

LREC-COLING 2024

**ParlaCLARIN IV**  
**Workshop on Creating, Analysing, and**  
**Increasing Accessibility of Parliamentary Corpora**

Proceedings

Editors  
Darja Fišer, Maria Eskevich, David Bordon

May 20, 2024  
Torino, Italia

**Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024): ParlaCLARIN IV Workshop on Creating, Analysing, and Increasing Accessibility of Parliamentary Corpora**

Copyright ELRA Language Resources Association (ELRA), 2024  
These proceedings are licensed under a Creative Commons  
Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-24-1  
ISSN 2951-2093 (COLING); 2522-2686 (LREC)

Jointly organized by the ELRA Language Resources Association  
and the International Committee on Computational Linguistics

# IMPAQTS: A Multimodal Corpus of Parliamentary and Other Political Speeches in Italy (1946-2023), Annotated with Implicit Strategies

Federica Cominetti, Lorenzo Gregori, Edoardo Lombardi Vallauri,  
Alessandro Panunzi

University of L'Aquila, University of Florence, University of Roma Tre, University of Florence  
federica.cominetti@univaq.it, lorenzo.gregori@unifi.it, edoardo.lombardivallauri@uniroma3.it,  
alessandro.panunzi@unifi.it

## Abstract

The paper introduces the IMPAQTS corpus of Italian political discourse, a multimodal corpus of around 2.65 million tokens including 1,500 speeches uttered by 150 prominent politicians spanning from 1946 to 2023. Covering the entire history of the Italian Republic, the collection exhibits a non-homogeneous consistency that progressively increases in quantity towards the present. The corpus is balanced according to textual and socio-linguistic criteria and includes different types of speeches. The sociolinguistic features of the speakers are carefully considered to ensure representation of Republican Italian politicians. For each speaker, the corpus contains 4 parliamentary speeches, 2 rallies, 1 party assembly, and 3 statements (in person or broadcasted). Parliamentary speeches therefore constitute the largest section of the corpus (40% of the total), enabling direct comparison with other types of political speeches. The collection procedure, including details relevant to the transcription protocols, and the processing pipeline are described. The corpus has been pragmatically annotated to include information about the implicitly conveyed questionable contents, paired with their explicit paraphrasis, providing the largest Italian collection of ecologic examples of linguistic implicit strategies. The adopted ontology of linguistic implicitness and the fine-grained annotation scheme are presented in detail.

**Keywords:** political discourse, multimodal corpus, pragmatic annotation, implicit content

## 1. The IMPAQTS corpus

### 1.1. Introduction

Linguistic implicit communication is a powerful means of persuasion, extensively characterizing manipulative discourse: indeed, it is used to convey deceptive content by reducing the receiver's attention to it, leading to its passive acceptance (Lombardi Vallauri, 2016a; Morency et al., 2008). This property makes linguistic implicit communication a potentially dangerous tool when it is used to influence people's choices and behaviors. The IMPAQTS project (Implicit Manipulation in Politics – Quantitatively Assessing the Tendentiousness of Speeches) is focused on this manipulative use of implicit content in political speeches: it aims to build a large multimodal corpus of Italian political discourse and annotate it per implicitly conveyed questionable content.

At the moment of writing, the corpus collection and annotation have been completed but the data processing and the building of a fully searchable web resource are still in progress.

The IMPAQTS corpus includes 1,500 speeches uttered by 150 Italian politicians throughout the history of the Italian Republic (1946-2023), totaling around 2.65 million tokens. Accordingly, the "political discourse" portrayed by the corpus is to be intended in the strict sense of "discourse by politicians", and not in the loose sense of "discourse on

political issues" (Van Dijk et al., 1997). Even in its strict sense, political discourse is a wide text genre, including very different textual and communicative types ranging from interventions in the Houses of Parliament to live recordings on social media. In the IMPAQTS corpus, political speeches have been classified according to channel (in presence vs. broadcast) and addressees (institutions, supporters, general public). Only monologues have been collected, thus focusing on the most typical structure of political discourse, excluding political dialogues and conversations.

### 1.2. Types of Speeches

Six types of political monologue were pinpointed and included in the corpus:

- Parliamentary speech (IMPAQTS\_PARL): speech given in presence, addressing institutions, typically in the Chambers or local councils. It is normally characterized by a formal or solemn register and a very high degree of planning.
- Rally: speech given in presence, addressing an audience mainly of supporters, typically during an election campaign or a public event. Degrees of formality and planning can vary; this variability is linked to the personal style of the speaker and the specific communicative situation, but it also appears highly sensitive

to diachronic variation. The rallies of the so-called "First Republic" (cf. §1.3) tend to be much more formal than the more recent ones. In any case, rallies are usually less formal and less planned than parliamentary interventions.

