



UNIVERSITÀ
DI PAVIA

UNIVERSITÀ DEGLI STUDI DI PAVIA

DIPARTIMENTO DI STUDI UMANISTICI
CORSO DI LAUREA MAGISTRALE IN LINGUISTICA TEORICA, APPLICATA E
DELLE LINGUE MODERNE

VELOCITÀ DI ELOQUIO E CONTENUTI IMPLICITI
NEL DISCORSO POLITICO ITALIANO CONTEMPORANEO:
UN'INDAGINE SOCIOLINGUISTICA

RELATRICE

Prof.ssa Claudia Roberta Combei

CORRELATRICE

Prof.ssa Chiara Zanchi

Tesi di laurea magistrale di

Claudia Tognini

Matricola n. 516337

Anno accademico 2023/2024

Abstract

La velocità d'eloquio, o *speech rate*, e le caratteristiche temporali del parlato sono oggetto di numerosi studi in fonetica, che spaziano dalla sociolinguistica all'apprendimento della L2, alla linguistica clinica e alla pragmatica, con un' enfasi sul legame tra la velocità del parlato e la sua potenzialità persuasiva: fenomeni di natura pragmatica, come implicature, presupposizioni, topicalizzazioni e istanze di vaghezza, influenzano le potenzialità persuasive di un discorso.

Il parlato politico è ampiamente studiato in linguistica. Nel presente elaborato sono analizzati 20 discorsi di 11 personalità politiche italiane contemporanee, tratti dal *corpus* IMPAQTS, per investigare il legame tra la velocità di eloquio e alcune variabili sociolinguistiche annotate nel *corpus*: si tratta di genere, età, provenienza e orientamento politico. Sfruttando *Whisper*, il sistema di *Automatic Speech Recognition* di OpenAI, le *pipeline* fornite dal servizio WebMaus, il programma Praat e la libreria *praatio* per Python, sono state estratte le metriche relative alle proprietà temporali del parlato di questi discorsi ed è stato creato un modello statistico che le metta in relazione con le variabili sociolinguistiche. Infine, per ogni parlante sono stati isolati tre enunciati contenenti presupposizioni grazie all'annotazione pragmatica fornita dal *corpus*. La velocità di eloquio di queste produzioni è stata comparata con quella di altri enunciati che ne sono invece privi.

Indice

Abstract	2
Indice	3
1. Introduzione	5
2. Velocità d'eloquio.....	9
2.1. Lingue iso-sillabiche e iso-accentuali	9
2.2. Variabili sociolinguistiche	14
2.3. Tendenze cross-linguistiche; contenuti informativi	19
2.4. Applicazioni in linguistica clinica.....	23
2.5. Velocità d'eloquio nell'acquisizione L2.....	25
3. Il discorso politico	29
3.1. Strategie linguistiche per il mantenimento del potere	29
3.2. Approcci multimediali.....	33
3.3. Populismo e nuovi media	36
4. Lo studio.....	39
4.1. Corpora multimediali e il corpus IMPAQTS	39
4.2. Annotazione pragmatica del corpus IMPAQTS	41
4.2.1. Implicature	42
4.2.2. Presupposizioni	44
4.2.3. Topicalizzazioni.....	46
4.2.4. Vaghezza.....	47
4.3. Selezione dei dati utilizzati e trascrizione	49
4.4. Metodi per il calcolo dello speech rate.....	52
4.5. Descrizione statistica dei dati	55
4.6. Modello di regressione lineare	57
4.7. Analisi dei dati: enunciati contenenti impliciti	60
5. Risultati e discussione	62

5.1. Lunghezza e distribuzione delle pause	62
5.2. Velocità di eloquio	64
5.3. Velocità di articolazione	69
5.4. Regressione lineare: speech rate	73
5.5. Regressione lineare: articulation rate	76
5.6. Enunciati contenenti impliciti	78
6. Conclusione e sviluppi futuri	96
Appendice.....	100
Bibliografia.....	104

1. Introduzione

In fonetica sperimentale, si è scritto molto sulla velocità di eloquio in relazione a numerose tematiche, anche molto diverse tra loro. Propriamente, si intende con velocità di eloquio la velocità con la quale viene prodotto un enunciato o una qualsiasi forma di parlato; esistono diversi modi per misurarla, a seconda delle necessità del singolo studio.

Nel secondo capitolo di questa tesi vengono approfondite alcune tematiche che sono state studiate in letteratura in relazione alla velocità di eloquio. Quest'ultima fa parte di una serie di parametri di natura temporale che possono essere utilizzati per descrivere la lingua. Anche per questo motivo, a lungo si è investigato il ruolo che la velocità ricopre nella distinzione tradizionale, e assai dibattuta, tra lingue a isocronia sillabica e lingue a isocronia accentuale. Recenti studi non escludono che la velocità di eloquio possa influenzare le proprietà ritmiche di una lingua, in misura diversa a seconda della lingua in questione (Dellwo & Wagner 2003; Dellwo & Volker 2006).

Lo studio delle proprietà temporali del parlato ha conosciuto anche applicazioni importanti in altri ambiti della linguistica, come indice di acquisizione (insieme ad altri) per l'apprendimento della L2 (Ordin et al. 2011; Ordin & Polyanskaya 2015b; Huang & Gráf 2020), come indizio nell'acquisizione di alcune distinzioni morfologiche nella L1 dei bambini (Lester et al. 2020), ma anche come possibile strumento diagnostico (*biomarker*) per l'identificazione precoce di individui affetti da morbo di Alzheimer e altri disturbi neurologici che causano *mild cognitive impairment* (Gagliardi & Tamburini 2021; Calzà et al. 2021): il rallentamento nella produzione sarebbe una conseguenza della crescente difficoltà nella selezione delle parole da parte dei soggetti affetti da tali patologie.

La velocità di eloquio (e di articolazione) è un parametro di per sé privo di significato indessicale; ogni parlante è in grado di controllare la propria velocità consciamente, almeno in maniera parziale, e i suoi ascoltatori sono in grado di accorgersi di queste differenze. Molti studi hanno ricercato relazioni tra la velocità nel parlato di un individuo e altre variabili di natura sociolinguistica come genere ed età: risultati consistenti supportano l'ipotesi che gli uomini, generalmente, parlino più velocemente delle donne (Pépiot 2014; Yuan et al. 2007), e che la velocità di eloquio cambi durante la vita naturale dell'individuo, andando incontro a una diminuzione con l'avanzare dell'età (Harnsberger et al. 2008).

Queste tendenze portano a risultati interessanti anche sul fronte percettivo: un parlato più veloce è più spesso attribuito dagli ascoltatori a individui giovani e di genere maschile (Harnsberger et al. 2008). A partire da questi risultati, risultati interessanti provengono anche

da studi di natura psicologica e sociologica, che indagano il rapporto tra la velocità con cui un individuo parla e le percezioni di competenza, intimità e attrattività sociale che lui/lei ingenera nel suo pubblico. Uniformemente sono riscontrati giudizi nettamente più positivi nei confronti di parlanti veloci (Miller et al. 1976), soprattutto quando l'ascoltatore percepisce una somiglianza tra la propria velocità usuale e quella a cui viene sottoposto durante l'esperimento (Street & Brady 1982; Feldstein et al. 2001); uniformemente, le voci dei parlanti uomini hanno suscitato valutazioni migliori. Non si tratta comunque di un'associazione di tipo automatico, e altri fenomeni, come quello dell'accomodamento, possono influenzare queste dinamiche comunicative; viene comunque affrontato il tema del legame tra la velocità di eloquio e il potenziale di persuasione del parlato, con studi che hanno riguardato anche specificamente personalità politiche (O'Connell et al. 1989). Al parlato politico è dedicato il terzo capitolo di questo elaborato, nel quale sono approfondite alcune linee di studio recenti, con una particolare attenzione per gli studi che hanno utilizzato dati di natura multimediale e per la tematica della persuasione.

Esistono numerose strategie di persuasione rese possibili dalla natura del linguaggio: ne vengono approfondite alcune appartenenti alla categoria degli impliciti nella sezione metodologica del capitolo quarto. Sono riprese nozioni fondamentali nell'ambito di studi della pragmatica, come implicature, presupposizioni, topicalizzazioni e istanze di vaghezza. Non ogni istanza di questi fenomeni pragmatici costituisce un tentativo di manipolare: molti sono ubiqui nella lingua per motivi legati all'economia degli sforzi e alla deissi. In ogni caso, si tratta di forme comunicative, di vario genere, in grado di trasmettere, offuscandoli, messaggi diversi dal significato letterale di un enunciato, riducendo al contempo l'attenzione prestata dall'ascoltatore nei confronti di questi contenuti; oppure, possono ridurre l'impegno epistemico del parlante per questi messaggi, diminuendo la responsabilità percepita nei loro confronti (Lombardi Vallauri 2019). Si è scelto di indagare il ruolo della velocità di eloquio in presenza di queste strategie linguistiche, nell'ambito del parlato politico contemporaneo in Italia.

Il legame tra lingua e politica è infatti noto fin dall'antichità, come testimoniato dall'interesse dimostrato nei confronti della retorica da parte dei maggiori oratori classici e dall'esistenza di numerosi trattati composti con l'obiettivo di formare figure professionali che fossero esperte nell'arte del parlare in pubblico. Numerosi studi di natura linguistica hanno analizzato il parlato politico, partendo però spesso da trascrizioni di discorsi o, in ambito contemporaneo, da contenuti scritti, come post pubblicati da personalità politiche su piattaforme *online* quali X (ex Twitter) o Facebook, che hanno rivoluzionato la comunicazione pubblica da parte delle personalità politiche e non solo. L'avvento dei *corpora* e degli strumenti

computazionali in grado di elaborare e analizzare automaticamente testi, purché in formato *machine readable*, ha reso possibile lo studio di una quantità sempre maggiore di materiali, raccolti in *corpora* specialistici preesistenti o creati *ad hoc* per questo fine (Combei & Reggi 2024).

Tuttavia, pochi studi finora hanno considerato gli aspetti fonetici del parlato politico e le sue caratteristiche segmentali e soprasegmentali, anche per la maggiore difficoltà di creazione e mantenimento di *corpora* definibili multimediali, che raccolgano cioè materiale audio ed eventualmente video contenente l'evento comunicativo in questione.

La metodologia con la quale è stato effettuato questo studio, che in parte ambisce a colmare le lacune appena menzionate, è dettagliata nel quarto capitolo; fondamentale è stato il *corpus* IMPAQTS (*Implicit Manipulation in Politics – Quantitatively Assessing the Tendentiousness of Speeches*), nato dalla collaborazione delle università di Firenze e di Roma Tre, che rientra in un progetto volto allo studio di elementi impliciti, manipolatori e tendenziosi nel parlato politico (Cominetti et al. 2022).

Sono stati selezionati 20 discorsi tra quelli presenti nella *corpus*, sia sotto forma di testo scritto che come collegamenti a materiale video, ospitato dalla piattaforma YouTube. A partire dalle registrazioni audio di questi discorsi e dalle relative trascrizioni si è investigato il ruolo della velocità di eloquio soprattutto in relazione a variabili sociolinguistiche quali genere, età, provenienza geografica, ma anche appartenenza politica. Poco, infatti, è stato scritto sulle caratteristiche temporali dell'italiano, soprattutto in relazione ai diversi italiani regionali; per questo motivo si è cercato un campione che raccogliesse parlanti provenienti dal Nord, dal Centro, e dal Sud del Paese.

I discorsi presenti nel *corpus* sono corredati da un'annotazione anche di natura pragmatica: si è sfruttata questa annotazione, isolando gli enunciati interessati da presupposizioni per indagare l'andamento della velocità di eloquio anche in corrispondenza di questi impliciti.

Per la preparazione dei dati sono stati utilizzati diversi strumenti di linguistica computazionale, come il sistema di riconoscimento vocale (*Automatic Speech Recognition*, o ASR) *whisper.ai*, fornito dall'azienda OpenAI, il software Praat, ampiamente usato in fonetica acustica, e le *pipeline* messe a disposizione *online* gratuitamente dal Dipartimento di Fonetica dell'Università di Monaco di Baviera con il servizio WebMaus.

Il calcolo delle pause, della velocità di eloquio, della velocità di articolazione e di altri parametri utili è stato effettuato attraverso uno *script* nel linguaggio di programmazione Python tratto da Combei & Didoni (2024).

Per l'analisi statistica dei risultati così ottenuti è stato utilizzato il linguaggio di programmazione e l'ambiente di sviluppo R, attraverso l'ambiente di lavoro RStudio, e alcuni *script* che sfruttano librerie quali *tidyverse* (Wickham et al. 2019) e *ggplot2* (Wickham et al. 2016).

La descrizione statistica dei risultati, sia per l'indagine sociolinguistica sia per quella pragmatica degli impliciti, è presentata nel capitolo quinto, insieme ai risultati delle regressioni lineari svolte a partire da essi.

2. Velocità d'eloquio

La velocità di eloquio, o *speech rate*, è una misura della velocità del parlato, esprimibile in diversi modi: è possibile calcolare il numero di foni prodotti per ogni secondo, le sillabe al secondo, le parole prodotte in un minuto. Nel presente elaborato si è scelto di utilizzare le sillabe al secondo come unità di misura. I valori della velocità di eloquio sono quindi stati ottenuti dal rapporto tra le sillabe totali prodotte durante la fonazione e la durata in secondi della fonazione stessa. Il tempo totale di fonazione include quindi pause, esitazioni e altri segmenti in cui la fonazione si arresta, che possono dipendere da diversi fattori: dall'esitazione del parlante a motivi di natura pragmatica, ma anche da fattori esterni, soprattutto quando si tratta di un discorso pronunciato pubblicamente.

Esiste un'altra misura delle proprietà temporali del parlato, la velocità di articolazione, o *articulation rate*: a differenza dello *speech rate*, al tempo totale di fonazione sono sottratti gli intervalli dove la fonazione non avviene. Rende dunque un'idea più precisa della velocità effettiva di fonazione, soprattutto laddove siano presenti numerose pause o esitazioni.

È anche possibile, a partire da queste misure, calcolare la durata media di ciascuna sillaba in secondi (*average syllable duration*).

2.1. Lingue iso-sillabiche e iso-accentuali

Oggetto di studio in letteratura è anche l'interazione tra il parametro della velocità di eloquio e la distinzione (o, meglio, il gradiente) tra lingue a isocronia sillabica (o iso-sillabiche, *syllable timed*) e lingue a isocronia accentuale (o iso-accentuali, *stress timed*). Questa distinzione è strettamente legata al concetto di periodicità: le lingue iso-sillabiche avrebbero come unità minima di segmentazione la sillaba, e si osserverebbe pertanto la tendenza a produrre sillabe della medesima durata; viceversa, le lingue iso-accentuali avrebbero come unità di segmentazione minima l'intervallo tra gli accenti, che rimarrebbe dunque costante.¹

Si tratta di una distinzione tradizionalmente applicata dai linguisti a partire dall'osservazione del comportamento ritmico di alcune lingue europee, che Abercrombie (1967) propone addirittura come universale. Essa è stata arricchita successivamente

¹ È interessante notare come varietà regionali della stessa lingua possano esibire caratteristiche diverse, come avviene nel caso dell'inglese: mentre l'inglese britannico sarebbe una lingua iso-accentuale, l'inglese di Singapore presenterebbe le caratteristiche di una lingua iso-sillabica (Grabe & Low 2002).

dall'aggiunta di una terza categoria, quella delle cosiddette lingue *mora-timed*, come il giapponese, che avrebbero come unità di segmentazione minima le *morae*, unità di livello inferiore alle sillabe (Bloch 1942; Han 1962).

