezzoraDataset`; *iii)* the first analysis of the relation between annotation layers, to investigate how gesture, posture and face annotation interact with lexical, semi-lexical and non-lexical information.

2. Related work

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 regards Italian, which is the language of interest for this study, only few corpora in the political domain are available. Indeed according to LRE Map[3] there are currently 24 monolingual corpora for Italian, two for the spoken language, i.e. VoLIP (Alfano et al., 2014) and the LUNA corpus (Dinarelli et al., 2009), and the other accounting for written documents. However, none of them pertains to the political domain. Furthermore, between the 286 multimodal resources certified for all the languages by the LRE map, 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 both from the political and the multimodal point of view, this language is not well represented. Although some studies related to corpus-based analysis of political discourse do exist also for Italian, they mainly deal with monological discourse (Bolasco et al., 2006; Cedroni, 2010; Longobardi, 2010; Catellani et al., 2010; Bongelli et al., 2010; Zurloni and Anolli, 2010; Sprugnoli et al., 2016; Moretti et al., 2016) and do not make the data available for further studies.

Concerning political corpora developed specifically for conversation analysis, other languages have been more extensively studied. In (Bigi et al., 2011), for example, the authors present a multimodal corpus of political debates at the French National Assembly, on May 4th, 2010 and introduce an annotation scheme for a political debate dataset which is mainly in the form of video and audio annotations. (Navarretta and Paggio, 2010) deal with the identification of interlocutors via speech and gestures in annotated televised political debates in British and American English. Other papers have focused primarily on visual aspects (gaze, gestures, facial expressions) of communicative interaction during political talk shows or parliamentary speeches (D'Errico et al., 2010). The most similar approach to ours is presented in (Koutsombogera and Papageorgiou, 2010), in which the authors analyse a Greek multimodal corpus of 10 face-to-face television interviews focusing on non-verbal aspects in order to study the attempts of persuasion and interruption during political interviews. Their work, however, is mainly aimed at studying the strategies for conversational dominance, and annotate specific traits accordingly. Our work, instead, is more general, includes a different set of tags and integrates also other linguistic features.

3. Description of the PoliModal corpus

In this section we briefly describe the PoliModal corpus (Trotta et al., 2019), on top of which we have added a novel annotation layer. The corpus includes 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 from 24 September 2017 to 14 January 2018. The show follows a fixed format, with interviews conducted by a journalist, Lucia Annunziata, to a guest, typically a prominent figure in the political or cultural scene. Each interview usually is done in the same limited time frame, 30 minutes (except few cases e.g. Matteo Renzi), and no audience is present, so that applause and any other type of reactions are not included in the corpus. The audio signal has been transcribed using a semi-supervised speech-to-text methodology (Google API + manual correction). All hesitations, repetitions and interruptions of the original interview have been included. The output has been further segmented into turns, and punctuation has been added, mainly to delimit sentence boundaries when they were not ambiguous. In PoliModal, annotation has been done using XML as markup language and following the TEI standard for Speech Transcripts in terms of utterances. The linguistic resource has currently 100,870 tokens and includes interviews to politicians covering all the Italian political spectrum. Beside politicians, also a small number of people with different backgrounds (students, academics, judges, economists, etc.) has been interviewed and is therefore included in the corpus.

For each interview the following information was manually annotated and is included in the XML resource file:

(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. Speakers differ very much in their rhythm and in particular in the amount of time they leave between words, so the following element is provided to mark occasions where the transcriber judges that a speech has been paused, irrespective of the actual amount of silence.

[3]LRE Map is a mechanism intended to monitor the use and creation of language resources by collecting information on both existing and newly-created resources, free available at `http://lremap.elra.info/`

Behaviour attribute	Behaviour value
General face	Smile, Laugh, Scowl, Other
Eyebrow movement	Frown, Raise, Other
Eye movement	Extra-Open, Close-Both, Close-One, Close-Repeated, Other
Gaze direction	Towards-Interlocutor, Up, Down, Sideways, Other
Mouth openness	Open mouth, Closed mouth
Lip position	Corners up, Corners down, Protruded, Retracted
Head movement	Down, Down-Repeated, BackUp, BackUpRepeated, BackUp-Slow, Forward, Back, Side-Tilt, Side-TiltRepeated, Side-Turn, Side-Turn-Repeated, Waggle, Other
Handedness	Both hands, Single hands
Hand movement trajectory	Up, Down, Sideways, Complex, Other
Body posture	Towards-Interlocutor, Up, Down, Sideways, Other

Table 1: List of gestures, following the list described in (Allwood et al., 2007). The presence of an underline means that the gesture has been found in our dataset (see Table 3).