- Party assembly: speech given in presence, addressing an audience of party colleagues, typically during a party congress. As in the case of the rally, notable interpersonal, intertextual, and intratextual variations can be observed in the register and are further influenced by the diachronic component. Not unlike what was observed for rallies, party assemblies in the First Republic tend to stick to a formal register, while the tone in more recent assemblies can be much more informal.
- Statement in presence: speech given in presence, before an institutional and/or general audience, typically including journalists, as in the case of statements released at a press conference. They may be well-planned speeches or spontaneous declarations; the register can also be more or less formal depending on the situation.
- Broadcast statement: speech delivered for video/audio transmission, intended for the general public, as in the case of messages to the nation from the President of the Republic or the Prime Minister and self-promotional messages broadcast by politicians on television or radio; the register is often medium-formal and the degree of planning tends to be high.
- New media statement: speech recorded and/or broadcast via new media, intended for an audience mainly of followers, such as in Facebook live broadcasts; the register is normally medium, and informal traits are possible; the degree of planning is usually low.

As the descriptions show, the Italian political language represented by the IMPAQTS corpus is not a monolithic entity but portrays instances of monologic speech of medium and even informal register.

To take into account the role of personal style in the linguistic phenomena witnessed by the corpus, 10 speeches for each speaker are included, balanced according to the text-type scheme: for each speaker, the corpus includes 4 parliamentary speeches, 2 rallies, 1 party assembly, and 3 declarations. Considering this, 150 politicians were selected, totaling 1,500 speeches.

Table 1 reports the number of speakers, speeches, tokens, and words per speech type.<sup>1</sup>

<sup>1</sup>These numbers refer to the part of IMPAQTS corpus processed so far, i.e. 1403 speeches out of 1,500 (93.5%).

### 1.3. Diachrony

The IMPAQTS corpus covers the entire history of the Italian Republic from its foundation to the year of resource release. To ease research taking into account the diachronic variable, the corpus has been divided into three sub-sections:

- the speeches delivered between June 25th 1946, the day of establishment of the republican institutions, and May 24th 1972, the closing day of the fifth legislature;
- the period from May 25th 1972 to April 14th 1994, corresponding to legislatures VI to XI;
- from April 15th 1994 to the spring of 2023, representing the legislatures from XII to XIX, still in progress at the time of project closure.

The first breaking point was set to account for the change in the themes and tones of the political debate observed in Italy in the early 1970s, in particular with the first bill for the regulation of abortion. The second breaking point corresponds to the transition from the proportional to the majoritarian electoral system, which marks in Italy a crucial turning point defined as the beginning of the so-called Second Republic. The consistency of the 3 sub-sections is not homogeneous but progressive towards contemporaneity, as shown in Table 2.

This responds to different needs. Firstly, the availability of audio-video resources (and even mere transcripts of speeches) falls dramatically the further we move away from the present. The limited availability of texts becomes even more significant if we consider the balancing between different types of political discourse described in Table 1. Secondly, the greater emphasis given to contemporaneity responds to one of the aims of the IMPAQTS project, namely the dissemination of the themes of linguistic implicitness and the education towards it. Such endeavor was reckoned to be more effective if applied to recent political texts, produced in cultural contexts better known by citizens and more impacting on their lives.

### 1.4. Political Orientation

For each period, well-known figures were favored, and the selection was also respectful of the composition of the parliamentary assemblies in the period considered in terms of gender and political affiliation. As a consequence of Italian political history, women are not represented in the corpus until the 60s. The whole corpus includes 340 speeches by women (23%, corresponding to 34 speakers) and 1,160 speeches by men (77%, 116 speakers). The average age of the speakers is 56. The youngest age is 27, while the oldest is 89.