Nonostante la distinzione tra lingue iso-sillabiche e iso-accentuali sia ormai in uso dagli anni Quaranta del XX secolo (Lloyd James 1940; Pike 1945), rimangono difficili da trovare correlati fonetici che forniscano una prova tangibile della sua sostanza, ovvero della presenza di unità della medesima durata alla base dell'organizzazione temporale del parlato nelle diverse lingue. Addirittura, numerosi studi hanno dimostrato che nelle lingue iso-sillabiche la durata delle sillabe non è realmente costante, mentre nelle lingue iso-accentuali l'intervallo tra gli accenti subisce spesso variazioni, e anche la durata delle *morae* in giapponese non sarebbe sempre costante (Nespor et al. 2011).

Se negli studi sopracitati l'interesse era di natura puramente quantitativa, secondo parametri misurabili in laboratorio, in altri lavori si è ricercata la correlazione tra le caratteristiche temporali di una lingua e la sua classificazione ritmica: secondo Dasher & Bolinger (1982), la distinzione tra lingue iso-accentuali e iso-sillabiche non sarebbe essa stessa un primitivo fonetico, bensì il risultato di altre proprietà di natura fondamentale.

Questo concetto venne parzialmente ripreso in uno studio dell'anno successivo (Dauer 1983): le lingue iso-accentuali ammetterebbero un numero maggiore di strutture sillabiche, e dunque le sillabe atone subirebbero un processo di riduzione vocalica², sia a livello di timbro vocalico, con una tendenza alla centralizzazione (con casi estremi nei quali ogni vocale atona sarebbe ridotta alla vocale indistinta), sia a livello di durata vocalica, fino alla sparizione. Per questo motivo, dunque, le sillabe sarebbero tutte ugualmente salienti nelle lingue iso-sillabiche, mentre non ciò non varrebbe per le lingue iso-accentuali (Ramus et al. 1999). Inoltre, le lingue iso-accentuali presenterebbero più spesso un accento lessicale mobile realizzato con una varietà di mezzi prosodici, mentre quelle iso-sillabiche sarebbero tendenzialmente prive di accento

² Si intende con riduzione vocalica una produzione nella quale gli organi fonatori non raggiungono pienamente la posizione *target*, con un procedimento denominato di *target undershoot*, generando dunque valori formantici diversi da quelli considerati ideali. Essa è stata variamente spiegata sia in termini di tendenza all'assimilazione, soprattutto nel caso delle vocali brevi (Lindblom 1963), sia in termini di ipoarticolazione (Van Bergem 1995). In generale, è comunemente accettato il legame con una durata ridotta della produzione. La riduzione vocalica è stata anche attribuita all'automatizzazione delle sequenze di produzione, in una modalità che è tipica del parlato degli adulti e che presenta un certo grado di dipendenza dal contesto: per esempio, i contesti in cui la riduzione è facilitata sono il parlato veloce, lo stile di parlato casuale, le strutture informative non marcate (Kapatsinski et al. 2020); viceversa, quando il parlante vuole enfatizzare un certo segmento, la riduzione tende a non avvenire.

lessicale, come il francese, o realizzerebbero l'accento con altri strumenti, come la variazione del contorno intonativo (Dauer 1983). Si tratterebbe dunque di differenze linguistiche dovute a fatti appartenenti ad altri livelli di analisi, e non alla tendenza ad avere unità che occupano intervalli uniformi di tempo; l'isocronia sarebbe quindi un fenomeno meramente percettivo.

La stessa Dauer (1987) sottolinea comunque come quella tra lingue iso-sillabiche e lingue iso-accentuali non sia una distinzione netta, strettamente categorica, com'era stata inizialmente concepita: non si tratta infatti di categorie discrete, ma di singole proprietà che una lingua può avere o meno, trovandosi quindi più o meno vicina alla definizione di lingua iso-sillabica o iso-accentuale lungo un *continuum*. Questa posizione, supportata dall'esistenza di lingue che presentano caratteristiche ritmiche intermedie (Nespor 1990), è quella comunemente chiamata *phonological account of rhythm*: Dauer (1987) propone una lista di tratti, concepiti come binari, dai quali sarebbe possibile attribuire a ciascuna lingua un punteggio riguardante il suo ritmo al fine di una classificazione.

Questi tratti riguardano diversi aspetti: la durata relativa delle sillabe toniche e atone; le strutture sillabiche ammesse; se il valore distintivo della quantità vocalica, qualora ammesso, sia espresso solo nelle sillabe toniche, o anche in quelle atone; il ruolo delle sillabe toniche nel contorno intonativo; la presenza o assenza di toni con valore distintivo nelle sillabe atone; la presenza o assenza di riduzione vocalica nelle sillabe atone; la presenza o assenza di ipoarticolazione consonantica nelle sillabe atone; il valore distintivo dell'accento e/o la sua possibilità di movimento (Dauer 1987). Molto rimane comunque da scoprire, data l'assenza di indagini approfondite su lingue meno studiate (Ramus et al. 1999); non è esclusa l'esistenza di altri profili ritmici delle lingue di alcune aree del mondo, che non sono ancora stati descritti.

In Ramus et al. (1999) non si ritiene che il *phonological account of rhythm*, come formulato in Dauer (1987), sia sufficiente per spiegare i fenomeni osservati. Esso non spiega in particolare la capacità dei bambini molto piccoli, dimostrata da Nazzi et al. (1998), di distinguere lingue diverse dalla loro L1 quando esse appartengono a categorie ritmiche diverse, ma non quando appartengono allo stesso profilo (nell'esperimento di Nazzi e i colleghi, i bambini francesi erano stati sottoposti all'ascolto di materiale linguistico in lingua inglese, olandese e giapponese). I parlanti sono dunque in grado di operare questa distinzione, intuitivamente, già durante la prima infanzia.

Il sistema proposto da Ramus et al. (1999) si concentra sulle vocali come unità di base per la segmentazione del parlato, data la loro importanza per la comprensione del parlato stesso da parte dei bambini molto piccoli. Sono proposte due misure, a partire dalle quali Dellwo & Wagner (2003) indagano l'interazione tra velocità di eloquio e proprietà ritmiche del parlato:

1. La percentuale di intervalli vocalici (%V), definita come la percentuale del tempo di produzione dell'enunciato occupata da vocali;
2. La deviazione standard degli intervalli consonantici (ΔC).

In Ramus et al. (1999) viene considerato un terzo fattore, la deviazione standard degli intervalli vocalici (ΔV), assente dallo studio di Dellwo & Wagner (2003); tuttavia, il rapporto di quest'ultimo parametro con le proprietà ritmiche di una lingua rimane poco chiaro.

Dei parametri sopracitati è possibile identificare valori medi per ciascuna lingua. In particolare, le lingue iso-accentuali, che ammettono un numero maggiore di strutture sillabiche, presenterebbero ΔC maggiore dal momento che la principale fonte di variazione tra strutture sillabiche è rappresentata dalla possibilità di avere più consonanti; a sua volta, un ΔC maggiore comporta una riduzione di %V. Le lingue tradizionalmente definite iso-accentuali presenterebbero quindi %V maggiore (Ordin & Polyanskaya 2015).

Lo studio di Grabe & Low (2002), pur continuando a considerare i fattori temporali e la loro rilevanza nella classificazione delle lingue come iso-accentuali o iso-sillabiche, si allontana dalla ricerca di un'isocronia, considerando nuovamente %V e ΔC , oltre a un indice che esprime la variabilità nella lunghezza vocalica in misurazioni successive (rPVI), non normalizzato per *speech rate*. Un'altra variante di questa misura, il nPVI, è invece normalizzato. Si tratterebbe, secondo Low et al. (2000), di una misurazione più adeguata del ritmo rispetto alle deviazioni standard degli intervalli vocalici (ΔV) e consonantici (ΔC). I risultati dello studio condotto da Grabe & Low (2002) sono compatibili con la classificazione tradizionale, pur non rilevando una distinzione netta tra i due tipi di lingue. La proposta è quindi quella di adottare una distinzione categorica debole e non rappresentativa di tutte le lingue del mondo.

Dellwo & Wagner (2003) indagano nuovamente l'interazione tra il comportamento ritmico delle sillabe, la percentuale di intervalli vocalici (%V) e la deviazione standard degli intervalli consonantici (ΔC). Queste misure sarebbero decisive nella definizione della tipologia ritmica, e dai loro risultati ha luogo la cosiddetta *cluster hypothesis*, in base alla quale le lingue iso-sillabiche e quelle iso-accentuali tendono a concentrarsi intorno a valori di %V e ΔC diversi, formando quindi dei *cluster*.

I risultati ottenuti da Dellwo & Wagner (2003) sull'interazione tra la velocità di eloquio e le misure ΔC e %V sono parzialmente coerenti con questa ipotesi. Il loro studio ha coinvolto 16 parlanti, di tre diverse L1 (inglese, sia scozzese che americano, considerati come un'unica lingua nonostante le differenze diatopiche; francese; tedesco), a cui è stato fatto leggere un passo tratto da un romanzo tedesco, tradotto da esperti nelle altre due lingue. Dapprima, ai partecipanti è stato chiesto di leggere ad alta voce il testo a loro assegnato senza essere registrati;

dopodiché è stato chiesto loro di produrre diverse registrazioni, una in cui leggevano il testo a una velocità da loro reputata normale, poi in modo sempre più veloce e sempre più lento.

I valori di ΔC e %V mantengono la tendenza postulata da Ramus et al. (1999) a formare dei *cluster* a seconda della lingua di enunciazione, con inglese e tedesco, entrambe lingue iso-accidentali, che si collocano in aree più vicine tra di loro rispetto al francese, lingua iso-sillabica, a supporto della distinzione tradizionale. Questa tendenza si verifica a ogni velocità: nonostante il parametro della velocità di eloquio influenzi i valori di ΔC e %V, questa influenza non sembra essere sufficientemente forte da impedire la formazione di *cluster* (Dellwo & Wagner 2003).

La variazione di ΔC in relazione alla velocità di eloquio può essere spiegata dalla tendenza a realizzare intervalli consonantici più lunghi nel parlato lento, e viceversa; questo fatto, se confrontato con la lunghezza e con le strutture sillabiche proprie di una lingua, può spiegare, per esempio, il motivo per cui il ΔC in francese alla velocità normale è inferiore rispetto a quello in inglese e tedesco. Per questo fenomeno è stato proposto un ulteriore coefficiente che misuri la variazione di ΔC , chiamato *varco* ΔC , dal rapporto tra il ΔC moltiplicato per 100 e il valore medio degli intervalli consonantici in quella lingua. Questo coefficiente sembrerebbe maggiormente in grado di supportare la distinzione tradizionale e rivelerebbe che alcune lingue subiscono l'influenza della velocità di eloquio sulle loro caratteristiche ritmiche più di altre (Dellwo & Volker 2006).

L'intenzione dei parlanti di modificare la loro velocità di eloquio è comunque rispecchiata dallo *speech rate* misurato in condizioni di laboratorio, suggerendo che questo parametro può essere almeno parzialmente controllato; inoltre, i risultati sembrano supportare l'ipotesi che i parlanti abbiano una nozione di cosa costituisce una velocità d'eloquio normale per la loro lingua. La correlazione tra lo *speech rate* misurato e quello richiesto rimane forte, anche se con qualche differenza legata alla lingua parlata: i parlanti di francese sembrano in grado di esercitare un maggior controllo, seguiti da quelli di tedesco e infine dai parlanti di inglese.

In Dellwo et al. (2006) viene approfondita la questione del ruolo dell'intenzionalità nella velocità di eloquio mediante un esperimento di natura percettiva, chiedendo a parlanti nativi di francese di valutare la velocità di segmenti registrati, tratti dal Bonn-Tempo Corpus, in francese, inglese e tedesco. Di queste due ultime lingue i parlanti avevano una conoscenza solo parziale. La correlazione tra *speech rate* misurato e percepito rimane forte in tutte e tre le lingue, anche se maggiore nel francese, e anche se l'accuratezza risulta migliore nella valutazione del parlato veloce: i soggetti possiederebbero dunque una nozione di quale sia una velocità d'eloquio normale anche per lingue diverse dalla propria L1. Gli autori propongono come indizi

maggiori il *vocalic rate*³, oltre che fattori soprasegmentali come l'intonazione e la presenza di pause.

2.2. Variabili sociolinguistiche

La velocità di eloquio è stata ampiamente studiata in relazione al ruolo di variabili sociolinguistiche come il genere nella lingua madre di un individuo. Molti studi supportano l'ipotesi che gli uomini siano parlanti più veloci delle donne, presentando tempi di produzione più brevi.

Lo studio di Pépiot (2014) si concentra sulle differenze tra il parlato maschile e femminile, considerando diversi parametri oltre alle misure temporali: vengono misurati anche i valori medi di F0, oltre al *range* che esso assume nella produzione di un individuo, e il cosiddetto *phonation type*, ottenuto dalla differenza tra le prime due armoniche. A 20 parlanti di inglese degli Stati Uniti d'America e di francese parigino sono state fatte pronunciare diverse parole della propria L1, bisillabiche, sia reali che inventate, per coprire ogni combinazione di fonemi. Dai risultati emergono differenze sostanziali sia di genere sia di lingua rispetto a ogni parametro, nonostante le ridotte dimensioni dello studio (20 partecipanti in totale). I risultati suggeriscono che il genere e la lingua madre siano alla base di queste differenze, con alcune difformità di genere che appaiono consistenti cross-linguisticamente, come la maggiore ampiezza del *range* di F0 nelle donne o la tendenza delle stesse a impiegare più tempo nell'articolazione delle singole parole, da cui deriva una velocità d'eloquio ridotta.

Differenze di genere emergono anche in studi che coinvolgono l'utilizzo di *Delayed Auditory Feedback* (DAF). Si tratta di un procedimento atto a prolungare artificialmente l'intervallo di tempo che intercorre tra la produzione di un segmento parlato e la sua percezione da parte del soggetto che lo ha prodotto. Stuart & Kalinowski (2015) studiano la presenza e la frequenza di *disfluencies* prodotte da 32 soggetti (16 uomini e 16 donne) in condizioni di laboratorio, che hanno accesso al loro *output* ritardato artificialmente di intervalli di tempo variabili da 25 a 200 millisecondi. Ai partecipanti è stato chiesto di parlare dapprima a una velocità normale, dopodiché più velocemente; all'aumentare del DAF aumentano anche i fenomeni di *disfluency*, che si fanno più frequenti anche con l'aumentare della velocità. A valori

³ Il *vocalic rate* esprime, in questo caso, gli intervalli vocalici per secondo, misura utilizzata nello studio per esprimere lo *speech rate* misurato, al quale viene paragonato quello percepito dai partecipanti.

di DAF elevati, le donne presentano una velocità di produzione minore rispetto a quella degli uomini.

In termini percettivi, lo studio di Harnsberger et al. (2008) trova nella velocità di eloquio uno degli indizi principali sfruttati dagli ascoltatori per attribuire un'età ai parlanti, con un certo grado di accuratezza: le produzioni degli anziani (e quelle, modificate artificialmente, che sono state identificate dai soggetti dell'esperimento come prodotte da anziani) presentavano una durata maggiore sia per quanto riguarda le frasi, sia in riferimento alle sillabe e ai dittonghi. Questa tendenza, come altre proprie degli individui di età avanzata, è riconducibile a fattori fisiologici.

Inoltre, è stato osservato (Feldstein et al. 1993), tramite esperimenti di laboratorio, che le donne tendono a essere avvertite come parlanti più lente rispetto agli uomini a parità di velocità effettiva di eloquio. Alcuni campioni di parlato sono stati sottoposti agli ascoltatori (in questo caso, studenti universitari) in condizioni di laboratorio: le percezioni riportate dai soggetti sembrano essere influenzate dal genere dei parlanti, come anche dal *pitch* (Bond et al. 1988), misure comunque correlate tra loro. Anche il genere degli ascoltatori sembra influenzare i risultati: le ascoltatrici donne tendono a sovrastimare la velocità di eloquio di tutti i parlanti.