(c) **vocal**: with this tag we mark any vocalized but not necessarily lexical phenomenon, for example non-lexical expressions (i.e. burp, click, throat, etc.) and semi-lexical expressions (i.e. ah, aha, aw, eh, ehm etc.).
(d) **del**: this tag covers different phenomena of speech management, specifically false starts, repetitions and truncated words. Since they are marked in the TEI Guidelines as 'editorially deleted', the corresponding tag is **del**.
(e) **overlap**: this phenomenon is present when the speaker conveys (in a verbal or non-verbal manner) that he/she is about to finish his/her turn and the co-locutor starts speaking so that there is a slight overlap of utterances.

4. New annotation layer

The goal of the novel annotation layer added on top of the PoliModal corpus was to enrich it with an additional mode and therefore a new level of meaning, expressed through facial displays, hand gesture and body posture. Adding this kind of information is very time-consuming, since it requires that the annotator watches the video interviews and marks traits derived from the video, while aligning them to the underlying text which was already transcribed. The novel annotation was therefore limited to a subset of 3 interviews with three politicians belonging to different political parties: Matteo Renzi, from the center-left party Partito Democratico, Matteo Salvini, from the right-wing party Lega, and Luigi di Maio, from the populist party Movimento Cinque Stelle. When the interviews took place, they were candidates for the presidency of the Council of Ministers.[4] Being therefore competitors on the Italian political scene, they had to establish an image for themselves as competent personalities, a goal which is considered equally important to the topic under discussion (Koutsombogera and Papageorgiou, 2010). At the same time, they had to respond to the interviewers' challenges and comments presenting their arguments and opinions in a persuasive way.
In the paper by (Allwood, 2008), the authors highlight that synchronization of information in different modalities is a crucial issue in assembling a multimodal corpus. Therefore tha 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 (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[5], and the CHIL project[6]), we did it manually since the recording and transcription of the corpus were done before knowing what layers would be exactly annotated. The video annotation was carried out using the ANVIL tool (Kipp, 2001) while the levels and labels used in the annotation scheme are mainly inspired by the MUMIN coding scheme notation (Allwood et al., 2007).
Table 1 summarizes the list of gestures, as described in (Allwood et al., 2007). The annotation – made at the moment by a single expert annotator – follows the criterion highlighted by (Allwood et al., 2007), claiming that anno-

[4]The Italian political elections referred to in the paper were held on Sunday, March 4, 2018. They followed the dissolution of the Chambers, which took place by decree of the President of the Republic Sergio Mattarella on December 28, 2017, a short time before the natural expiry of the 17th legislature, scheduled for March 14, 2018. The results saw the centre-right establish itself as the most voted coalition, with about 37 percent of the preferences, while the single most voted list, the Movimento 5 Stelle, collected more than 32 percent of the votes.

[5]http://www.amiproject.org
[6]http://chil.server.de/servlet/is/101/

tators are expected to select gestures[7] 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. For example, mechanical recurrent blinking due to dryness of the eye might not be annotated because it does not seem in a given context to have a communicative function. As regards the annotation guidelines - as specified in (Allwood et al., 2007) - the attributes concerning the shape or dynamics of the observed phenomena are not fine-grained, because they only seek to capture features that are significant with respect to the functional level of the annotation. Once a gesture has been selected by an annotator because of its communicative role, it is annotated with functional values, as well as features that describe its behavioural shape and dynamics: this is what we call the modality-specific annotation level. An additional, multimodal annotation level concerns the relation that the gesture has either with other gestures or with the speech modality. The scheme provides a number of simple categories for the representation of multimodal relations. However, it does not include tags for the specific annotation of verbal expressions since its focus is on the study of gestures, which is why we have integrated them in order to study - in the future - the relationship between verbal and non-verbal expressions. Following this principle, we do not annotate all gestures, focusing on what follows:

(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.