Speech Type	Speakers	Speeches	Token	Words
Parliamentary speech	150	561 (39.99%)	1,015,495	889,769 (43,11%)
Rally	150	283 (20.17%)	557,902	480,983 (23,30%)
Part assembly	137	137 (9.76%)	264,920	229,379 (11,11%)
Statement in person	133	231 (16.46%)	345,558	299,404 (14,51%)
Broadcast statement	108	164 (11.69%)	145,286	126,427 (6,13%)
New media statement	24	27 (1.92%)	44,429	37,971 (1,84%)
<b>Total</b>	<b>150</b>	<b>1403</b>	<b>2,373,590</b>	<b>2,063,933</b>

Table 1: IMPAQTS numbers per speech type (data derived from 93.5% of total corpus)

Period	Speakers	Speeches
1946-1972	25	88 (6.27%)
1972-1994	57	327 (23.31%)
1994-2023	124	988 (70.42%)

Table 2: Consistency of the diachronic sub-section (data derived from 93.5% of total corpus)

Orientation	Speakers	Speeches
Independent	23	153 (10.91%)
Left	28	199 (14.18%)
Center-Left	50	385 (27.44%)
Center	34	236 (16.82%)
Center-Right	38	294 (20.96%)
Right	20	136 (9.69%)

Table 3: Speeches by political orientation (data derived from 93.5% of total corpus)

The political affiliation was expressed with reference to the party to which the speaker belonged at the time of utterance. Due to the remarkable fragmentation of the Italian political history, no less than 65 different parties were included in the metadata. To ease research, the additional metadatum "political orientation" was added, including six possible values: left, center-left, center, center-right, right, independent. The distribution of the speeches according to this variable is reported in Table 3.

### 1.5. Multimodality

The IMPAQTS corpus was conceived as, and mainly is, a multimodal corpus. However, the ambition to cover the entire history of the Italian Republic in diachrony made it necessary to include in the corpus some speeches for which no video nor audio recording is available. Specifically, this is the case with 63 speeches, whose transcripts were found only in parliamentary stenographs or in printed publications. Numerous speeches – around 600 – are available only in audio format, which means that over 800 speeches are available in video format. Recordings were sourced from different archives, including the Chambers' web TVs and parties and politicians' YouTube channels.

An invaluable source for old speeches was Radio Radicale's archive, a very large collection including not only parliamentary recordings starting from 1976 (while the Chambers' web TVs are available only from the XIV legislature, i.e., from 2001) but also a very large collection of rallies, party assemblies, broadcast messages, and press conferences, some of which dating back to the 60s.

## 2. Criteria for the Annotation of Implicitness

### 2.1. General Aims and Motivation

The IMPAQTS corpus is entirely annotated with information about the implicitly conveyed questionable contents. The collection of a large catalog of spontaneous, ecologic examples of linguistic implicit strategies in Italian is indeed one of the main aims of the project. Political discourse is a text genre particularly suitable for the collection of linguistic implicit strategies. Theoretical and experimental accounts have shown that implicit strategies have strong persuasive power, being able to reduce the critical vigilance that addressees use, as compared to when they are aware of being the target of persuasion attempts (typically, explicit). Accordingly, linguistic implicit strategies are extensively used in text genres characterized by persuasive aims, of which political discourse is a typical representative (Van Dijk, 1992, 1997; Van Dijk et al., 2000; Van Dijk, 2011; Sbisà et al., 1999; Chilton, 2005; Danler, 2005; Rocci, 2002; Charaudeau, 2005; Reisigl, 2008; Lombardi Vallauri et al., 2020; Cominetti et al., 2022, 2023).

### 2.2. Implicit Strategies

The model adopted for the annotation of the implicit strategies includes four main categories:

- presupposition;
- implicature;
- vagueness;
- topicalization.

In the following, the annotated categories are presented with examples extracted from the parliamentary section of the IMPAQTS corpus.

**Presupposition** The presupposition is an implicit strategy included in practically all analyses and taxonomies on implicitness (Bertuccelli Papi, 2009).

To presuppose something is to take it for granted, or at least to act as if one takes it for granted, as background information - as common ground among the participants in the conversation (Stalnaker, 2002)

- (1) Il rapporto tra individuo e Stato con un rafforzamento degli elementi di dialogo e di consulenza preventiva per i cittadini, con una sottolineatura del principio di irretroattività delle norme di sfavore, quindi davvero elementi di fisco amico e di *uno Stato che deve smettere non solo di essere ma anche di apparire sleale e nemico rispetto al cittadino contribuente.*

*The relationship between the individual and the State with a strengthening of the elements of dialogue and preventive consultancy for citizens, with an underlining of the principle of non-retroactivity of unfavorable regulations, therefore true elements of friendly taxation and of a State that must stop not only being but also appearing disloyal and hostile towards the tax-paying citizen.* [DCAP13-A1]

In (1), the change of state predicate *smettere* ("to stop") presupposes that the State is currently being disloyal and hostile towards the tax-paying citizen.