È dunque possibile che la diversa percezione della velocità del parlato, sia propria che altrui, sia alla base di alcuni stereotipi di genere: lo studio di Miller et al. (1976), dedicato alla persuasività del parlato, rileva una percezione di maggiore oggettività e intelligenza del parlato percepito come veloce; la velocità d'eloquio sarebbe così percepita come indice di credibilità. Dunque, una velocità percepita maggiore potrebbe generare un'apparenza di maggiore credibilità e intelligenza nei soggetti di genere maschile.

Queste conclusioni sembrano supportare Brown et al. (1974), che due anni prima identificava proprio la velocità di eloquio come il miglior predittore dei giudizi che i soggetti esprimono nei confronti della produzione. A partire da due campioni manipolati artificialmente in 27 combinazioni di *speech rate* e F0 media, l'esperimento rileva tuttavia che il parlato veloce e quello lento sono percepiti come meno "benevolenti" rispetto a quello che presenta una velocità normale (secondo i giudizi espressi dai partecipanti, a cui era stato chiesto di scegliere, tra alcuni aggettivi, quelli che meglio rispecchiavano lo stimolo). Il parlato veloce continua comunque a ottenere giudizi più positivi per quanto riguarda la competenza percepita del parlante.

Risultati simili sono ottenuti da Smith et al. (1975), di nuovo manipolando artificialmente sei voci, questa volta solo per quanto riguarda la velocità: la relazione tra *speech rate* e

percezione di competenza è lineare, la percezione di benevolenza è maggiore a una velocità di eloquio definita “normale”.

Tuttavia, in parziale disaccordo con quanto osservato da studi precedenti, Apple et al. (1979) rilevano che valutazioni migliori in termini di competenza, fluenza e persuasività sono attribuiti al parlato che presenta una velocità normale e che il parlato lento ottiene risultati nettamente inferiori, ed è percepito come passivo e meno veritiero; invece, gli studiosi non registrano miglioramenti significativi quando la velocità viene aumentata. Anche questo studio utilizza campioni manipolati artificialmente, in cui anche il *pitch* viene modificato; emerge la tendenza, già dimostrata, a giudicare meno competenti e meno veritieri i parlanti che presentano un *pitch* più alto. Tuttavia, sono considerate solo produzioni di parlanti uomini.

Le parlanti rimangono sottorappresentate in molti degli studi sopracitati. Feldstein et al. (2001) indagano il ruolo del genere e della velocità di eloquio nella percezione della personalità dell'interlocutore: i risultati supportano l'ipotesi che i soggetti giudicano come parlanti più competenti e socialmente più attraenti coloro che parlano con una velocità simile alla loro, ma il genere (sia dei parlanti che di coloro che esprimono i giudizi) si dimostra un fattore importante che influenza significativamente i giudizi. Tra coloro che erano chiamati a esprimere i giudizi, le donne hanno espresso valutazioni generalmente più elevate rispetto ai soggetti di genere maschile; al contempo, i parlanti uomini hanno ricevuto giudizi migliori in termini di attrattività sociale rispetto alle parlanti di genere femminile.

Negli esperimenti sopraelencati, i campioni erano stati manipolati per ottenere produzioni più veloci o più lente di quelle usuali, pratica criticata da molti per l'effetto di artificialità che ne risulta. L'esperimento di Feldstein et al. (2001) ha utilizzato campioni di parlato di tre studenti maschi e tre studentesse (già registrati per un altro esperimento), con una velocità di eloquio compresa tra 125 e 154 parole al minuto, e ha coinvolto 17 ascoltatori e 28 ascoltatrici, che hanno partecipato come giudici. È importante, anche se difficile da un punto di vista sperimentale, considerare il contesto in cui una produzione avviene e l'argomento che viene discusso: i campioni sottoposti agli ascoltatori da Feldstein et al. (2001) sono stati valutati per l'emotività tramite un sistema denominato *Thematic Apperception Test* (TAT) teorizzato da Murray (1943), e sono stati utilizzati solamente campioni che non risultavano esprimere forti emozioni, né positive né negative. Anche gli argomenti trattati e l'intensità vocalica sono stati controllati nello svolgimento dell'esperimento. Ai partecipanti è stato chiesto di valutare su una scala da 1 a 10 la velocità di eloquio degli stimoli, e la somiglianza tra la velocità percepita e la propria; sempre su una scala da 1 a 10, è stato chiesto loro di valutare i campioni rispetto a 10 aggettivi, sui quali si è poi basato il calcolo della competenza e dell'attrattività sociale percepita.

Le ascoltatrici donne hanno percepito tutti i parlanti, indipendentemente dal genere, come più competenti rispetto agli ascoltatori uomini; tutti gli ascoltatori hanno comunque mostrato di riporre maggiore fiducia nella competenza dei parlanti con velocità di eloquio simili alla propria velocità percepita, e non più veloci o più lenti. Tuttavia, quando la velocità reale dei parlanti era veramente più alta di quella degli ascoltatori, la valutazione era più positiva, soprattutto da parte delle ascoltatrici donne. Per quanto riguarda l'attrattiva sociale, di nuovo i parlanti uomini erano giudicati più positivamente, soprattutto da ascoltatori uomini con una velocità di eloquio inferiore, anche se i punteggi migliori erano attribuiti a coloro che mostravano una velocità simile a quella degli ascoltatori.

Non tutti, comunque, sostengono l'idea che una velocità di eloquio maggiore comunichi un maggior senso di competenza e persuasività, e O'Connell et al. (1989) cita proprio il parlato politico come ambito per eccellenza nel quale un parlato più lento può, al contrario, dimostrarsi più persuasivo. A 24 soggetti, parlanti di italiano e tedesco, è stato chiesto di leggere discorsi politici nella propria lingua madre originariamente pronunciati da personaggi politici di primo piano, e la stessa cosa è stata chiesta a due attori. I soggetti dell'esperimento hanno impiegato globalmente meno tempo sia dei politici che degli attori a pronunciare gli stessi discorsi, sia per *articulation rate* che per numero e durata delle pause; attori e politici hanno impiegato pressoché la stessa velocità e hanno fatto un uso simile delle pause (O'Connell et al. 1989). Questa tendenza è compatibile con l'assenza di una volontà di persuadere e di un *training* specifico; rimangono comunque notevoli differenze negli stili dei singoli politici.

Molta della ricerca su velocità di eloquio e persuasività interagisce con ipotesi legate alla ricerca di teorici della *speech accommodation*, come Giles et al. (1987) e Gallois et al. (2005), a loro volta debitori della teoria della cosiddetta *similarity-attraction* (Byrne 1997). Secondo la teoria della *speech accommodation*, una convergenza negli stili di parlato di due interlocutori porta a una percezione più positiva dello scambio e delle attitudini dell'altro. L'*accommodation* è un processo attraverso il quale gli interlocutori, quindi, modificano la propria produzione per mostrarsi più simili (o, in altri casi, dissimili, a seconda dell'esigenza comunicativa corrente) al proprio interlocutore. Alcuni studi hanno rilevato, per esempio, che nella comunicazione tra parlanti di inglese britannico di status socioeconomico disuguale i parlanti di status più elevato, oltre a rendere il proprio accento meno standard, tendevano anche a rallentare la propria velocità di eloquio, entrambi tratti tipici della lingua dell'interlocutore di status meno elevato. Il contrario avveniva da parte dei parlanti di status basso durante la conversazione (Thakerar et al. 1982). In molte culture, infatti, una velocità di eloquio elevata è percepita come socialmente prestigiosa, e può essere un tratto desiderato dai parlanti quando essi costruiscono la propria

immagine di sé nell'interazione. La velocità di eloquio percepita, unita ad altri fattori di natura linguistica, può influire sul trattamento che una persona riceve in una vasta gamma di situazioni, per esempio nel ricevere valutazioni in un contesto scolastico (Cargile et al. 1994).

È importante notare che nella maggior parte degli studi trattati finora non si può parlare realmente di convergenza, in quanto non avviene comunicazione quando ai soggetti dell'esperimento sono sottoposti stimoli registrati. Tuttavia, il paragone con la velocità di eloquio del soggetto dell'esperimento che agisce da giudice (o almeno con la sua stessa percezione di essa) è parte integrante di numerosi studi sulla velocità di eloquio.

Competenza e attrattività sociale sono investigate da Street & Brady (1982), dimostrando, con 261 studenti intervistati, che esiste una correlazione positiva tra i giudizi espressi dagli ascoltatori e la somiglianza percepita con la propria velocità di eloquio, soprattutto per quanto riguarda la percezione di competenza. Gli autori lamentano l'inadeguatezza dei modelli per la valutazione percettiva del parlato e cercano di implementare un elemento contestuale che influenza i margini di accettabilità della variazione di velocità. Street et al. (1983) ottengono risultati simili, indagando di nuovo il ruolo del contesto, e nello specifico la differenza tra un contesto formale come un colloquio di lavoro e una conversazione informale. Nei colloqui di lavoro, i soggetti, nuovamente studenti universitari, trovano che velocità inferiori siano più accettabili e dimostrano di essere generalmente più consapevoli delle differenze in termini di velocità di eloquio. Giudicano comunque più competenti e affidabili i parlanti che presentano una velocità simile alla propria, o leggermente superiore.

Rimane dunque fondamentale la definizione di un'ulteriore variabile: la percezione di similarità tra lo *speech rate* dei campioni a cui i soggetti sono esposti e il proprio, oltre al dato assoluto della velocità di eloquio, che rimane comunque stabilmente legato a una percezione di maggiore competenza.

Buller & Aune (1988) offrono una lettura del fenomeno influenzata dalla *speech accommodation theory*, che postula una relazione tra l'abilità di un soggetto nella decodifica e la sua velocità di eloquio. Coloro che si mostrano più abili nella decodifica (e parlano quindi più velocemente) sarebbero maggiormente bendisposti nei confronti di richieste formulate velocemente, e viceversa. L'abilità nella decodifica è stata calcolata mediante una misura proposta da Rosenthal nel 1979, il cosiddetto *Profile of Nonverbal Sensitivity* (PONS). L'esperimento, che ha coinvolto 168 partecipanti, sembra confermare le ipotesi; la rilevanza di questo dato è legata parzialmente anche al fatto che richieste percepite favorevolmente hanno maggiori possibilità di essere soddisfatte.

Buller & Aune (1992) riprendono queste idee nella cornice teorica della *communication accommodation theory* e considerando attrattività sociale e soddisfazione delle richieste in relazione alla velocità di eloquio con la quale esse sono espresse e alla sua somiglianza con quella degli ascoltatori; postulano inoltre la necessità per gli interlocutori di soddisfare le richieste che vengono loro poste per favorire lo sviluppo positivo del loro rapporto interpersonale (Roloff 1987). Dall'esperimento, condotto con 320 studenti universitari del Sud-Ovest degli Stati Uniti, è emerso che velocità di eloquio simili portano a una maggiore percezione di intimità e immediatezza, che a sua volta porta a una maggiore probabilità di un esito favorevole. Non è però riscontrato un rapporto diretto tra velocità di eloquio e probabilità di soddisfazione della richiesta.

Il ruolo della somiglianza di velocità di eloquio in relazione all'attrattività sociale e alle richieste, e alla probabilità che esse hanno di essere ricevute con successo, è ripreso in Buller et al. (1992) con l'aggiunta di una variabile importante: il beneficio che l'ascoltatore trae dal soddisfare le richieste dell'altro. La velocità con cui sono prodotte le richieste avrebbe un effetto favorevole sulla probabilità che esse vadano a buon fine, ma solo quando l'ascoltatore ritiene che obbedire sia nel proprio interesse. I risultati non sono comunque definitivi, e altri fattori (intimità, personalità) sembrerebbero influenzare i processi interpersonali coinvolti.

2.3. Tendenze cross-linguistiche; contenuti informativi

Uno studio in senso cross-linguistico delle caratteristiche del parlato, tra cui anche la velocità d'eloquio, è interesse dello studioso Frank Seifart, che a tale scopo ha curato il *corpus* DoReCo (Language Documentation Reference Corpus), che presenta oltre 100 ore di registrazioni di testi, principalmente di natura narrativa, parlati in 51 lingue, delle quali molte sono considerate *endangered*. Si tratta di una raccolta che vede le sue origini nel lavoro sul campo a stretto contatto con lingue non ampiamente documentate e studiate (Seifart 2022). Le lingue presenti nella raccolta coprono tutti i continenti abitati e tutte le macrofamiglie linguistiche identificate.

A partire dai dati contenuti nel DoReCo, per esempio, Paschen et al. (2022) studiano il fenomeno dell'allungamento finale (*final lengthening*) su un *corpus* di 25 lingue, molte delle quali scarsamente documentate o in via di estinzione. È comprovata la tendenza interlinguistica all'allungamento dei confini di unità prosodiche di livello superiore rispetto alla parola, e soprattutto del loro segmento terminale (Wightman et al. 1992). Questa tendenza è stata interpretata come universale e incorporata in molti modelli della produzione parlata. L'analisi di Paschen et al. (2022) si concentra sui fattori dell'allungamento finale che possono essere

considerati particolari e propri di una specifica lingua, e sulle interazioni tra questo fenomeno e alcune caratteristiche fonologiche delle lingue, come la lunghezza vocalica con valore distintivo. Emerge dunque che, mentre il fenomeno è generalizzato a tutte le lingue osservate, anche quelle che non presentano durata vocalica fonologica, varia l'entità dell'allungamento vocalico finale (limitato alla penultima e ultima sillaba). Delle lingue che invece ammettono contrasti tra la durata delle vocali, tale allungamento poteva essere relegato a specifiche posizioni, come la penultima sillaba, o interagire in modo più complesso con la variabile della posizione, senza però che sia attestato un allungamento unicamente in ultima posizione.

La velocità di eloquio è legata anche alla densità d'informazione contenuta nei segmenti pronunciati (Priva 2017): contenuti più informativi sono spesso accompagnati da un rallentamento nel parlato di individui sia di genere maschile che di genere femminile, e la velocità di eloquio stessa è spesso utilizzata come indicatore dello sforzo cognitivo coinvolto nella pianificazione di un determinato segmento di parlato.

In un *corpus* creato *ad hoc*, contenente parlato conversazionale spontaneo in inglese, olandese, e diverse lingue scarsamente documentate appartenenti a tutti i continenti del mondo, Seifart et al. (2018) rilevano per ciascun idioma una tendenza al rallentamento nel segmento immediatamente precedente un nome, posizione nella quale sono anche più frequenti le pause. Il rallentamento avviene dunque con maggiore frequenza prima dei nomi, e non prima dei verbi: questo dato è almeno parzialmente in opposizione con quanto precedentemente rilevato, per esempio, da Szekely et al. (2005), che riscontrano invece tempi di produzione maggiori per i verbi quando a parlanti adulti di inglese sono state mostrate illustrazioni che raffiguravano oggetti e azioni.

In uno studio sulla lingua Chintang, una lingua polisintetica della famiglia kiranti parlata nel Nepal orientale, Lester et al. (2020) indagano il ruolo della velocità di eloquio nell'acquisizione da parte dei bambini della capacità di distinguere verbi e nomi; i mezzi prosodici sono normalmente i canali privilegiati per effettuare tale distinzione nel parlato rivolto agli adulti. Esiste un corpus audiovisivo longitudinale dedicato all'acquisizione della lingua Chintang, che è considerata *endangered* secondo gli standard dell'UNESCO (2003). I verbi in Chintang sono più complessi morfologicamente rispetto ai nomi (in entrambi i casi si tratta di parole multimorfemiche: si tratta dunque di materiale molto diverso rispetto alle lingue più comunemente studiate, come l'inglese e il francese), per i quali ci si aspetta dunque qualche fenomeno di compressione fonetica. Tuttavia, lo studio non rileva differenze significative in termini di velocità di eloquio tra verbi e nomi, in netta opposizione con quanto riscontrato in altre lingue come l'inglese e il cinese mandarino, ed è coerente con Seifart (2018) la tendenza

a rallentare prima dei nomi, rivolgendosi sia ad adulti che a bambini; entrambe queste tendenze sono comuni ad adulti e bambini. La differenza tra nomi e verbi risiederebbe dunque nel segmento immediatamente precedente anziché nella parola stessa.