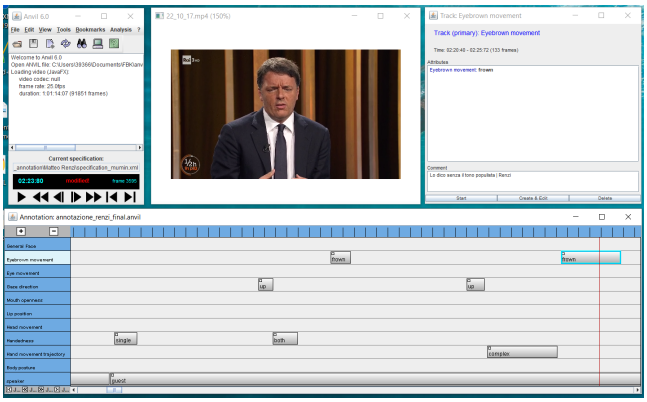

Figure 1: Example of facial display: frowning.

As an example, we report in Fig.1 the annotation of the interview to Matteo Renzi. The leader of Partito Democratico frowns when discussing the defeat of his proposal in the constitutional referendum, at minute 00:02:23:80. This gesture, which - according to (Poggi, 2005) - can take on four main meanings (surprise, emphasis, contrasting, perplexity/doubt), takes here a contrasting meaning, because it occurs when the politician expresses his disagreement with what the interviewer just said about the referendum. Renzi's words uttered when making this facial expression are:

"Però, giusto per non perdere l'abitudine, non è che sia d'accordissimo sulla lettura che lei dà, nel senso che il referendum l'ho perso io."

En: *"But - just so as not to lose the habit - I don't agree with your interpretation, that is, I lost the referendum."*

(b) **Hand gesture**: we follow a simplification of the scheme from the McNeill Lab (Duncan, 2004). 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.

In Fig.2 we show an example annotation of hand movements, in particular the use of both hands. At minute 00:01:55:72 Matteo Renzi, still discussing the defeat at the referendum, uses both hands – which could assume a batonic value [8] in this circumstance – while uttering the following sentence:

"Io quei politici che tutte le volte danno la responsabilità, la colpa, si nascondono dietro gli alibi personalmente non li sopporto."

En: *"Personally, I can't tolerate those politicians who always blame and hide behind alibis."*

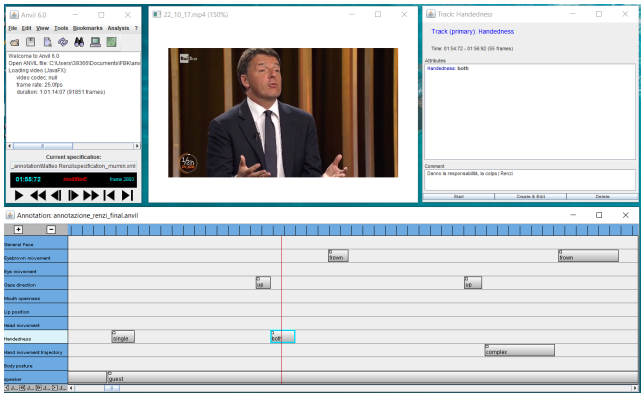

Figure 2: Example of hand gesture: double-handed.

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

Fig.3 shows a third example – taken again from the interview to Matteo Renzi, at minute 00:00:41:12 – in which the position of the interviewee's bust appears slightly sideways. In this case, the gesture occurs while the interviewee listens to a question and therefore outside of a sen-

[7](Duncan, 2004) defines a gesture as a movement that is always characterised by a stroke, and may also go through a preparation and a retraction phase. Each stroke corresponds in MUMIN to an independent gesture.

[8]According to (Allwood et al., 2007): "*baton gestures* are those in which the hands move rhythmically from top to bottom to scan and emphasize the accented syllables in a sentence".

tence. This annotation is therefore temporally aligned with the transcribed turn of the journalist.