Change of state predicates (Sellars, 1954; Karttunen, 1973) is only one of many presupposition triggers pinpointed in the literature, including factive predicates (Kiparsky and Kiparsky, 1971; Karttunen, 1971), verbs of judgment (Fillmore, 1969), iteratives (Levinson, 1983), some adverbial clauses (Frege, 1892; Lombardi Vallauri, 2000, 2009), definite descriptions (Frege, 1892), etc.

**Implicature** Implicatures are the second cornerstone of linguistic implicitness, famously defined by Grice (1975) as propositions that can be communicated through an utterance without being explicitly said, as in (2).

- (2) Dobbiamo uscire da questa crisi e dobbiamo uscirne più forti come italiani. Tutti sapete che la corsa non di un governo, ma di una lunga fase politica, durata quindici anni, è finita. *Lo dicono quei sondaggi che un tempo venivano tanto citati e oggi tanto nascosti.*

*We must emerge from this crisis and we must emerge stronger as Italians. You all know that the race not of a government, but of a long political phase, which lasted fifteen years, is over. Those polls that were once so often cited and today are so hidden say so.* [WVEL11-A1]

In (2), the speaker – a member of the opposition – is implying that the majority is aware of its loss of consensus and is deliberately hiding polls to conceal it. This is an example of conversational implicature, a type of implicit content arising as a consequence of the obedience in discourse to the four maxims of conversation (the Gricean Maxims, Grice 1975), which jointly express a general cooperative principle. In the specific case, the utterance in itself would violate the maxim of quantity, by apparently giving insufficient information about why polls today are hidden. The maxim is only respected if the mentioned implicature is added to the explicit content of the message.

The literature pinpoints two other types of implicature: conventional and generalized implicatures. The former arise from the use of certain expressions (often connectives and adverbs) to which they are conventionally associated. Generalized implicatures are conversational implicatures that tend to apply frequently in the same way, also in different contexts.

**Vagueness** Vagueness is an implicit strategy contiguous to implicatures, in that it also leaves the completion of the explicitly expressed content to addressees. More specifically, persuasive vagueness is based on the deliberate omission of a relevant detail to assure an advantage to the source (Lombardi Vallauri, 2016a,b, 2019). Typically, speakers resort to vagueness when they want to charge rivals with (often exaggerated) accusations, or when they are making (often exaggerated) promises. Vagueness can be obtained through semantic means, as in (3), or syntactic means.

- (3) Qua c'è *gente* che chiacchiera di mafia ma poi se la dà a gambe quando si deve intervenire con durezza contro la mafia, eh. Qua l'antimafia dei chiacchieroni.  
*Here there are people who chat about mafia but then run for the hills when it is time to intervene with rigidity against mafia, huh. Here, the big mouths' antimafia.* [MSAL20-A1]

In (3), the collective noun "gente" (*people*) is used to avoid explicitly mentioning the actual people who are supposedly responsible for the mentioned behavior.

**Topicalization** Finally, topicalization is a category of implicitness based on the prosodic and/or syntactic framing of some content as a topic information unit. The topic is defined in opposition to the comment, the part of the utterance that realizes the informative purpose of the utterance and conveys the utterance's illocutionary force. Not differently from presuppositions, topics tend to receive shallower processing, because they tend to encode information already active in the short-term memory of the addressees (Lombardi Vallauri, 2009; Lombardi Vallauri and Masia, 2014). Accordingly, they can be considered an implicit strategy. Specifically (and not differently from presupposition), what topicalization leaves implicit is the epistemic responsibility of the source for introducing its content. An example of a tendentious topic is presented in (4).

- (4) Abbiamo compiuto un gesto vero, immaginando sensatamente di confrontarci con interlocutori veri. *Poiché siamo condotti a constatare che le cose non stanno così e che non ci si vuole più paragonare su una misura di verità*, non possiamo avere più dubbi sulla inesorabile esigenza di un gesto reciso.

*We made a real gesture, sensibly imagining that we were dealing with real interlocutors.* Since we are led to realize that things are not like this and that we no longer want to compare ourselves on a measure of truth, *we can no longer have doubts about the inexorable need for a decisive gesture.* [MMAR87-A1]<sup>2</sup>

In the IMPAQTS corpus, only potentially manipulative contents are annotated. In fact, linguistic implicitness is not per se a dishonest linguistic device. On the contrary, it may be a legitimate strategy allowing for conciseness and politeness. For example, it is licit on the part of a source to presuppose that the Italian Republic exists: on the contrary, it would be uneconomical to state it explicitly. The criterion adopted to distinguish potentially manipulative from legitimate implicitness relies on the concept of *bona fide* true information, which applies to contents that any speaker can legitimately think to be shared by any other. Accordingly, the mentioned implicit strategies are annotated only when they convey non-*bona fide* true contents.