Inoltre, nella ricerca di correlati fonetici per le categorie grammaticali, emerge la tendenza propria dei nomi ad essere prodotti con una durata maggiore. Il fenomeno è stato variamente spiegato, principalmente attraverso l'osservazione del fatto che verbi e nomi mostrano una forte tendenza a comparire in sedi diverse all'interno dell'enunciato: sedi, come quella finale all'interno del sintagma, portatrice privilegiata dell'accento di frase, che ne influenzerebbero le caratteristiche temporali (Cooper & Paccia-Cooper, 1980).

Nel loro studio, Lohmann & Cowell (2020) cercano ulteriori spiegazioni per la differente durata della produzione di verbi e nomi, partendo da considerazioni legate alla probabilità di incontrare il confine di un'unità intonativa dopo verbi e nomi e alla loro frequenza differente, in termini di *type* e *token*, senza escludere la possibilità che si tratti di un effetto legato puramente alla categoria di appartenenza della parola in questione. Gli studiosi hanno utilizzato una lista di coppie di parole che presentassero caratteristiche comuni (monosillabiche, con attacco e coda sillabica piene, che avessero come nucleo una vocale fra le tre selezionate), delle quali è stata indicata anche la frequenza. Le parole *target* così delineate sono state poi inserite in contesti sintattici di due tipi diversi, per ottenere diversi tipi di sintagmi preposizionali (di lunghezza variabile) e di confini prosodici nel segmento immediatamente successivo alla parola. Anche la prevedibilità di ogni parola dato il contesto è stata presa in considerazione. Le frasi così costruite sono state fatte leggere a quattro soggetti, due uomini e due donne, selezionati tra 78 partecipanti. I risultati confermano la tendenza dei nomi ad essere seguiti da confini prosodici maggiormente rilevanti rispetto ai verbi; soprattutto, è confermata la durata maggiore della produzione dei nomi rispetto a quella dei verbi, di un margine che si rivela sempre significativo. I risultati dello studio suggeriscono che la rilevanza del confine prosodico che segue la parola rivesta il ruolo più importante, entro un modello che si rivela però complesso, comprendente diversi fattori.

Lo studio di Lohmann & Cowell (2020) considera dunque anche la frequenza, che sembra avere una forte influenza sulla durata della produzione sia delle parole che delle vocali in esse contenute. La velocità di eloquio sarebbe infatti legata anche alla frequenza delle parole, che influenzerebbe positivamente la velocità a cui le parole stesse tendono a essere prodotte nel parlato spontaneo (Dell 1990; Jurafsky 2003).

Non solo la frequenza di determinate parole, ma anche dei *pattern* fonologici in esse presenti, sembra giocare un ruolo: ulteriori studi si sono spinti infatti al di sotto dei confini della

parola, analizzando le sequenze di fonemi e la loro frequenza. Si tratta della cosiddetta probabilità fonotattica, e le restrizioni fonotattiche sarebbero incluse nei meccanismi legati alla produzione e alla processazione del parlato. La frequenza dei *pattern* fonologici è strettamente legata ai giudizi di plausibilità attribuiti alle pseudo-parole in una lingua, e numerosi studi hanno dimostrato, per esempio, che parlanti adulti di inglese presentano meno difficoltà a produrre e processare sequenze fonologiche frequenti, indipendentemente dalla loro complessità a livello articolatorio (Goldrick & Larson 2008).

Le minori difficoltà nella produzione sembrano portare a una velocità di articolazione più elevata: bambini e adulti a cui è stato chiesto di produrre e ripetere parole inesistenti le producono più lentamente quando le sequenze fonologiche in esse contenute sono inusuali, e solo i bambini dimostrano difficoltà nel ripetere sequenze rare (Munson 2001). Gli stessi soggetti, quando viene chiesto loro di giudicare la plausibilità di pseudo-parole, danno punteggi che correlano con la frequenza delle sequenze fonologiche presenti in esse.

Nel caso di una coppia di omofoni, dei quali uno è nettamente più frequente dell'altro, l'ipotesi della cosiddetta *frequency inheritance* prevede che l'omofono più raro sia accessibile quanto quello più comune, e dunque presenti un comportamento simile, poiché essi condividono la stessa forma fonologica, in cui risiederebbe l'informazione relativa alla frequenza. Effettivamente, nell'esperimento condotto da Jescheniak & Levelt (1994), i partecipanti, a cui era stato chiesto di tradurre in olandese alcune parole inglesi, impiegano lo stesso tempo a produrre parole comuni e parole rare con omofoni comuni. Anche la tendenza a produrre meno errori nel caso delle parole rare con omofoni comuni è confermata.

Tuttavia, l'impossibilità di replicare questi risultati in altri studi che hanno coinvolto altre lingue ha portato a formulare ulteriori ipotesi relative alla localizzazione delle informazioni circa la frequenza e le possibili differenze tra lingue diverse (Jescheniak et al. 2003). Navarrete et al. (2006) rilevano che la frequenza delle parole influisce sulla velocità con la quale i soggetti sono in grado di attribuire loro un genere grammaticale in spagnolo, *task* per il quale viene attivato il lemma, e non la forma fonologica di una parola.

I risultati relativi alla durata dell'articolazione variano dunque in base alla frequenza, e in relazione a una serie di variabili difficilmente quantificabili come il contesto prosodico, la ripetizione entro uno stesso discorso, la prevedibilità. Nel caso delle 223 coppie di omofoni presenti nel corpus di inglese americano Switchboard, indagate da Gahl (2008), la studiosa rileva che la più comune delle due parole tende a essere pronunciata più velocemente, in modo statisticamente significativo, suggerendo dunque che le informazioni relative alla frequenza potrebbero risiedere a livello del lemma, e non della forma fonologica (l'esempio citato dalla

studiosa è la coppia *time* ‘tempo’ e *thyme* ‘timo’). È necessario, tuttavia, considerare anche la velocità di eloquio nei segmenti adiacenti alla parola, oltre che la sua prevedibilità nel contesto (calcolata con un modello a bigrammi), l’ortografia, la sua categoria sintattica, la prossimità alla fine del segmento, e dunque il fenomeno dell’allungamento finale.

Questi risultati sono comunque attribuibili anche all’abitudine, e dunque a fattori di natura puramente articolatoria. Infine, tipi di studi e di dati diversi possono portare a significative differenze nei risultati: Whalen (1996) riscontra differenze nella produzione degli omofoni rari quando essi sono contenuti in liste di parole rare e in liste miste; tuttavia, si tratta di dati radicalmente diversi da quelli utilizzati da Gahl (2008), tratti da un *corpus* di telefonate, e dunque esempi di parlato principalmente spontaneo.

2.4. Applicazioni in linguistica clinica

La velocità di eloquio è stata anche ampiamente studiata in linguistica clinica, soprattutto in relazione alla diagnosi precoce di demenza senile e malattia di Alzheimer (Gagliardi & Tamburini 2021). La demenza senile e le altre patologie neurodegenerative raggruppate sotto questa etichetta causano deficit in diverse aree, tra le quali anche quella linguistica: possono causare diversi tipi di disturbi acquisiti del linguaggio, che riguardano diverse aree della produzione, dalla conoscenza semantica agli aspetti pragmatici (Calzà et al. 2021). Sono intaccati da tali patologie anche gli aspetti temporali del parlato, nella misura in cui una velocità di eloquio ridotta, unita nella maggior parte dei casi a difficoltà nella selezione delle parole, potrebbe essere un buon predittore di una patologia in atto.

Numerosi studi si sono concentrati sulle caratteristiche temporali del parlato degli individui affetti dagli stadi iniziali della malattia, quando esibiscono cioè un *mild cognitive impairment*. L’identificazione di caratteristiche che correlano con malattie neurodegenerative ai primi stadi è utile, in primo luogo, per lo sviluppo di strumenti di elaborazione automatica del parlato che possono fungere da strumenti diagnostici in ambito clinico. La diagnosi precoce di queste malattie neurodegenerative è vitale per la loro gestione sia a livello individuale, data la maggiore efficacia delle terapie disponibili ai primi stadi, sia a livello sociale, dal momento che il numero di pazienti, complice l’aumento dell’aspettativa di vita media, rimane molto elevato a livello mondiale e in continuo aumento. Inoltre, l’identificazione di pazienti affetti da *mild cognitive impairment* offre numerose opportunità per studi di natura farmacologica e nuovi approcci clinici (Gagliardi & Tamburini 2021).

La maggior parte degli studi trattati sono opera di linguisti ungheresi e riguardano parlanti di ungherese; gli studi longitudinali rimangono rari, per la difficoltà nella loro attuazione e per la natura recente degli strumenti impiegati. In un primo studio sui parametri temporali del parlato e sul loro rapporto con la manifestazione di questi disturbi, Hoffmann et al. (2010) identificano un primo set di *features* utili per la diagnosi precoce nel parlato di soggetti di lingua ungherese: si tratta di *articulation rate*, *speech tempo*, tasso di esitazioni, numero di errori grammaticali prodotti. L'*articulation rate* differisce notevolmente tra soggetti sani e soggetti affetti da patologie, ulteriormente suddivisi in tre classi in base alla severità del sintomo; lo *speech tempo* in particolare presenta differenze apprezzabili tra soggetti sani e soggetti affetti in modo lieve, così come il numero e la frequenza delle esitazioni (la malattia si fa più evidente nel parlato spontaneo, quando dunque è assente qualsiasi tipo di pianificazione del discorso che può sopperire alle difficoltà nel cercare le parole). La difficoltà nella selezione delle parole, che emerge in *task* diagnostiche come il TOT (*tip of the tongue*) test, è stata ampiamente documentata nei soggetti affetti dalla malattia di Alzheimer. Il numero di errori grammaticali diventa invece veramente significativo solo nella produzione dei soggetti agli ultimi stadi della malattia.

Gosztolya et al. (2016) identificano alcune *features* ritenute importanti per la diagnosi precoce di questi disturbi, a partire dalla produzione di parlanti di lingua ungherese. Utilizzando un riconoscitore addestrato *ad hoc* sull'ungherese e alcuni modelli del linguaggio, è possibile effettuare un paragone con i parametri identificati in precedenza, con misure di *precision*, *recall*, *F-measure* oltre a una misura, definita *Unweighted Average Recall*, che esprime la media dei valori assunti dal *recall* per tutte le classi, ed è dunque meno sensibile alla frequenza delle classi rispetto a *F-measure*. È confermata l'importanza già attribuita alle pause, sia piene che vuote, alle esitazioni, e ai fattori demografici; inoltre, grazie all'uso di strumenti automatici, emergono numerose *features* relative ai singoli fonemi, finora ignorati. Si tratta di fonemi tipicamente correlati alla realizzazione delle esitazioni, per esempio l'allungamento della consonante finale di una parola in una situazione di parlato spontaneo.

Lo studio di Vincze et al. (2016) ha un obiettivo simile, limitandosi però all'analisi delle trascrizioni delle produzioni dei soggetti, parlanti nativi di ungherese. Per rendere il parlato il più possibile spontaneo, ai soggetti sono mostrati dei film e sono poste domande relative ai contenuti, sia immediatamente dopo la visione sia il giorno successivo. Le trascrizioni sono state effettuate da linguisti, e comprendono quindi informazioni relative alle pause e alle esitazioni. Emerge la tendenza delle persone affette da *mild cognitive impairment* (di qualsiasi natura e origine) a produrre più esitazioni, più pause sia vuote sia piene, per sopperire alle

difficoltà legate alla selezione lessicale; tali soggetti ricorrono inoltre più spesso a parafrasi, pronomi e altre espressioni vaghe.

Altre *features* identificate dagli studiosi includono il numero di *token* e di lemmi utilizzati; la frequenza delle diverse parti del discorso, dei verbi alla prima persona, dei neologismi; il numero e la frequenza di espressioni indefinite e di termini legati all'attività della memoria, che possono essere indice di qualche tipo di difficoltà; il numero di negazioni; il rapporto tra parole-funzione e parole-contenuto, oltre che il numero di parole tematicamente legate ai contenuti dei filmati guardati. Questo insieme di *features* è quindi testato tramite un modello a *Support Vector Machine* (SVM) sulle trascrizioni della produzione di 84 parlanti, ottenendo un punteggio di accuratezza del 69.1%.

Anche lo studio di Gosztolya et al. (2019) utilizza modelli computazionali per individuare i soggetti affetti da Alzheimer o *mild cognitive impairment*, con l'obiettivo di affinare l'identificazione delle *features* fonetiche rilevanti, comparando questi modelli ad altri che utilizzano soltanto le trascrizioni (e, dunque, *features* di natura unicamente morfologica e sintattica). Sono stati utilizzati dati di lingua ungherese, tratti da un *database* creato dal dipartimento di psichiatria dell'università di Szeged (Ungheria), contenenti parlato di persone affette da Alzheimer ai primi stadi, da *mild cognitive impairment*, e di persone prive di disturbi che hanno fatto da gruppo di controllo per la valutazione dei modelli. Anche in questo caso ai soggetti è stato mostrato un filmato e sono state poste loro alcune domande immediatamente dopo la visione e nei giorni successivi. Il test ha coinvolto in totale 75 parlanti, 25 per gruppo; il numero dei parlanti è stato ridotto a 50 in alcune fasi dello studio, per la necessità di accorpate i soggetti affetti da Alzheimer e quelli affetti da *mild cognitive impairment*, permettendo dunque di utilizzare un classificatore di tipo binario. In tutti i casi, l'aggiunta di *features* di natura fonetica migliora i risultati ottenuti rispetto a quanto osservato usando i soli elementi semantici e morfologici ottenuti dalle trascrizioni, e l'utilizzo di entrambi i *feature set* porta ai risultati migliori. I parametri fonetici sono dunque un'utile aggiunta nello sviluppo di modelli di NLP che possano essere utilizzati come supporto in ambito diagnostico.

2.5. Velocità d'eloquio nell'acquisizione L2

Anche nell'ambito dell'acquisizione della L2 la velocità di eloquio è stata studiata come parametro che correla con la fluenza dell'apprendente. Il Quadro comune europeo di riferimento per la conoscenza delle lingue (CEFR) fa esplicita menzione alla velocità di eloquio, all'uso delle pause e ad altri fenomeni riguardanti i fattori temporali del parlato, a

livello sia di comprensione che di produzione orale. La Tabella 3 delle linee guida europee prevede che, al livello C2, l'apprendente sia in grado di «esprimersi spontaneamente, dettagliatamente, con un andamento colloquiale naturale»⁴ e che già al livello B2 sia in grado di «produrre segmenti linguistici con un ritmo abbastanza regolare»⁵. Il livello B1, immediatamente precedente, prevede invece che la presenza delle pause sia ancora molto evidente⁶. Uno studio del fenomeno è reso possibile dall'esistenza di *corpora* specialistici di lingua degli apprendenti: ad alti livelli di competenza linguistica, ancora tra il livello B2 e C1 come definiti dal CEFR, si registra effettivamente un aumento della velocità di eloquio e una riduzione sostanziale delle pause (Huang & Gráf 2020).

In particolare, le *features* soprasegmentali del parlato legate all'organizzazione temporale della produzione sono indizi centrali sia per il discernimento di lingue diverse sia per l'identificazione degli accenti dei parlanti L2.

Risultati simili sono ottenuti da Ordin & Polyanskaya (2015b) in uno studio che ha coinvolto parlanti nativi di francese e tedesco, apprendenti di inglese a vari livelli: la differenza in termini di velocità tra la produzione nella L1 e nella L2, e tra apprendenti a livelli diversi di competenza linguistica, è altamente significativa e percepibile; inoltre, il ritmo dell'inglese parlato dagli apprendenti presenta sempre più caratteristiche proprie delle lingue iso-accentuali col procedere dell'acquisizione, indipendentemente dalla L1, anche se i parlanti di tedesco arrivano mediamente a produrre *pattern* più simili a quelli dell'inglese, la lingua *target*. La somiglianza ritmica tra la L1 e la L2 aiuterebbe dunque il processo di acquisizione, mentre le caratteristiche temporali della produzione possono essere usate come predittori del livello di acquisizione ottenuto da un parlante (Ordin et al. 2011).