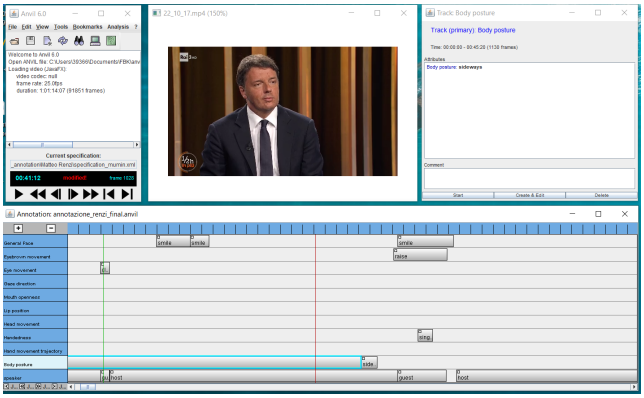

Figure 3: Example of body posture: sideways.

5. Corpus Statistics and Discussion

While the annotation is still ongoing to extend the corpus size, we report in this Section some statistics related to the three interviews that have been first completed. Table 2 shows the number of turns and the overall duration of the interview for each politician. The duration refers only to the interviewees' utterances, therefore excluding the time used by the journalist to make questions. The interviews to Luigi Di Maio and Matteo Salvini have a comparable duration both in terms of time and of turns. The interview to Matteo Renzi, instead, is longer (1 hour in total) but the turns are considerably shorter because he was being interrupted more frequently by the interviewer.

Politician	Turns	Duration (sec.)
Matteo Renzi	149	2143.32
Luigi Di Maio	30	1113.92
Matteo Salvini	29	1070.28

Table 2: Corpus content: turns per speaker and total duration

As regards the new annotation layer, we report in Table 3 the statistics for all annotated phenomena in the three interviews. Some traits that are present in the annotation scheme have not been reported because they have not been observed in any of the three interviews. For example, no occurrences of the extra-open and close-one eye movement types have been observed, nor the scowl among the facial expressions. Overall, Matteo Renzi shows the highest expressiveness through the use of gestures, facial displays and posture, with more than double occurrences compared to the opponents.

An interesting phenomenon is the movement of eyebrows, which has been extensively discussed also in the literature. In particular the frowning of the eyebrows, which as (Poggi, 2006) suggests indicates the rapprochement of the eyebrows, forming vertical wrinkles on the forehead, may be used for a range of purposes, such as asking a question, communicating to an interlocutor that that s/he is not clear, expressing indirectly disagreement with the other party, looking at something very carefully, trying to remember something, asserting something with confidence, expressing concern or anger about something, giving a peremptory order.

In our specific case, the politician who shows the most frowning (30) is Matteo Renzi, and from the context we can argue that this signal is used by the former Prime Minister to show confidence in his assertions and exhibit attention to what is being said. The raising of the eyebrows – defined by (Poggi, 2006) as "a signal of the gaze that is produced by lifting both eyebrows in a symmetrical manner" – may instead take on four main meanings: surprise, emphasis, adversity, perplexity/doubt. The semantic element shared by all these interpretations is the presence of new information, as a matter of unexpected knowledge.

Overall, the different communication strategies adopted by the three politicians are evident in the corpus: Matteo Renzi's gesture, facial displays and body posture express an extrovert attitude, but also an evident attempt to please the audience and to be convincing at all costs. This is confirmed also by the lexical and semi-lexical traits annotated in this interview that include a high number of repetitions and truncations (0.21 and 0.37 per turn on average, respectively) and no pauses, as if the interviewee could not organise well the discourse and was too much involved in trying to convince the audience.

On the contrary, Luigi di Maio shows only 0.19 repetitions and 0.19 truncations per turn on average, while gaze, head and eye movements are almost not present. The only traits that are more present in his speech than in the others' are facial displays to convey a positive attitude through smiles and laughs. As for other lexical features, he makes a remarkably higher use of overlaps, 0.43 per turn (vs. 0.13 for Renzi and 0.34 for Salvini), probably because Movimento Cinque Stelle was openly critical of journalists, and Di Maio tends to overlap the interviewer in the discussion. The overall impression is that Di Maio has a good control over the conversation and does not let emotions interfere much with the flow of the debate. Also when he smiles or laughs, his body and eyes do not move much and are not used to emphasize a message.