The table 4 presents the full set of pragmatic annotation classes.

<sup>2</sup>The typical intonation of topic in Italian can be appreciated in the corresponding audio fragment: <https://www.radioradicale.it/scheda/22225?p=2&s=1528&t=1550&f=2>.

## 2.3. Communicative Functions

Following the model proposed by Brocca et al. (2016), and Garassino et al. (2022), any implicitly conveyed questionable content is reckoned to perform some communicative function. In particular, five possible functions are identified:

- Stance-taking: conveying one's position or stand on a particular issue (Evans, 2016);
- Attack: a blast of unfavorable characteristics or flaws of a political opponent or group (Lee and Xu, 2018);
- Self-praise: a positive content about oneself or one's own (or one's allies') policy (Dayter, 2014);
- Praise to others: a positive content about other people's ideas, intentions, or deeds (Garassino et al., 2019);
- Defence: conveying one's righteousness and non-guilt (Cominetti et al., 2022).

Accordingly, implicit strategies in the corpus are tagged for the communicative function(s) they perform. For example, the conversational implicature described in (2) functions as an attack towards the majority.

## 3. Building and Annotating the IMPAQTS Corpus

### 3.1. Processing Pipeline

Even if the core part of the corpus collection and annotation are made manually by experts, a set of computational linguistics tools is used during the corpus creation process. Each video or audio source passes through the following steps:

1. Transcription of the speech source
2. Time-alignment of the transcription to the source
3. Cooperative pragmatic annotation and curation
4. Export of the XML file with annotation.

### 3.2. Transcription and Alignment

The spoken datum is the obvious starting point of a spoken corpus. Nonetheless, for the large part of the IMPAQTS corpus consisting of parliamentary speeches (IMPAQTS\_PARL), obtaining transcripts was eased by the availability of the stenographic reports of all parliamentary sessions. For the other

<b>Implicatures (IMPL)</b>
Conventional implicature
Generalized implicature
Conversational (particularized) implicature
Conversational implicature by metaphor
Conversational implicature by list
<b>Presuppositions (PPP)</b>
Pragmatic presupposition
Semantic presupposition by definite description
Sem. pres. by restrictive relative clause
Sem. pres. by anaphoric indefinite description
Sem. pres. by adverbial subordinate clause
Sem. pres. by second term of comparison
Sem. pres. by change of state predicate
Sem. pres. by factive predicate
Sem. pres. by adverb
Sem. pres. by adjective
Sem. pres. by wh- question
Sem. pres. by alternative question
Sem. pres. by counterfactual construct
<b>Vagueness (VAG)</b>
Syntactically triggered vagueness
Semantically triggered vagueness
Vagueness triggered by metaphor
<b>Topicalization (TOP)</b>
Syntactically triggered topicalization
Prosodically triggered topicalization

Table 4: Types of implicit annotation in IMPAQTS corpus.

types of text, the speeches were automatically transcribed through the Google Speech-to-Text tool.<sup>3</sup> Both types of transcripts – stenographic reports and automatic transcripts – were then reviewed by at least two members of the IMPAQTS team to eliminate errors and deliberate interventions by stenographers.

Two versions of the written section of the corpus will be released: in the first one, orthographic punctuation is inserted to ease readability (see below); in the other one, prosodic breaks are inserted according to the Lablita/C-ORAL-ROM conventions (Cresti and Moneglia, 2005).

Transcribed texts are automatically aligned to their audio through Aeneas, an open-source tool<sup>4</sup> that performs forced alignment.

### 3.3. Protocol for Implicit Annotation

Pragmatic annotation is a task highly influenced by personal sensitivity and encyclopedic knowledge. In the IMPAQTS project, the protocol includes the study of the relevant literature, an ad hoc vademecum, and training by the project manager.

cum, and training by the project manager.

In the pragmatic annotation, not only are the strings of text marked with the tags corresponding to the implicit strategy and its pragmatic function, but an explicit version of the implicit content is made available (a procedure whose importance was highlighted by Sbisà 2021). The IMPAQTS protocol leads to extremely comprehensive explicitation, avoiding anaphorics and deictics to untie the implicit content completely.