I parlanti a livelli più avanzati di acquisizione tendono inoltre a parlare più velocemente (Ordin & Polyanskaya 2015b). Dal momento che alcune delle misure trattate in precedenza, come ΔC e rPVI, sono dipendenti dallo *speech rate*, le differenze tra i vari livelli di competenza emergono dai parametri normalizzati come nPVI e varco ΔC , ed emerge in generale una tendenza alla stabilizzazione di questi parametri con l'aumentare della competenza linguistica di un parlante L2 (Ordin et al. 2011).

⁴ «[...]express him/herself spontaneously at length with a natural colloquial flow»

⁵ «[...]produce stretches of language with a fairly even tempo»

⁶ <https://www.coe.int/en/web/common-european-framework-reference-languages/table-3-cefr-3.3-common-reference-levels-qualitative-aspects-of-spoken-language-use>

La velocità e il ritmo sono anche fattori in base ai quali i parlanti nativi giudicano l'accento straniero degli apprendenti. Sono state date diverse definizioni alla nozione di accento straniero (Combei 2023). Alcune di esse prendono avvio dalle caratteristiche oggettive della produzione di parlanti L2, riconoscibili come tali dai parlanti nativi: caratteristiche che deviano dalla produzione dei parlanti nativi per una serie di elementi appartenenti al dominio fonetico, prosodico e paralinguistico (Jenner 1976), di cui è stato evidenziato il carattere “non patologico” (Munro 1998).

Altre definizioni si sono soffermate sulla percezione dei parlanti nativi quando sono esposti alla produzione di un parlante L2, e non su qualche caratteristica misurabile sperimentalmente: Flege (1981) definisce l'accento straniero come una percezione esperita dall'ascoltatore, mentre Jilka (2000) aggiunge che perché un accento sia percepito come straniero, deve ricordare ai nativi un'altra lingua. Distingue inoltre tra accento straniero limitato alla componente fonetica, quando un apprendente che ha acquisito il repertorio fonologico della L2 non è in grado di ricrearne i suoni nello stesso modo di un parlante nativo, e un accento straniero fonologico, quando l'acquisizione del repertorio non è ancora avvenuta.

La tematica dell'accento straniero si è affermata come campo di indagine nell'ambito degli studi di *Second Language Acquisition* (SLA) negli ultimi anni, come conseguenza della maggiore mobilità di persone nel mondo contemporaneo. La disciplina si è a lungo interrogata su quale sia il modo migliore di insegnare la pronuncia agli apprendenti, sperimentando diversi metodi: l'utilizzo di alfabeti fonetici, come l'IPA, o esercizi mirati che sfruttano le coppie minime della lingua *target*. Più recentemente però è stata messa in discussione la necessità di minimizzare l'accento straniero degli apprendenti: l'obiettivo dovrebbe essere un accento che sia il più possibile comprensibile (Jenkins 2002), soprattutto quando la lingua *target* è l'inglese, data la funzione che questa lingua ormai riveste nel mondo contemporaneo.

Ciò non significa che gli accenti stranieri non possano essere causa di stigma sociale, e in generale venire utilizzati per comunicare diverse sensazioni, anche in ambito pubblicitario: Lombardi Vallauri (2019) nota la tendenza, comune nelle pubblicità e nel materiale promozionale, a utilizzare il francese per veicolare eleganza, raffinatezza. Questo almeno fino ai primi anni '90: dopodiché, sarebbe l'inglese a rivestire lo *status* di massimo prestigio. Viceversa, il fenomeno della discriminazione linguistica porta alcuni accenti stranieri, come per esempio quello romeno in Italia, ad essere percepiti come portatori di connotazioni negative (Combei 2023).

I parlanti nativi sono comunque in grado di riconoscere l'accento straniero dei propri interlocutori, come già verificato da Flege (1984) sottoponendo la produzione in inglese di

parlanti francesi al giudizio di parlanti nativi di inglese americano. I parlanti nativi sono stati in grado di identificare correttamente l'accento anche quando non avevano esperienza di inglese con accento francese e gli stimoli utilizzati consistevano in segmenti molto brevi, anche una frazione di sillaba.

Per quanto riguarda la natura fonetica degli "indizi" che portano all'identificazione degli accenti, quando le produzioni dei parlanti L2 sono sottoposte al giudizio di parlanti nativi, questi ultimi sembrano basarsi più sulla velocità di eloquio che su altre caratteristiche ritmiche per determinare il livello di competenza linguistica (Ordin & Polyanskaya 2015a). Tuttavia, come rilevato da Polyanskaya et al. (2017), il ritmo rimane per i parlanti nativi un indizio più forte nell'identificazione del parlato L2. In questo esperimento sono state utilizzate voci artificiali, create a partire dalle produzioni reali di apprendenti francesi di inglese a diversi livelli di competenza linguistica. Dalle registrazioni sono state però rimosse differenze di contorno intonativo e di F0, sia separatamente che congiuntamente, in modo che i giudizi potessero essere formulati unicamente sulla base di parametri temporali quali *articulation rate* e caratteristiche ritmiche. Gli stimoli così ottenuti sono stati sottoposti ai soggetti, parlanti nativi di inglese. I soggetti sono stati meno in grado di percepire differenze a livello di velocità di eloquio, ma maggiormente in grado di percepire differenze ritmiche, ascoltando le tracce in cui il contorno intonativo non era stato neutralizzato artificialmente; ciò suggerisce un ruolo centrale rivestito dal contorno intonativo nell'identificazione degli accenti stranieri, soprattutto nel caso dei parlanti francesi.

Di nuovo mediante l'utilizzo di stimoli sintetizzati, e non di produzioni reali di apprendenti, con un procedimento simile a Polyanskaya et al. (2017), Polyanskaya & Ordin (2019) studiano l'influenza della velocità d'eloquio e del ritmo nella percezione dell'accento da parte di ascoltatori nativi monolingui. Lo studio conferma quanto già rilevato nel caso dei parlanti francesi: il ritmo costituisce l'indizio più importante laddove L1 e L2 sono lingue che presentano caratteristiche ritmiche diverse. Nel caso dei parlanti di tedesco, lingua ritmicamente più simile all'inglese, la velocità si rivela invece molto più influente.

Ancora, sulla modulazione della velocità di eloquio da parte di parlanti nativi durante la comunicazione con parlanti non nativi, è stata osservata una riduzione della velocità di eloquio, accompagnata però da una velocità di articolazione paragonabile a quella impiegata durante la comunicazione con altri parlanti nativi e dalla presenza di pause più frequenti e più lunghe (Derwing 2008).

3. Il discorso politico

Molto è stato detto in ambito linguistico sul discorso politico e sulle sue peculiarità; il rapporto tra politica e lingua sembra essere per molti versi indissolubile e così è stato teorizzato fin dagli albori, con l'importanza attribuita alla retorica classica. Il linguaggio politico è stato spesso letto attraverso la lente della semantica dei *frame*, come nell'opera di Robin T. Lakoff, *The language wars*. Alcuni esempi di analisi di questo genere, che riguardano l'Italia contemporanea, si trovano anche in Zanchi (2022) e Coschignano et al. (2023). Ancora, nell'ambito della *Critical Discourse Analysis*, il lavoro di Fairclough (2000) raccoglie diversi materiali linguistici prodotti dal partito laburista inglese e ne analizza i contenuti in rapporto ad alcune tematiche.

Molti studi hanno invece riguardato singoli elementi o parametri linguistici e paralinguistici impiegati nella comunicazione politica; dal punto di vista puramente fonetico, Dietrich et al. (2019) cercano correlazioni tra alcuni parametri fonetici, l'argomento trattato dai discorsi pronunciati al Congresso degli Stati Uniti d'America e l'esito delle votazioni successive. L'attenzione degli studiosi è stata rivolta anche al ruolo del corpo e della gestualità dei politici, che agiscono in contemporanea col parlato per la creazione del messaggio (Maricchiolo et al. 2014; Scherer et al. 2012). A questo proposito gli studi si sono avvalsi anche di risorse come i *corpora* (Poggi et al. 2010).

Inoltre, l'avvento dei nuovi media ha modificato le forme della comunicazione pubblica preesistenti (Bouvier & Way 2021; Graham et al. 2016), in particolare per la natura interattiva di molte piattaforme, che permettono una comunicazione diretta e in tempo reale tra il politico e la sua base elettorale (Combei & Reggi 2024). Infine, alcuni studi hanno indagato le modalità di utilizzo della piattaforma Twitter da parte dei politici (Graham et al. 2016; Macagno 2022), con analisi anche di natura pragmatica (Garassino et al. 2022).

3.1. Strategie linguistiche per il mantenimento del potere

Molti studi si sono interessati al parlato politico dal punto di vista della semantica e della pragmatica. Robin T. Lakoff (2000) applica ai discorsi che hanno caratterizzato momenti chiave della politica e dell'opinione pubblica americana la teoria, nata nell'ambito della linguistica cognitiva, della semantica dei *frame*, individuata come strategia linguistica di mantenimento del potere. La semantica dei *frame* prevede che gli elementi linguistici siano in grado di evocare nella mente dei partecipanti all'interazione un *frame*, che può essere concepito come una

struttura cognitiva che porta con sé una serie di aspettative e di ruoli reciproci. La conoscenza di queste situazioni è tipicamente condivisa dai partecipanti all'interazione, e da esse deriva la nozione di "buonsenso"; quando la condivisione viene meno, possono nascere fraintendimenti.

Lakoff (2000) individua il dibattito politico come un *frame* a sé stante, che porta dunque con sé una serie di aspettative. Il titolo stesso del libro di Lakoff, *The language war*, fa riferimento alla guerra attorno alla lingua che sarebbe in corso, legata alla difesa di vecchi *frame* e al tentativo di crearne di nuovi attraverso la lingua. L'esempio che la linguista porta come più importante, per interesse mediatico e controversie ad esso legate, è il paradigma del politicamente corretto (*politically correct*), definito come l'insieme delle forme linguistiche, ai tempi relativamente nuove, coniate da gruppi e individui che di norma fino a quel momento non avevano avuto la possibilità di scegliere le parole con le quali venivano definiti. I *frame* culturali in età contemporanea non sono quindi più decisi unilateralmente e senza possibilità di controversia da parte di un solo gruppo dominante.

Il concetto di *frame* è legato a un altro concetto, molto utilizzato in linguistica: si tratta della marcatezza. È già saussuriana la distinzione tra elementi linguistici non marcati e le loro controparti marcate, che sarebbero portatrici di una maggiore quantità di informazioni, più rare, e più complesse. Ogni elemento che rientra in un *frame* sarebbe infatti non marcato, e in quanto tale non richiederebbe spiegazioni e giustificazioni; inoltre, una volta che un partecipante è immerso in un *frame*, diventa molto difficile uscire da esso. Anche solo l'atto di rendere espliciti i *frame* soggiacenti a un'idea o a un discorso può essere fonte di conflitto: «spesso, i tentativi di rendere visibili i *frame* o di modificare idee largamente accettate vengono accolti con derisione. Dal momento che la derisione è supportata dalle credenze convenzionali, essa "ha senso" e non richiede una spiegazione o una giustificazione esplicita»⁷ (Lakoff 2000, p. 52).

Lakoff identifica, a partire da eventi di comunicazione pubblica degli Stati Uniti d'America degli anni '90, fenomeni di creazione e di mantenimento di nuovi *frame* che comunichino implicitamente messaggi specifici, col fine ultimo di ottenere e mantenere il potere e il consenso.

Uno dei *frame* citati dall'autrice, che rappresenta un esempio di terminologia controversa, è quello legato al concetto di *dipendenza* in relazione al grado di libero arbitrio che viene attribuito a coloro che ne soffrono, e al dibattito su quali sostanze debbano essere regolate

⁷ «Often, attempts to make frames visible or to change generally accepted ideas are met with derision. Because the derision is supported by conventional beliefs, it "makes sense" and does not require explicit explanation or justification».

legalmente. Se al giorno d'oggi è più o meno accettato che, nel caso delle sostanze stupefacenti, si tratta di uno stato patologico che, quindi, va oltre la volontà dell'individuo (con un cambiamento già notevole rispetto a pochi decenni prima), non è invece chiaro quanta responsabilità debba essere attribuita a coloro le cui dipendenze sono di altra natura (gli esempi citati dall'autrice sono lo *shopping* e il sesso). Il modo in cui questi individui sono considerati è molto importante nel decidere come essi debbano essere trattati all'interno della società, poiché chi pratica comportamenti negativi volontariamente deve essere punito mentre chi lo fa involontariamente deve essere aiutato (Lakoff 2000).

Il concetto di *frame* è legato a quello di implicito linguistico, approfondito nel Capitolo 4, e spesso co-occorre con fenomeni appartenenti alla tipologia delle presupposizioni, che possiamo preliminarmente definire come messaggi impliciti il cui contenuto deve essere dato per vero da chi accetta un enunciato come appropriato. Sempre Lakoff (2000) cita come esempio l'utilizzo di presupposizioni negative (entro un discorso altrimenti veritiero) per distanziare il pubblico da un gruppo specifico di persone: sarebbe il caso del termine *vittima*, connotato negativamente nella società statunitense, soprattutto nelle fazioni più conservatrici, che tengono in grande considerazione l'individualismo e l'imprenditorialità (Lakoff 2000), qualità che si vuole presupporre che l'individuo definito come *vittima* non abbia. Ancora, il medesimo effetto è ottenuto tramite l'utilizzo di aggettivi negativi quali *stridulo* per descrivere il modo di parlare di qualcuno, che possono riverberare la loro connotazione negativa sull'individuo e sul suo gruppo di appartenenza. Un'altra comune presupposizione individuata dall'autrice, di cui sarebbe intrisa la società americana del tempo, è quella della presunta superiorità di un passato idealizzato, che porta alla strenua difesa dello *status quo*: è bene fidarsi di quanto è familiare e noto, e per questo rientra nel *frame* di riferimento.

Tutti questi concetti, e molti altri, sono veicolati attraverso la lingua, nei discorsi pronunciati dalle personalità politiche e non: la lingua sarebbe dunque in grado non solo di descrivere la realtà, ma anche di costruirla.

Per una recente applicazione della semantica dei *frame* all'ambito italiano, si veda per esempio Luraghi & Stringa (2024), che offrono una panoramica sul dibattito intorno ad alcuni argomenti geopolitici di particolare interesse come la pandemia da Covid-19, la guerra e le migrazioni, concentrandosi sul ruolo della comunicazione e in particolare sull'utilizzo dei *social media*. Coschignano et al. (2023) analizzano la rappresentazione dei migranti e del fenomeno delle migrazioni a partire da titoli di giornali italiani pubblicati tra il 2013 e il 2021. Dallo studio emerge la tendenza a parlare di migranti come numeri, problemi o oggetti; per esempio, essi compaiono spesso entro il *frame* della rimozione, con verbi come *espellere*, o del

collocare, con verbi come *sgomberare*. Inoltre, si parla di migranti spesso entro *frame* che hanno a che fare con la quantificazione, nei quali viene specificato il loro numero o incremento. Esistono interessanti differenze tra le modalità utilizzate dalle diverse pubblicazioni nel parlare di migranti: i giornali di orientamento religioso si discostano dagli altri per una maggiore frequenza di *frame* positivi, come l'assistenza e la somiglianza, mentre alcuni *frame* sono propri dei giornali di destra, come quelli legati alla rimozione e al ricollocamento.