This kind of control is even more evident in Matteo Salvini's interview. The only non-verbal devices he uses to convince the audience are smiles and hand movements, especially complex hand trajectories. The gaze, the eyes and the eyebrows do not move at all. As regards lexical and semi-lexical traits, he uses repetitions slightly more frequently than Renzi (0.22 per turn on average) and only few truncations (0.09 per turn). The overall impression he gives is that of a cold-blooded person who is in control of the situation, whose persuasion strategy relies on his seriousness, paired with the worried attitude for the future of the country that he expresses throughout his arguments.

For the records, Luigi di Maio and Matteo Salvini won the following elections and became the Minister of Economic Development and the Minister of the Interior respectively.

	Matteo Renzi		Luigi Di Maio		Matteo Salvini	
	count	duration	count	duration	count	duration
Face						
laugh	9	51.2	7	40.56	1	4.04
smile	32	163.96	13	185.20	7	36.20
scowl	2	43.96	0	-	0	-
Eyebrown movement						
frown	30	120.8	4	53.20	0	-
raise	20	126.08	0	-	0	-
Eye movement						
close-both	4	7.76	0	-	0	-
close-repeated	10	56.6	2	61.56	0	-
Gaze direction						
up	3	3.36	0	-	0	-
sideways	2	7.52	0	-	0	-
towards-interlocutor	4	47.92	0	-	0	-
down	6	11.48	0	-	0	-
Mouth openness						
open	2	2.96	0	-	0	-
Head movement						
down-repeated	3	6.56	0	-	1	3.12
Handedness						
single	4	9.20	4	109.20	1	0.72
both	17	83.32	4	82.92	0	-
Hand movement trajectory						
complex	42	672.52	8	226.32	20	989.72
up	5	13.80	0	-	4	22.96
sideways	13	107.56	5	103.28	2	4.12
down	3	11.56	1	4.52	0	-
Body posture						
sideways	2	46.6	0	-	0	-
down	1	0,4	0	-	0	-

Table 3: Statistics on annotated information comparing number of occurrences and average duration in milliseconds.

6. Conclusions

In this work we present the extension of the PoliModal corpus to include an additional annotation layer for gesture, posture and facial displays. After introducing the corpus and the annotation scheme, we provide some analyses related to three interviews that have been fully annotated, involving three politicians having a different political orientation: Matteo Renzi, Matteo Salvini and Luigi Di Maio.

Although the corpus annotated so far does not allow for generalisations, we can already observe how the three politicians adopt different communication strategies, with Renzi being more emotional and showing more multimodal traits, while the other two are colder and Salvini tends to express his thoughts exclusively through lexical and semi-lexical traits. This preliminary analysis shows the potential of putting different modalities in relation, as a means to have a wider prespective on political discourse and persuasive strategies.

In the future, we plan to extend the corpus by adding gesture, facial and posture annotation to more interviews. The transcripts and the lexical and semi-lexical annotations in the PoliModal corpus include 56 face-to-face interviews. It would be interesting to have at least three politicians for each party, so as to perform some analyses at party and not at politicians' level.

7. References

Alfano, I., Cutugno, F., Rosa, A. D., Iacobini, C., Savy, R., and Voghera, M. (2014). Volip: a corpus of spoken italian and a virtuous example of reuse of linguistic resources. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).

Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C., and Paggio, P. (2007). The mumin coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation*, 41(3-4):273–287.

Allwood, J. (2001). Dialog coding-function and grammar:

Göteborg coding schemsas. *rapport nr.: Gothenburg Papers in Theoretical Linguistics 85*.

Allwood, J. (2008). Multimodal Corpora. In Lüdeling, et al., editors, *Corpus Linguistics. An International Handbook*, pages 207–225. Mouton de Gruyter.

Bartolini, R., Quochi, V., Felice, I. D., Russo, I., and Monachini, M. (2014). From synsets to videos: Enriching italwordnet multimodally. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).