Each speech is annotated by three independent annotators, one of which subsequently adopts the role of curator, comparing the three annotated versions and validating the definitive one.

To this aim, a WebAnno-MM<sup>5</sup> instance has been set up on a local server. WebAnno-MM is the multimodal version of the WebAnno<sup>6</sup> cooperative annotation tool: in addition to providing an online user-friendly annotation environment, it allows playing the video/audio segments during annotation. Submission of the transcription into HIAT-TEI format (Rehbein et al., 2004) is necessary to upload text and video for multimodal annotation. Annotation analysis and curation are also performed through the WebAnno-MM platform. At the end of this process, annotated files are exported to XMI,<sup>7</sup> tagged with parts of speech and lemmas with TreeTagger,<sup>8</sup> and converted to VRT to be further inserted in the search engine platform. After the annotation is finished, all the VRT files will be indexed and, together with the corresponding video or audio source, loaded into EMMAcorp (Cominetti et al., 2022) for linguistic searches.

Although the inter-annotator agreement has not been evaluated yet, a few main issues can be mentioned. Curators noticed that less expert annotators tend to go through a phase of "hyper-annotation", in particular when wrongly tagging as implicatures merely re-elaborated content and logical implications, or on the contrary full deductions. Implicit strategies with clear linguistic triggers (including some kinds of presupposition and vagueness and conventional implicatures) tend to show larger agreement, even if hyper-annotation may still be present due to the sometimes uncertain recognition of bona fide true content. The most difficult implicit strategy to manage seems to be topicalization, especially when only activated by prosodic cues.

The whole corpus with implicit annotation is stored in XML format. Figure 1 shows the annotation of the implicature of example 2 in section 2.2. The implicature is annotated with the tag <impl>.

<sup>3</sup><https://cloud.google.com/speech-to-text/>

<sup>4</sup><https://github.com/readbeyond/aeneas>

<sup>5</sup><https://github.com/webanno/webanno-mm>

<sup>6</sup><https://webanno.github.io/webanno/>

<sup>7</sup><https://www.omg.org/spec/XMI/>

<sup>8</sup><https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

Sp. Type	Words	Implicits	/100Kw
Parliam.	887,965	19,538	2,200
Rally	479,053	11,602	2,422
Party ass.	229,379	4,583	1,998
Statements	462,277	8,053	1,742
<i>Total</i>	<i>2,058,674</i>	<i>43,776</i>	<i>2,126</i>

Table 5: Number of implicits per speech type.

Sp. Type	IMPL	PPP	VAG	TOP
Parliam.	845	873	419	221
Rally	895	868	557	191
Party ass.	616	805	515	169
Statements	609	717	386	209

Table 6: Relative frequency of implicit strategies per speech type (number of implicits per 100Kw).

along with its classification (type), its communicative function (function), and an explanation of the implicit content (comment).

### 3.4. Preliminary Results on Implicit Strategies and Types of Speech

Table 5 shows the number of words and implicit per speech type, along with the relative frequency of implicits, estimated per 100,000 words (last column). Table 6 reports, for each speech type, the relative frequency of the different implicit strategies: implicature (IMPL), presupposition (PPP), vagueness (VAG), and topicalization (TOP). All these numbers refer to 93.5% of the whole IMPAQTS corpus.

As Tables 5 and 6 show, IMPAQTS\_PARL is above the average political discourse for global implicitness. This is due to a relatively high presence of the two most common implicit categories: implicatures (a trait shared with rallies) and presuppositions (a trait shared with rallies and party assemblies). If compared with the single most implicit political genre, rallies, parliamentary speeches prove to be significantly less vague but higher in topicalizations. On one side, this may be linked to the tendency of rallies to include many promises (a linguistic act often tending to vagueness). On the other, parliamentary speeches are the most carefully planned type of political speech, and accordingly often show elaborate syntax, in which subordinates and other circumstantial phrases may be framed as topics.

Certainly, this is merely a preliminary outline of an analysis of such data, and further elaboration would be necessary for comprehensive development.

## 4. Further Research

Subsections of the IMPAQTS corpus and its pragmatic annotation have already been used for the de-

scription of pragmatic phenomena, including the in-depth analysis of under-described linguistic implicit triggers (Lombardi Vallauri et al., 2021; Cominetti and Giunta, 2022), and the interaction of linguistic implicitness and different aspects of grammar (Cominetti, 2023; Cimmino and Cominetti, 2023). The large collection of texts has allowed us to extend to political discourse a kind of study that was limited so far to other text types. Once made available to the scientific community, the large size of the corpus, its diachronic and multimodal nature, and the unprecedented pragmatic annotation will certainly be useful for an array of research in all the fields of linguistics.