Parlare di migranti in termini numerici, di masse o incrementi, implicitamente li disumanizza: Zanchi (2022) definisce questo processo disumanizzante in quanto i migranti non sono più visti come singoli individui. Inoltre, la stampa amplificherebbe la portata del fenomeno delle migrazioni, che tende a presentare come una minaccia. Sempre Zanchi (2022) affronta il tema delle 'ingiustizie discorsive' perpetrate attraverso il discorso ai danni di alcuni gruppi, come migranti e donne, anche attraverso l'attivazione di *frame*. Questi soggetti fanno parte di quello che è chiamato *outgroup*, esterno rispetto all'*ingroup*, dominante, e rispetto ad esso più debole. In quanto appartenenti all'*outgroup*, queste categorie sociali sono spesso vittime di rappresentazioni e atteggiamenti negativi nei loro confronti.

La persuasività rimane uno degli obiettivi principali del parlato politico, e l'abilità di parlare rimane una capacità fondamentale per chi pratica la comunicazione pubblica in ogni ambito. Molta della ricerca nell'ambito della comunicazione politica si inserisce a pieno titolo nella corrente denominata *Critical Discourse Analysis*, che «indaga l'uso della lingua oltre il livello della frase, oltre ad altre forme di creazione del significato quali elementi visivi e sonori, visti come irriducibili nella (ri)produzione della società attraverso la semiosi»⁸, e «mira a snaturare il ruolo che i discorsi rivestono nella (ri)produzione di strutture non inclusive e non egualitarie, e sfida le condizioni sociali nelle quali queste strutture sono incastonate»⁹ (Wodak 2015a). Si tratta di un orientamento per sua natura interdisciplinare, con vaste intersezioni con la storia, la demografia e la sociologia; l'analisi non si ferma al discorso, ma arriva a toccare il legame che esso intrattiene con le istituzioni, le ideologie e le stesse strutture di potere (Fairclough 2013). La lingua sarebbe dunque uno strumento utilizzato consciamente per ottenere e mantenere il potere, inteso come rapporto asimmetrico tra fazioni diverse della società, oltre che per costruire e rinforzare visioni del mondo composte da rappresentazioni, opinioni, giudizi.

⁸ «investigates language use beyond the sentence level, as well as other forms of meaning-making such as visuals and sounds, seeing them as irreducible elements in the (re)production of society via semiosis».

⁹ «aims to denaturalize the role discourses play in the (re)production of noninclusive and nonegalitarian structures and challenges the social conditions in which they are embedded».

Il volume di Fairclough (2000) *New labour, new language?* conduce un'analisi del discorso politico britannico a partire da diverse fonti, sia materiali a stampa che discorsi pronunciati da personalità politiche appartenenti al partito laburista dal 1994 al 1999. Fairclough ha creato per il suo studio un vero e proprio *corpus* di materiali provenienti da personalità e canali ufficiali del partito laburista, anche in formato leggibile elettronicamente, per poter fornire informazioni legate alle occorrenze di determinate parole; inoltre, ha raccolto materiali simili provenienti da periodi storici precedenti, aggiungendo una dimensione diacronica all'indagine. A partire da questi materiali ha condotto il suo studio concentrandosi su alcuni temi particolarmente significativi. Per esempio, dall'analisi delle occorrenze della parola *valore* entro questo *corpus*, emerge che la sua frequenza è aumentata rispetto ai periodi storici precedenti, ma essa ha perso la connotazione economica che aveva in passato, oltre che i legami con il socialismo, e compare prevalentemente in contesti legati al *cambiamento*.

3.2. *Approcci multimediali*

Nel caso del libro di Fairclough (2000), è lo studioso ad aver creato un *corpus* ad hoc per il proprio studio, a partire dalla raccolta dei dati. In anni recenti, la linguistica dei *corpora* ha reso disponibili i mezzi attraverso i quali è possibile effettuare ricerche sistematiche sul parlato politico a partire da raccolte di testi in formati leggibili dalle macchine. Al contempo, l'analisi ha abbandonato la sola parola scritta, finora dominante, in favore di una molteplicità di canali che agiscono sincronicamente per trasmettere un certo messaggio (Maricchiolo et al. 2014; Poggi et al. 2010; Scherer et al. 2012).

Con l'avvento dei nuovi *media* gli elementi visivi e paralinguistici hanno assunto un'importanza ancora maggiore: è oggi possibile avere materiali video di qualità molto elevata delle personalità politiche, registrati in occasione di dibattiti o interviste, che includono anche primi piani di specifiche parti del corpo (Maricchiolo et al. 2014): c'è motivo di postulare un rapporto tra quanto il pubblico vede del corpo di un politico e i giudizi e le valutazioni che esprime, non sempre consapevolmente.

Lo studio di Poggi et al. (2010) si concentra sulla gestualità corporea e, in particolare, offre una rassegna di tipologie e funzioni rivestite dall'atto di annuire in una raccolta di dibattiti televisivi svizzeri di argomento politico tratti dal *corpus* SSPNet (*Social Signal Processing Network*). Si tratta di uno studio di natura principalmente qualitativa, atto a sviluppare una tassonomia di cenni basata sulla funzione e sul ruolo di chi li effettua nel dibattito (parlante, interlocutore, ascoltatore terzo, esterno al dibattito), a partire da 100 gesti presenti nel *corpus*.

L'annotazione dei dati è molto articolata, e include informazioni quali l'identità del parlante e dell'interlocutore, il contesto verbale che ha accompagnato la produzione del gesto, e la descrizione di come il gesto è stato prodotto fisicamente. Anche i gesti, come la parola, sono soggetti a polisemia, complicata dalla presenza di numerosi altri canali che agiscono contemporaneamente: per esempio, lo stesso cenno può assumere un significato molto diverso se è accompagnato da un sorriso. È dunque necessario dapprima identificare tutti i possibili significati di un cenno, per poi cercare i segnali (linguistici e non) che più spesso ricorrono con ciascuna funzione: per esempio, quando è l'interlocutore ad annuire dopo che è stata fatta una proposta, il cenno può indicare che essa è stata accettata. Annuire può anche fungere da ringraziamento, dopo che si è ricevuto un gesto di cortesia o un favore, ma anche da *backchannel*, per segnalare che il messaggio è stato recepito correttamente, o per dare enfasi a un segmento della produzione verbale, come avviene per gli elementi di una lista, insieme ai gesti delle mani che sono tipici di questo tipo di enumerazione.

Ancora, Scherer et al. (2012) effettuano un'analisi simile a partire da sedute plenarie del parlamento tedesco, distinguendo tra movimenti della testa, del corpo, e delle mani. I materiali così raccolti sono stati sottoposti a sette soggetti, sia in formato audio-video sia solo video; in aggiunta, durante gli esperimenti è stato utilizzato un dispositivo in grado di effettuare l'*eye tracking*. Ai soggetti è stato poi chiesto di valutare su una scala numerica quanto i parlanti che avevano appena visto mostrassero di essere *insicuri, monotoni, espressivi, persuasivi*, e *globalmente ottimi oratori*. Lo studio, seppure di scala ridotta, rivela legami incoraggianti tra il tipo di movimento e le qualità percepite dai partecipanti; inoltre, conferma l'ipotesi che in assenza di audio i soggetti si concentrano meno sul volto e più sui movimenti corporei dei politici, come rilevato dall'*eye tracking*.

Si è quindi iniziato a indagare il ruolo di altri elementi della comunicazione, diversi dalla parola, allontanandosi dal primato assoluto dell'analisi delle trascrizioni, dovuto anche alla relativa facilità nel processare testi scritti. L'operazione non è priva di criticità, dovute principalmente alla difficoltà nel processare automaticamente materiale multimediale e nel produrre uno schema di annotazione.

Il *corpus* PoliModal, curato da Trotta & Guarasci (2021), raccoglie filmati di interviste andate in onda durante un *talk show* di argomento politico sul canale Rai 3 tra il 2017 e il 2018, corredate da trascrizioni, metadati, annotazione di fenomeni relativi alla produzione (come false partenze, pause, sovrapposizioni di turni), ma anche da un'annotazione relativa a fenomeni non linguistici come la posizione e la gestualità del corpo, del viso, delle mani. Per il *corpus*

IMPAQTS, da cui sono stati tratti i materiali utilizzati per il presente studio, si rimanda al Capitolo 4.

Nonostante l'esistenza di *corpora* anche multimediali di discorso politico, e la relativa facilità nel reperire il materiale grazie agli archivi televisivi e degli organi istituzionali, pochi studi si sono concentrati sulle caratteristiche puramente fonetiche del parlato politico.

Anche alcuni studi di cui si è già parlato hanno utilizzato come dati proprio le produzioni di personalità politiche, esempi di parlato, peraltro, difficilmente definibile come spontaneo per la tendenza a pianificare, a grandi linee se non parola per parola, i discorsi da pronunciare nelle occasioni pubbliche: si veda a tal proposito O'Connell et al. (1989), di cui si è parlato nel Capitolo 2, per un'analisi anche cross-linguistica, dal momento che sono stati utilizzati materiali sia in inglese che in tedesco. È interessante come gli autori abbiano voluto paragonare le modalità con cui i politici pronunciano i loro discorsi con quelle utilizzate da attori dotati di formazione scenica e anni di esperienza: entrambe le categorie hanno pronunciato i discorsi più lentamente rispetto al gruppo di controllo, costituito da studenti privi di formazione teatrale, con un utilizzo quantitativamente maggiore, ma consapevole, delle pause.

Lo studio di Dietrich et al. (2019) fa uso di una vastissima quantità di materiale audio proveniente da discorsi tenuti nel Congresso degli Stati Uniti d'America tra il 2009 e il 2014, raccolti grazie al servizio *HouseLive*¹⁰. Grazie al sottotitolaggio automatico e alle informazioni contenute in esso, è stato possibile utilizzare più di 74 mila discorsi: si tratta di un campione eccezionalmente ampio per il tipo di studio, che ha preso in analisi la variazione di *pitch* e l'intensità emotiva in relazione agli argomenti dei singoli discorsi e al genere dei parlanti. L'intensità emotiva è stimata dagli studiosi sulla base di minime variazioni nel *pitch* che non sarebbero controllabili dal soggetto che parla; è stata dunque calcolata F0 per ogni parlante, considerabile come la *baseline* a partire dalla quale sono calcolate le eventuali variazioni. Per un'annotazione degli argomenti è stato utilizzato un filtro che, a partire dalle trascrizioni generate automaticamente dei discorsi, annota quali di essi contengono parole quali *woman*, *female*, *girl*: è stato creato così uno schema di annotazione binario, che distingue i discorsi che contengono queste parole da quelli che non le contengono. Più del 17% dei discorsi pronunciati da donne sono stati etichettati dal programma, contro il 9% di quelli prodotti da uomini. Tutte le donne, indipendentemente dall'orientamento politico, presentano *pitch* più alto quando parlano di altre donne, mentre negli uomini non si registra una tale differenza; le donne che

¹⁰ <https://live.house.gov/>

dimostrano una maggiore variazione di *pitch* quando discutono questo argomento ricevono inoltre un maggior supporto elettorale da parte di gruppi orientati agli interessi delle donne. Gli autori postulano che ascoltare discorsi pronunciati da legislative donne, che presentino tratti di maggiore intensità emotiva, possa trasmettere tale intensità emotiva anche agli ascoltatori, anche quando si tratti di uomini: questa ipotesi sarebbe confermata dai risultati elettorali che seguono i discorsi presi in analisi.

3.3. Populismo e nuovi media

In anni recenti, comunque, l'avvento e la rapida diffusione dei cosiddetti nuovi media ha completamente rivoluzionato le modalità e la frequenza della comunicazione pubblica, grazie a nuove piattaforme quali Facebook, X (precedentemente Twitter), e Instagram, talvolta utilizzate come canali ufficiali di comunicazione da parte di personalità e organizzazioni politiche (Graham et al. 2016).

Si tratta in ogni caso di forme di comunicazione pubblica vera e propria, e dunque non spontanee, bensì studiate scrupolosamente alla luce del pubblico a cui sono rivolte, che è ora in grado di fornire un *feedback* in tempo reale. Infatti, questi nuovi media si distinguono rispetto alla televisione e alle altre forme più tradizionali per la possibilità di interazione tra le personalità politiche e il pubblico, sotto forma di risposte, reazioni a ciascun *tweet*, citazioni dirette dei messaggi dei politici tramite *retweet*. Graham et al. (2016) studiano, per esempio, il ruolo che Twitter ha avuto nelle elezioni politiche del 2010 in Olanda e Regno Unito: i politici olandesi coinvolti hanno utilizzato, in media, il *social network* in misura maggiore rispetto alle loro controparti britanniche, e in maniera molto più interattiva, rispondendo più frequentemente a commenti e interventi da parte dei cittadini.

Si rimanda al Capitolo 4 per lo studio di Garassino et al. (2022) sui *tweet* prodotti da capi di governo e opposizione in Italia e nel Regno Unito, e sul ruolo degli impliciti nei messaggi in essi contenuti. Lo studio di Garassino et al. (2022), che usa i *tweet* di quattro personalità politiche (due italiani e due inglesi), rivela l'esistenza di veri e propri profili comunicativi individuali, che influenzano sia il processo di annotazione che i risultati ottenuti. Le personalità studiate dimostrano di utilizzare diversi tipi di impliciti con finalità diverse, entro strategie comunicative differenti legate anche al ruolo istituzionale che ricoprono.

In un contesto più ampio, il ruolo di queste piattaforme è stato ampiamente studiato in relazione alla nascita e alla diffusione di informazioni false (*fake news*) e del discorso d'odio (*hate speech*), soprattutto per quanto riguarda il contesto socio-politico statunitense

(Matamoros-Fernández & Farkas, 2021). Rispetto ai media tradizionali, piattaforme come X (ex Twitter) e Facebook contribuirebbero alla diffusione di questi contenuti per la loro natura interattiva e per il meccanismo per il quale ciascun utente è in grado di personalizzare la propria esperienza, circondandosi dunque di altri utenti con interessi e valori simili ai propri, generando un vero e proprio effetto di *echo chamber* (Combei & Reggi 2024; Graham et al. 2016).

Tuttavia, la semplice esplicitazione di contenuti falsi o di contenuti esprimenti odio è un fenomeno molto diverso, e circoscritto, rispetto al linguaggio manipolatorio, che riguarda i contenuti impliciti, l'utilizzo di espressioni cariche emotivamente, e gli schemi argomentativi (Macagno 2022). Queste strategie sono in grado di ridurre il livello di attenzione prestato dall'ascoltatore, e/o di allontanare la responsabilità nei confronti del messaggio da chi lo pronuncia: è più probabile, dunque, che il contenuto venga accettato passivamente, senza che venga meglio indagato, quanto non viene esplicitato, nei casi in cui, se lo fosse, non sarebbe immediatamente accettato (Lombardi Vallauri 2019). L'analisi si muove dunque nella direzione dello studio dei contenuti che non vengono direttamente esplicitati, in particolare quelli veicolati da fenomeni di presupposizione: si tratta di contenuti più difficili da identificare, e dunque da moderare, su piattaforme come i *social network*.

Proprio per cercare una visione d'insieme di questo fenomeno, Macagno (2022) parte da alcuni concetti propri dell'analisi dell'argomentazione e della pragmatica per ottenere uno schema di annotazione che renda conto della tipologia di argomento utilizzato e dell'eventuale presenza di fallacie logiche, impiegate come strumenti di manipolazione. Un esempio di strategia manipolatoria è quella di sfruttare le presupposizioni per trasmettere un messaggio che è in contraddizione con quanto accettato dagli interlocutori come *common ground*, per esempio attraverso l'utilizzo di aggettivi o strutture sintattiche che implicano concetti che non sono altrimenti esplicitati o accettati come veri. Uno degli esempi citati dall'autore è il *tweet* con cui Donald Trump dichiara che «il 20 gennaio 2017 sarà ricordato come il giorno in cui il popolo è tornato ad essere al comando di questa nazione»¹¹, con la presupposizione che il popolo non fosse più al comando della nazione fino al momento del suo insediamento. È possibile ottenere un effetto simile anche attraverso una falsa dicotomia, dove due opzioni alternative sono rappresentate come contraddittorie, anche quando non lo sono realmente.