Bigi, B., Portès, C., Steuckardt, A., and Tellier, M. (2011). Multimodal annotations and categorization for political debates. In *ICMI Workshop on Multimodal Corpora for Machine learning*, pages 1–4.

Bolasco, S., de'Paratesi, N. G., and Giuliano, L. (2006). *Parole in libertà: un'analisi statistica e linguistica dei discorsi di Berlusconi*. Manifestolibri.

Bongelli, R., Riccioni, I., and Zuczkowski, A. (2010). Certain-uncertain, true-false, good-evil in Italian political speeches. In *International Workshop on Political Speech*, pages 164–180. Springer.

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

Catellani, P., Bertolotti, M., and Covelli, V. (2010). Counterfactual communication in politics: Features and effects on voters. In *International Workshop on Political Speech*, pages 75–85. Springer.

Cedroni, L. (2010). Politolinguistics. Towards a New Analysis of Political Discourse. In *International Workshop on Political Speech*, pages 220–232. Springer.

D'Errico, F., Poggi, I., and Vincze, L. (2010). Discrediting body. A multimodal strategy to spoil the other's image. In *International Workshop on Political Speech*, pages 181–206. Springer.

Dinarelli, M., Quarteroni, S., Tonelli, S., Moschitti, A., and Riccardi, G. (2009). Annotating spoken dialogs: From speech segments to dialog acts and frame semantics. In *Proceedings of SRSL 2009, the 2nd Workshop on Semantic Representation of Spoken Language*, pages 34–41, Athens, Greece, March. Association for Computational Linguistics.

Duncan, S. (2004). Coding manual. `http://mcneilllab.uchicago.edu/pdfs/Coding_Manual.pdf`. Accessed: 2020-03-05.

Esposito, F., Basile, P., Cutugno, F., and Venuti, M. (2015). The CompWHoB Corpus: Computational construction, annotation and linguistic analysis of the white house press briefings corpus. *Proceedings of CLiC-it*.

Guerini, M., Giampiccolo, D., Moretti, G., Sprugnoli, R., and Strapparava, C. (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.

Itakura, H. (2001). Describing conversational dominance. *Journal of Pragmatics*, 33(12):1859–1880.

Kipp, M. (2001). Anvil-a generic annotation tool for multimodal dialogue. In *Seventh European Conference on Speech Communication and Technology*.

Koutsombogera, M. and Papageorgiou, H. (2010). Multimodal indicators of persuasion in political interviews. In *International Workshop on Political Speech*, pages 16–29. Springer.

Longobardi, F. (2010). Linguistic factors in political speech. In *International Workshop on Political Speech*, pages 233–244. Springer.

Moneglia, M., Brown, S., Frontini, F., Gagliardi, G., Khan, F., Monachini, M., and Panunzi, A. (2014). The IMAGACT Visual Ontology. an Extendable Multilingual Infrastructure for the Representation of Lexical Encoding of Action. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).

Moretti, G., Sprugnoli, R., Menini, S., and Tonelli, S. (2016). ALCIDE: Extracting and visualising content from large document collections to support Humanities studies. *Knowledge-Based Systems*, 111:100–112.

Navarretta, C. and Paggio, P. (2010). Multimodal behaviour and interlocutor identification in political debates. In *International Workshop on Political Speech*, pages 99–113. Springer.

Poggi, I. (2005). The goals of persuasion. *Pragmatics & Cognition*, 13(2):297–335.

Poggi, I. (2006). Le parole del corpo. introduzione alla comunicazione multimodale.

Sprugnoli, R., Moretti, G., Tonelli, S., and Menini, S. (2016). Fifty years of European history through the Lens of Computational Linguistics: the De Gasperi Project. *IJCOL*, pages 89–100.

Trotta, D., Tonelli, S., Palmero Aprosio, A., and Elia, A. (2019). Annotation and Analysis of the PoliModal Corpus of Political Interviews. In *Proceedings of the Sixth Italian Conference on Computational Linguistics, Bari, Italy, November 13-15, 2019*.

Zurloni, V. and Anolli, L. (2010). Fallacies as argumentative devices in political debates. In *International Workshop on Political Speech*, pages 245–257. Springer.