## 5. Acknowledgements

The IMPAQTS project was funded by the Italian Ministry of University and Research within the Italian national research program PRIN "IMPAQTS: Implicit Manipulation in Politics – Quantitatively Assessing the Tendentiousness of Speeches" (2020-2023, PRIN code: 2017STJCE9), coordinated by Edoardo Lombardi Vallauri from Università Roma Tre, and involving Università di Firenze as computational linguistics unit.

We are grateful to Radio Radicale<sup>9</sup> for allowing the offline processing of many of their audio-video documents.

## 6. References

- Marcella Bertuccelli Papi. 2009. Implicitness. *Key notions in pragmatics*, pages 139–162.
- Nicola Brocca, Davide Garassino, and Viviana Masia. 2016. Politici nella rete o nella rete dei politici? l'implicito nella comunicazione politica italiana su twitter. *PhiN-Beiheft*, 11(2016):66.
- Patrick Charaudeau. 2005. *Le discours politique: les masques du pouvoir*. Vuibert.
- Paul A Chilton. 2005. Manipulation, memes and metaphors. *Manipulation and ideologies in the twentieth century*, pages 15–43.
- Doriana Cimmino and Federica Cominetti. 2023. Italian davvero ('really') as a trigger of implicit contents in persuasive discourse. *Journal of Pragmatics*, 211:84–95.
- Federica Cominetti. 2023. Nominalization as an enhancer of linguistic implicitness in political discourse. *Lingue e Linguaggi*, 56:69–88.

<sup>9</sup><https://www.radioradicale.it/>

```

<doc id='WVEL11-A1' parlante='Walter Veltroni' [...]>
[...]
Dobbiamo uscire da questa crisi e dobbiamo uscirne più forti come italiani.
Tutti sapete che la corsa non di un governo, ma di una lunga fase politica,
durata quindici anni, è finita.
<impl type='cvrs' comment='Implica che il governo [...] ' function='TT'>
Lo dicono quei sondaggi che un tempo venivano tanto citati e oggi tanto
nascosti.
</impl>
[...]
</doc>

```

Figure 1: XML annotation of the implicit phenomenon reported in the example (2) (implicature).

- Federica Cominetti, Doriana Cimmino, Claudia Coppola, Giorgia Mannaioli, and Viviana Masia. 2023. Manipulative impact of implicit communication: A comparative analysis of French, Italian and German political speeches. *Linguistik online*, 120(2):41–64.
- Federica Cominetti and Giulia Giunta. 2022. Change of state and factive nominals and nominalizations as presupposition triggers. *Italian Journal of Linguistics*, 34:59–102.
- Federica Cominetti, Lorenzo Gregori, Edoardo Lombardi Vallauri, Alessandro Panunzi, et al. 2022. IMPAQTS: un corpus di discorsi politici italiani annotato per gli impliciti linguistici. In *Corpora e Studi linguistici. Atti del LIV Congresso della Società di Linguistica Italiana (Online, 8–10 settembre 2021), a cura di Emanuela Cresti e Massimo Moneglia. Milano, Officinaventuno*, pages 151–164.
- Emanuela Cresti and Massimo Moneglia. 2005. *C-ORAL-ROM: integrated reference corpora for spoken romance languages*. John Benjamins Publishing.
- Paul Danler. 2005. Morpho-syntactic and textual realizations as deliberate pragmatic argumentative linguistic tools. *Manipulation and ideologies in the twentieth century*, pages 45–60.
- Daria Dayter. 2014. Self-praise in microblogging. *Journal of Pragmatics*, 61:91–102.
- Ash Evans. 2016. Stance and identity in Twitter hashtags. *Language@internet*, 13(1).
- Charles J Fillmore. 1969. Verbs of judging: An exercise in semantic description. *Research on Language & Social Interaction*, 1(1):91–117.
- Gottlob Frege. 1892. Über sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100(1):25–50.
- Davide Garassino, Nicola Brocca, and Viviana Masia. 2022. Is implicit communication quantifiable? a corpus-based analysis of British and Italian political tweets. *Journal of Pragmatics*, 194:9–22.
- Davide Garassino, Viviana Masia, and Nicola Brocca. 2019. Tweet as you speak: the role of implicit strategies and pragmatic functions in political communication: Data from a diamesic comparison. *RILA: Rassegna Italiana di Linguistica Applicata: 2/3*, pages 187–208.
- Herbert P Grice. 1975. Logic and conversation. In *Speech acts*, pages 41–58. Brill.
- Lauri Karttunen. 1971. Some observations on factivity. *Research on Language & Social Interaction*, 4(1):55–69.
- Lauri Karttunen. 1973. Presuppositions of compound sentences. *Linguistic inquiry*, 4(2):169–193.
- Paul Kiparsky and Carol Kiparsky. 1971. Fact'in semantics dd steinberg and la jakobovits, eds. *Semantics (1 971)*, pages 345–69.
- Jayeon Lee and Weiai Xu. 2018. The more attacks, the more retweets: Trump's and clinton's agenda setting on Twitter. *Public Relations Review*, 44(2):201–213.
- Stephen C Levinson. 1983. *Pragmatics*. Cambridge university press.
- Edoardo Lombardi Vallauri. 2000. *Grammatica funzionale delle avverbiali italiane*. Carocci.
- Edoardo Lombardi Vallauri. 2009. La struttura informativa: forma e funzione negli enunciati linguistici. (*No Title*).
- Edoardo Lombardi Vallauri. 2016a. Implicits as evolved persuaders. *Pragmemes and Theories of Language Use*, pages 725–748.
- Edoardo Lombardi Vallauri. 2016b. The “exaptation” of linguistic implicit strategies. *SpringerPlus*, 5(1):1106.