L'annotazione delle tipologie di argomenti e delle strategie di manipolazione è stata effettuata indipendentemente da due annotatori e ha ricevuto punteggi di *inter-annotator*

¹¹ «January 20th, 2017, will be remembered as the day the people became the rulers of this nation again».

agreement, come il kappa di Cohen¹², elevati. Questa analisi è in grado di delineare un profilo argomentativo di quattro personalità politiche, a partire dai loro *tweet*, nei primi sei mesi dalla loro elezione. Le personalità politiche “populiste” farebbero uso di strategie quali fallacie logiche, argomentazioni non solide e parole cariche emotivamente con una frequenza maggiore rispetto alle altre, oltre a presentare un maggior numero di attacchi diretti.

Di populismo già parlava Wodak (2015b) nel suo volume *The politics of fear*; la letteratura in merito è ampia e ci limiteremo a fornire alcune definizioni. Alla base del populismo sarebbe una concezione manichea della società, nettamente divisa in gruppi dei quali uno, autoctono e uniforme, sarebbe minacciato dall’altro, percepito come estraneo e pericoloso, che funge da capro espiatorio: è questa la politica della paura di cui parla la studiosa. La lingua è uno strumento fondamentale per creare e rafforzare questa visione della società e per definire questi gruppi: ne è un esempio la tendenza a utilizzare un linguaggio carico di emotività per stabilire un contatto con il proprio pubblico. Tuttavia, non si tratta comunque di strategie universali e molto dipende dal singolo Paese e dalla sua cultura (Combei & Reggi 2024).

Per i moderni partiti populistici, l’utilizzo dei nuovi *social media* sarebbe fondamentale, con strategie di costruzione dell’immagine pubblica che spesso si allineano a quelle utilizzate dalle celebrità; viceversa, gli esponenti dei partiti tradizionali avrebbero più difficoltà nell’utilizzo di questi nuovi canali comunicativi (Wodak 2015b).

Data la specificità di queste nuove forme di comunicazione rappresentate dai moderni *social media*, si è scelto di escludere i discorsi tenuti *online* dal presente studio, limitando la selezione a discorsi pronunciati in parlamento, durante i comizi o in televisione. Discorsi di questo genere sono comunque presenti all’interno del *corpus* IMPAQTS, che prevede un’etichetta apposita per l’annotazione di questo tipo di materiali.

¹² Si intende con kappa di Cohen una misura di *inter-annotator agreement* ampiamente usata in linguistica computazionale per misurare la bontà di uno schema di annotazione e delle sue linee guida, partendo dall’assunto che un buono schema e linee guida chiare e valide porteranno annotatori diversi a fornire le medesime etichettature dei dati. Più in generale, si tratta di una misura che stima il grado di accordo tra due valutazioni di natura qualitativa. Il kappa di Cohen si ottiene dal rapporto tra l’accordo osservato (le istanze, cioè, che i diversi annotatori hanno etichettato nello stesso modo), a cui viene sottratto l’accordo atteso (quello cioè che si avrebbe se gli annotatori avessero etichettato ogni segmento in maniera del tutto casuale), e la massima differenza possibile tra le annotazioni, se esse fossero casuali.

4. Lo studio

In questo studio si è voluta investigare la relazione tra la velocità di eloquio e alcuni fattori di natura sociolinguistica nel contesto del discorso politico; in particolare, da una selezione di discorsi tratti dal *corpus* IMPAQTS, trascritti e annotati con i dati di ciascun parlante, si è voluto indagare il rapporto tra *speech rate*, *articulation rate* e alcune variabili quali genere, orientamento politico, fascia d'età del parlante e provenienza geografica.

Per la parte sperimentale dello studio, sono stati utilizzati *tool* di linguistica computazionale quali sistemi di riconoscimento vocale (*Automatic Speech Recognition*, o ASR), *webapp* e altre risorse disponibili online gratuitamente, e programmi ampiamente utilizzati in fonetica sperimentale, come Praat. Fondamentale è stato IMPAQTS, un *corpus* multimediale che raccoglie discorsi politici pronunciati nell'Italia repubblicana.

In un secondo momento, partendo dall'annotazione pragmatica fornita dal corpus IMPAQTS, che riguarda diversi tipi di impliciti linguistici, sono state analizzate eventuali differenze in termini di velocità di eloquio tra enunciati contenenti presupposizioni ed enunciati che sono privi di qualsiasi tipologia di implicito. L'utilizzo di questi dispositivi pragmatici in un discorso di natura politica spesso non è una scelta neutrale, e può essere motivato dall'intenzione di distrarre da alcuni contenuti, lasciati impliciti e spesso discutibili, o non dimostrati, che il parlante trasmette.

4.1. *Corpora* multimediali e il corpus IMPAQTS

Nelle indagini di natura fonetica sono molto utili risorse, e in particolare *corpora*, contenenti materiale audio e talvolta anche video; si tratta purtroppo di risorse poco comuni, poiché archiviare e annotare materiale multimediale è un'operazione molto dispendiosa, dal punto di vista del tempo e delle risorse, oltre che dello spazio di archiviazione richiesto, necessariamente molto maggiore rispetto a quello necessario per immagazzinare soli testi scritti. In particolare, la necessità di trascrivere ed eventualmente elaborare ulteriormente il materiale audio è particolarmente dispendiosa in termini di tempo: il processo può essere almeno parzialmente automatizzato tramite programmi che forniscono una trascrizione automatica, ma è comunque necessario l'intervento umano per il controllo e l'aggiunta di altre informazioni linguistiche e di metadati. La maggior parte dei *corpora* esistenti, soprattutto di lingua italiana, contiene esclusivamente testi scritti, o trascrizioni di eventi comunicativi parlati.

Nell'ambito degli studi sull'italiano parlato, il *corpus* LIP, raccolto negli anni '90, rappresenta una delle tappe più importanti (De Mauro & Mancini, 1993): da circa 60 ore di registrazione di lingua parlata, con vari livelli di spontaneità, si intendeva creare un lessico di frequenza per l'italiano parlato contemporaneo. Il *corpus* CLIPS (Corpora e Lessici dell'Italiano Parlato e Scritto), trascritto e annotato dall'Università di Napoli Federico II, offre, nelle varie sottosezioni (determinate dal genere di parlato in esse contenuto), *file* in formato .wav corredati da metadati e trascrizioni in formato .txt. I dati di parlato sono divisi in sottocartelle a seconda della città di provenienza dei parlanti, per limitare l'influenza della variazione diatopica o, al contrario, per permetterne lo studio.

Il *corpus* KiParla, nato dalla collaborazione delle Università di Bologna e di Torino, raccoglie oltre 150 ore di parlato suddiviso in 3 moduli (le sottosezioni KIP, ParlaTO e KIPasti), registrato a microfono palese (Mauri et al. 2019). Il materiale è annotato per quanto riguarda la tipologia di interazione (conversazioni durante i pasti, conversazioni libere, esami, interviste semistrutturate, lezioni, ricevimento studenti), il numero di partecipanti, il tipo di rapporto, simmetrico o asimmetrico, che intercorre tra i partecipanti, la presenza di un moderatore, l'anno di raccolta e il luogo dove è stato registrato. Dei parlanti sono annotati elementi utili per l'analisi sociolinguistica quali l'occupazione, il genere, la regione di provenienza, l'età e il titolo di studio. Il materiale è annotato seguendo una versione semplificata del sistema Jefferson (2004) comunemente usato per l'analisi della conversazione; il *corpus* è consultabile grazie alla piattaforma NoSketch Engine (Rychlý 2007).

Una risorsa importante per lo studio dei discorsi politici in italiano è rappresentata dal *corpus* IMPAQTS (*Implicit Manipulation in Politics – Quantitatively Assessing the Tendentiousness of Speeches*), un *corpus* multimediale nato dalla collaborazione dell'Università di Firenze con l'Università di Roma Tre all'interno del Progetto di Rilevante Interesse Nazionale per il triennio 2020-2022 che porta lo stesso nome del *corpus* (Cominetti et al. 2022).

Si tratta di un *corpus* che, una volta ultimato, conterrà 1500 discorsi pronunciati da 150 personalità politiche italiane nel periodo che va dal 1946 al 2022, accompagnati quando possibile da *file* audiovisivi. Con un totale stimato di 3,5 milioni di parole, sarebbe uno dei maggiori *corpora* multimediali esistenti per la lingua italiana. L'annotazione prevede che tutto il materiale venga lemmatizzato e annotato per parti del discorso; i metadati includeranno informazioni di tipo sociolinguistico, sui singoli parlanti, e di tipo pragmatico, relativamente al discorso.

Riguardo ai singoli politici sono annotati nome, sesso, età, ruolo politico, forza politica, orientamento politico; quest'ultimo parametro rappresenta una classe chiusa con sei possibili valori. Relativamente al discorso, sono annotati canale, pubblico, struttura, lunghezza (in caratteri) oltre al tipo di discorso. Quest'ultimo parametro prevede i seguenti valori: assemblea, comizio, riunione di partito, dichiarazione trasmessa, dichiarazione in presenza, dichiarazione trasmessa sui nuovi media (Cominetti et al. 2022). Per ognuno dei 150 politici selezionati si è cercato di includere una varietà di tipi di discorso.

4.2. Annotazione pragmatica del corpus IMPAQTS

Inoltre, è disponibile nel *corpus* IMPAQTS un'annotazione pragmatica riguardante la presenza o assenza di contenuti impliciti nei discorsi, effettuata seguendo le linee guida riprese e ampliate da Lombardi Vallauri & Masia (2014) e Lombardi Vallauri (2019).

La presenza di impliciti è comunemente riconosciuta come una strategia persuasiva efficace nel discorso politico; si tratta comunque di un fenomeno comune a tutta la comunicazione parlata, che non sempre ha fini manipolatori (Lombardi Vallauri 2019). L'utilizzo, per esempio, di implicature e presupposizioni è estremamente comune e spesso impiegato per non esplicitare informazioni o referenti che sono già noti all'interlocutore, o che non richiedono grandi sforzi cognitivi da parte sua per essere individuati, al fine di rendere la comunicazione più agile e veloce: in questi casi, dunque, l'uso di impliciti non costituisce una strategia atta a manipolare o nascondere qualcosa. Questi dispositivi possono però anche essere utilizzati per non rendere esplicito l'intero messaggio veicolato da una proposizione, affinché l'interlocutore non diventi mai pienamente consapevole di esso e non possa così rifiutarlo o dubitare del suo valore di verità (Lombardi Vallauri & Masia 2014). Un'altra conseguenza dell'utilizzo degli impliciti, e in particolare delle presupposizioni, è l'allontanamento metaforico del parlante dalla propria proposizione, da cui consegue la mancata presa di responsabilità nei confronti del contenuto: egli si comporta come se quell'informazione fosse già nota all'interlocutore, a causa di qualche altra situazione o evento diverso dall'evento comunicativo in atto.

Nel *corpus* IMPAQTS, si è cercato di annotare questi fenomeni soltanto quando l'affermazione resa implicita dal loro utilizzo è stata giudicata non *bona fide* vera.

Nel processo di annotazione del *corpus* sono state segnalate implicature (convenzionali e conversazionali, sia generalizzate che particolari), presupposizioni, topicalizzazioni e istanze di vaghezza. Vediamo brevemente questi elementi, uno per uno.

4.2.1. *Implicature*

Con implicature convenzionali intendiamo quelle implicature che sono attivate da qualche elemento di natura linguistica contenuto nella proposizione stessa (Levinson 1983): si tratta, per esempio, di quelle attivate dal connettivo *ma*, o da *quindi*, in grado di generare un contrasto o un rapporto di causa-effetto tra i due concetti che vengono affiancati.

Per esempio, una delle implicature convenzionali annotate nel corpus è la seguente:

1. *si inserisca la magistratura stabilendo addirittura come atto dovuto un'incriminazione per sequestro di persona aggravato* [ABER18-T1].

Di ogni implicatura è fornita un'esplicitazione: in questo caso, l'implicazione è che sia stato eccessivo stabilire come atto dovuto un'incriminazione per sequestro di persona aggravato, ed è attivata dall'avverbio *addirittura*.

Con implicature conversazionali intendiamo quelle implicature attivate dall'enunciato che viene prodotto e dal dare per assodato che il parlante si stia attenendo a quelli che Grice definisce il principio di cooperazione e le massime conversazionali. Il principio di cooperazione (o di cooperazione conversazionale) recita: «conforma il tuo contributo conversazionale a quanto è richiesto, nel momento in cui avviene, dall'intento comune accettato o dalla direzione dello scambio verbale in cui sei impegnato» (Grice 1975: 26 (tr. it. 1993: 60)). Non si tratta comunque di un principio di natura prescrittiva: nella formulazione di Grice è più che altro una norma costitutiva della comunicazione umana. Esso si articola nelle quattro massime che Grice definisce conversazionali e che prescrivono quanto segue:

1. di fornire un contributo che sia appropriato per lunghezza e informatività alle necessità della conversazione,
2. che il contributo fornito sia vero (nei limiti delle conoscenze e della consapevolezza del parlante),
3. che il contributo fornito sia appropriato e rilevante nel contesto,
4. che sia formulato in modo chiaro e comprensibile.

Un enunciato che viene prodotto, alla luce del principio di cooperazione e delle quattro massime, oltre che di altri elementi del contesto esterni alla conversazione, può attivare un'implicatura conversazionale nel momento in cui una massima è apparentemente violata pur continuando a rispettare il principio di cooperazione. Tale violazione non è nascosta dal parlante, che vuole quindi ostentare il suo rifiuto di seguirla per veicolare un messaggio, senza

però esplicitarlo: questo fenomeno è definito da Grice ‘implicatura da sfruttamento’, in quanto deriva dallo sfruttamento delle quattro massime da parte del parlante.

Un esempio di implicatura conversazionale presente nel *corpus* è il seguente:

2. Potremmo fare riferimento ai tanti casi in cui uomini delle istituzioni sono stati iscritti al registro degli indagati e la loro dimissione non è stata richiesta, ma non lo facciamo [AALF10-A1].

Il contenuto implicato è che non sia necessario elencare i precedenti di indagini nei confronti di uomini delle istituzioni per argomentare la propria posizione.

La distinzione tra implicature conversazionali generalizzate e particolari è legata alla dipendenza dal contesto: le implicature conversazionali generalizzate dipendono unicamente dall’enunciato e dalla presupposizione di cooperazione tra i parlanti. Un esempio di implicatura generalizzata presente nel *corpus* è il seguente:

3. Napoli avrebbe avuto bisogno, avrebbe bisogno di una linea amministrativa e politica diversa, fatta di collaborazione con le altre istituzioni e con il Governo nazionale, e fatta anche di necessaria autonomia [ABAS16-F1].

Con questa affermazione si vuole implicare che l’amministrazione di Napoli non ha agito con correttezza nella gestione della città, poiché non ha collaborato con le istituzioni e non ha avuto l’autonomia necessaria.

Le implicature particolari richiedono invece, in aggiunta a ciò, la conoscenza di elementi specifici del contesto in cui avviene la conversazione e non sono quindi interpretabili in altre situazioni. Un esempio di implicatura particolare è il seguente:

4. scrivete il suo cognome perché è fondamentale evitare che qualcuno metta altro rispetto a quello che è il nostro diritto [FSTO13-C1].

In questo caso, come chiarito dalle esplicitazioni fornite dal *corpus*, l’implicazione è che ci sia il rischio che qualcuno falsifichi i voti.

4.2.2. Presupposizioni

Le presupposizioni sono un tipo di implicito il cui contenuto deve essere necessariamente accettato come vero da chi accetta come appropriata la produzione dell'enunciato.