- Edoardo Lombardi Vallauri. 2019. La lingua disonestà: Contenuti impliciti e strategie di persuasione, il mulino.
- Edoardo Lombardi Vallauri, Laura Baranzini, Doriana Cimmino, Federica Cominetti, Claudia Coppola, and Giorgia Mannaioli. 2020. Implicit argumentation and persuasion: A measuring model. *Journal of Argumentation in Context*, 9(1):95–123.
- Edoardo Lombardi Vallauri, Federica Cominetti, and Laura Baranzini. 2021. Presupposing indefinite descriptions. *Journal of Pragmatics*, 180:173–186.
- Edoardo Lombardi Vallauri and Viviana Masia. 2014. Implicitness impact: measuring texts. *Journal of Pragmatics*, 61:161–184.
- Patrick Morency, Steve Oswald, and Louis De Saussure. 2008. Explicitness, implicitness and commitment attribution: A cognitive pragmatic approach. *Belgian journal of linguistics*, 22(1):197–219.
- Jochen Rehbein, Thomas Schmidt, Bernd Meyer, Franziska Watzke, and Annette Herkenrath. 2004. Handbuch für das computergestützte transkribieren nach hiat. *Working papers in multilingualism*, 56.
- Martin Reisigl. 2008. 11. rhetoric of political speeches. *Handbook of communication in the public sphere*, 4:243.
- Andrea Rocci. 2002. Are manipulative texts coherent? In *New perspectives on manipulation and ideologies: theoretical aspects*, Amsterdam: John Benjamins (selected papers from the conference “Manipulation in the totalitarian ideologies of the twentieth century”, Monte Verità/Ascona).
- Marina Sbisà. 2021. Presupposition and implicature: Varieties of implicit meaning in explicitation practices. *Journal of Pragmatics*, 182:176–188.
- Marina Sbisà et al. 1999. Ideology and the persuasive use of presupposition. In *Language and ideology. Selected papers from the 6th International Pragmatics Conference*, volume 1, pages 492–509. International Pragmatics Association Antwerp.
- Wilfrid Sellars. 1954. Presupposing. *The Philosophical Review*, 63(2):197–215.
- Robert Stalnaker. 2002. Common ground. *Linguistics and philosophy*, 25(5/6):701–721.
- Teun A Van Dijk. 1992. Discourse and the denial of racism. *Discourse & society*, 3(1):87–118.
- Teun A Van Dijk. 2011. Discourse and ideology. *Discourse studies: A multidisciplinary introduction*, pages 379–407.
- Teun A Van Dijk et al. 1997. What is political discourse analysis. *Belgian journal of linguistics*, 11(1):11–52.
- Teun A Van Dijk et al. 2000. New (s) racism: A discourse analytical approach. *Ethnic minorities and the media*, 37:33–49.
- Teun Adrianus Van Dijk. 1997. *Discourse as social interaction*, volume 2. Sage.