Sono definiti attivatori (*preposition triggers*) gli elementi linguistici, che possono essere di natura lessicale o sintattica, da cui scaturisce la necessità di presupporre (Levinson 1983); negli esempi sottostanti, gli attivatori sono stati riportati in grassetto. Gli attivatori possono essere di natura lessicale ma non solo: anche elementi appartenenti al livello sintattico possono attivare presupposizioni. In Figura 1 sono riportati gli attivatori annotati nel *corpus* IMPAQTS.

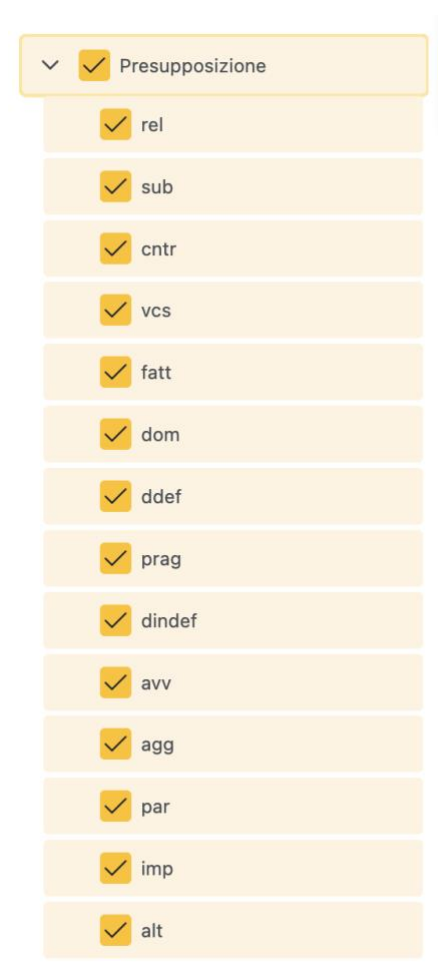


Figura 1. Tutti gli attivatori di presupposizione annotati nel corpus IMPAQTS.

Un caso tipico, annotato nel *corpus*, è la presupposizione di esistenza del referente delle descrizioni definite (Lombardi Vallauri 2019), come nell'esempio seguente, in cui viene data

per scontata l'esistenza di un 'sacrificio' da parte dell'uomo nelle coppie che fanno uso di fecondazione eterologa:

5. *questa donna possa avere la gioia della maternità, e l'uomo il **sacrificio** di avergliela consentita* [ABIO04-A1].

All'interno del *corpus*, le presupposizioni sono ulteriormente etichettate in sottogruppi a seconda della natura dell'elemento linguistico che fa da attivatore (Figura 1), come pronomi relativi, marcatori di subordinazione, avverbi, aggettivi, elementi di natura pragmatica, come nell'esempio sottostante:

6. *Questa è storia, **amici** di Siracusa, questa è storia vissuta* [RSCH13-P1].

In questo caso, è presupposto che il parlante sia in un rapporto di amicizia con i siracusani. Ancora, l'attivatore è rispettivamente un avverbio e un aggettivo nei casi sotto riportati:

7. *Anche perché un tale comportamento può determinare un rischio, ossia che i cittadini si allontanino **ancora di più** dalla vita politica* [DTON18-A2].

In questo caso, *ancora di più* presuppone che i cittadini si siano già effettivamente allontanati dalla politica.

8. *il **primo** tema su cui c'è una diversità grande come una casa* [MREN18-C1].

Con l'espressione "il primo tema", viene presupposto che ci siano altri argomenti sui quali è presente una tale diversità.

Anche la sintassi può generare una presupposizione:

9. *ma di misure concrete per contrastare la crisi e per venir fuori dalla disperazione **in cui** lo ha gettato questa classe politica* [RLOM13-A2].

In questo caso, la subordinata relativa vuole presupporre che il popolo sia stato effettivamente gettato in una condizione di disperazione per le azioni della classe politica.

Le presupposizioni rappresentano un fenomeno molto diffuso, anche per motivi legati all'economia degli sforzi nella produzione linguistica e alla necessità di evitare quanto è superfluo o ridondante (o anche, semplicemente, poco rilevante). Tuttavia, può avvenire che esse siano utilizzate per distrarre l'attenzione del destinatario da quanto viene taciuto proferendo la presupposizione; in questi casi, il contenuto presupposto potrebbe essere discutibile, infondato o altrimenti difficile da accettare, qualora venisse esplicitato.

4.2.3. *Topicalizzazioni*

Un'altra categoria di implicito annotata nel *corpus* sono le topicalizzazioni, forme linguistiche che agiscono al livello della struttura informativa dell'enunciato. Spesso, il soggetto o altre parti essenziali di un enunciato possono non essere espressi, perché disponibili all'interlocutore nel contesto linguistico (cotesto) o extralinguistico in cui avviene la comunicazione. Questo elemento, disponibile all'interlocutore e dunque passibile di omissione, è il Dato, mentre quanto viene aggiunto su di esso attraverso l'enunciato costituisce il Nuovo (Chafe 1987).

Il Dato tende a costituire nell'enunciato il *topic*, l'"argomento", mentre quanto è Nuovo costituisce di norma il *focus* e segue il *topic*¹³. Diversamente dalla presupposizione, che pure considera il proprio contenuto come noto, il *topic* non necessariamente richiama alle conoscenze generali condivise, ma può anche riferirsi a elementi contestuali, presenti nella memoria a breve termine dell'interlocutore (Lombardi Vallauri, 2019).

Nel momento in cui qualcosa viene presentato come *topic* di un enunciato, ciò può essere processato meno attentamente e meno accuratamente, poiché viene "dato per scontato": questo emerge da diversi studi, tra i quali citiamo Brédart & Modolo (1988), nei quali i partecipanti dimostrano di prestare maggiore attenzione (ed essere in grado di individuare più errori) nei segmenti di enunciato presentati come Nuovi e come *focus*.

Un esempio di topicalizzazione annotato all'interno del *corpus* IMPAQTS è il seguente:

10. *Votare a favore di un simile monstrum giuridico, non solo sarebbe ingiusto, ma costituirebbe soprattutto un gesto irresponsabile* [RLOM13-A2].

¹³ Diversa è la questione nel caso degli enunciati marcati con dislocazioni, che presentano spesso anche particolari contorni intonativi: il picco d'intensità che normalmente si registra nella parte finale dell'enunciato è anticipato e corrisponde all'elemento oggetto di *focus*.

In questo caso viene dato per accettato e condiviso con il pubblico che la proposta di legge costituisca un *monstrum* giuridico.

Un'altra somiglianza con le presupposizioni è dunque la capacità di presentare un contenuto come già noto e condiviso tra i partecipanti all'interazione, e Lombardi Vallauri (2019) considera questi due casi come «impliciti della responsabilità», in quanto la responsabilità del contenuto viene attribuita parzialmente anche al destinatario del proferimento, che è portato a considerare l'implicito come qualcosa che sia stato accettato in precedenza.

Non a caso, questi due tipi di impliciti possono essere attivati contemporaneamente, e quando il *topic* presenta una presupposizione, l'attenzione dell'interlocutore è minima nei confronti del contenuto sottaciuto.

E dunque «tornando alla lingua dei discorsi politici, essa è fitta di enunciati in cui proprio i contenuti più discutibili sono presentati come Topic, per dare ai destinatari l'impressione che trasmetterli non costituisca un'iniziativa dell'emittente» (Lombardi Vallauri 2019: 181).

4.2.4. *Vaghezza*

La vaghezza è in stretta relazione con l'estensione delle espressioni usate e con i suoi confini: riprendendo la definizione fornita da Tullio De Mauro (1982), Miriam Voghera scrive che la vaghezza «consiste nel fatto che i segni non instaurano un rapporto deterministico tra piano dell'espressione e del contenuto tale che i parlanti possano automaticamente decodificarli» (Voghera 2012: 343). La vaghezza richiama la ricerca della pertinenza, richiesta dalla terza massima conversazionale di Grice, anche chiamata Massima di Relazione (cfr. Capitolo 4.2.1); normalmente, espressioni vaghe rimangono comunque comprensibili grazie a elementi del contesto circostante che aiutano nell'identificazione del referente. Il fenomeno della vaghezza è infatti estremamente comune nel parlato e non sempre usato a fini persuasivi o manipolatori (Lombardi Vallauri 2019). Spesso essa è impiegata perché l'informazione specifica non è nota, o non vuole essere esplicitata perché ridondante e ampiamente accessibile nel contesto, o per motivi legati all'economia del discorso, e un'espressione vaga può addirittura funzionare meglio di una definita (Voghera 2012).

Per la vaghezza sono state annotate tre categorie: vaghezza semantica, vaghezza sintattica, e vaghezza derivante dall'utilizzo di una metafora.

Un esempio di vaghezza semantica è il seguente:

11. *Ci sono molti problemi, voi sapete io poi come la penso, ci sono molti problemi di partecipazione, ci sono problemi gravissimi nel nostro Paese che affronteremo* [AMUS01-P1].

In questo caso, non è chiarita la natura esatta dei problemi di cui si parla.

Un caso annotato di vaghezza sintattica è invece il seguente:

12. *non mi convince la distinzione tra una fase nella quale si discute di noi e una di quello, o è una nella quale si discute dell'opposizione* [AORL18-P1].

In questo esempio, l'autore della distinzione spiegata non è esplicitato.

Infine, un caso di vaghezza derivante da metafora presente nel *corpus* è quello qui riportato:

13. *aver messo in sicurezza le partecipate del comune di Napoli e lo stesso comune di Napoli, che se ascoltavamo certe volte le sirene di questi giorni, ci davano per falliti* [LDEM19-A1].

Di nuovo, il parlante non esplicita chi sia la “sirena” che reputa fallito il comune di Napoli.

È inoltre possibile e frequente che allo stesso segmento venga attribuita più di un'etichetta: il caso seguente unisce vaghezza semantica (poiché non è esplicitato chi debba interrogarsi) e una presupposizione derivante da una descrizione definita (l'esistenza di un 'solito tasto').

14. *Ma voglio specificare almeno per dare un'idea, perché forse qualcuno non sa e qualcuno si interrogherà prima di pigiare il solito tasto* [ABON15-A1].

Garassino et al. (2022) introducono una tassonomia relativa alla finalità dell'implicito veicolato dalle presupposizioni nel loro studio che ha riguardato i *tweet* come forma di comunicazione politica in Italia e nel Regno Unito. Come attivatori, sono identificati come più frequenti le descrizioni definite; i verbi aspettuali; alcuni operatori di *focus*; i cosiddetti predicati fattivi, che presuppongono il valore di verità positivo di quanto segue; le subordinate relative; le proposizioni subordinate in generale (Garassino et al. 2022). Le categorie proposte per le finalità delle presupposizioni sono le seguenti:

1. Critica (CRT);
2. *Stance-taking* (STK);

3. Contenuto informativo (INF);
4. Elogio, che può essere orientato nei confronti di chi parla (SPR) o rivolto verso qualcun altro (OPR);
5. Altro (OTH) per quanto non rientra nelle casistiche sopra elencate.

Lo studio di Garassino et al (2022) ha riguardato i *tweet* prodotti da soli quattro politici, al tempo capi di governo e opposizione nei due Paesi. L'annotazione è stata effettuata da tre annotatori e sono stati calcolati coefficienti di *inter-annotator agreement* come il kappa di Cohen (cfr. Capitolo 3, Nota 12), con risultati variabili per diversi annotatori e politici. Anche la prevalenza di diverse finalità sembra dipendere altamente dal singolo politico.

Nel caso del *corpus* IMPAQTS, l'annotazione è stata svolta indipendentemente da tre annotatori e infine rivista e approvata da un curatore.

4.3. Selezione dei dati utilizzati e trascrizione

Nel presente studio viene utilizzata una selezione di discorsi tratti dal *corpus* e in particolare dal *sotto-corpus* C, dedicato ai discorsi pronunciati dal 1994 al 2022; per tutti questi dati sono disponibili i file audio dei discorsi. La sottosezione più recente del *corpus* che utilizzeremo contiene più materiali dei *sotto-corpora* A e B¹⁴, compatibilmente con gli obiettivi del progetto IMPAQTS in ambito di disseminazione, tra cui «la diffusione nella società civile di uno spirito critico nei confronti degli intenti manipolatori dei politici» (Cominetti et al. 2022).

In particolare, sono stati analizzati 20 discorsi appartenenti alle categorie T (dichiarazione trasmessa), F (dichiarazioni in presenza) e A (discorsi pronunciati in sede di assemblea). Sono stati volutamente esclusi dalla selezione discorsi appartenenti alla tipologia dei nuovi media (N), quali dirette Facebook o Instagram, per le caratteristiche peculiari e uniche delle nuove tecnologie di comunicazione. Di questi nuovi mezzi di comunicazione non si vuole comunque sminuire la rilevanza: basti pensare all'importanza che i profili Facebook o X (precedentemente Twitter) rivestono nelle strategie di comunicazione di individui o enti di pubblica rilevanza.

Le variabili considerate ai fini dello studio sono le seguenti:

1. Genere: sono state considerate le due categorie *maschi* e *femmine*, le uniche due previste dallo schema di annotazione del *corpus* IMPAQTS.

¹⁴ La sezione C contiene infatti circa sei volte il materiale della prima sezione, dedicata ai discorsi dalla nascita della Repubblica fino al 1973, e circa tre volte il materiale della seconda, che contiene i discorsi pronunciati dal 1973 al 1993.

2. Fascia d'età al momento della produzione del discorso: sono state identificate tre fasce d'età (meno di 40 anni; dai 41 ai 60 anni; dai 61 anni).
3. Orientamento politico: sono state identificate tre categorie (sinistra e centro-sinistra; centro; destra e centro-destra).
4. Provenienza geografica: sono state considerate tre macroregioni (Nord; Centro; Sud e isole).

Si è cercato di ottenere un campione bilanciato rispetto a tutte queste variabili, nei limiti delle condizioni nelle quali il *corpus* IMPAQTS si trovava al momento della consultazione (per la maggior parte dei discorsi, non era ancora presente il *file* video corrispondente).

Per la trascrizione dei discorsi è stato utilizzato il riconoscitore automatico *whisper.ai*, uno strumento *open source* di ASR (*Automatic Speech Recognition*) messo a disposizione da OpenAI¹⁵. Si tratta di uno strumento addestrato su 680.000 ore di parlato, di cui circa un terzo in lingue diverse dall'inglese, ed è in grado, oltre che di fornire trascrizioni, di riconoscere la lingua del materiale *audio* e di produrre una traduzione simultanea del contenuto.

Il programma di riconoscimento vocale *whisper.ai* sfrutta una rete neurale basata sull'architettura dei *transformer*. In *machine learning* si intende con rete neurale (*neural network*) un oggetto di calcolo in grado di effettuare *task* molto complessi nell'ambito della linguistica computazionale e non solo. Le reti neurali sono in grado di effettuare in pochi minuti operazioni complesse, come la trascrizione di numerose ore di parlato, che manualmente risulterebbero molto dispendiose in termini di tempo e di risorse. Il loro nome deriva da un'ipotesi sul funzionamento del cervello, che esse dovrebbero replicare. Le reti neurali sono in grado di generare una rappresentazione dei dati di *input* in modo molto accurato, grazie alla distribuzione del calcolo in diversi nodi, suddivisi in livelli. Dei dati di *input* sono in grado di selezionare autonomamente le *features* rilevanti, e stimarne i pesi, che esprimono l'importanza di ciascun elemento, senza che vi sia necessità di un intervento esterno. Alla base del loro funzionamento vi è il procedimento di regressione logistica. Le reti neurali con architettura a *transformer*, introdotte nel 2015, segnano un progresso rispetto alle reti delle generazioni precedenti grazie all'introduzione di un meccanismo di *attenzione* (Vaswani et al. 2017) che permette loro di processare dati in maniera non sequenziale, sfruttando i rapporti reciproci tra gli elementi dell'*input* (Jaderberg et al. 2015). I pesi sono dunque stimati simultaneamente e in diversi *layer*, permettendo di mappare accuratamente anche il contesto di ciascun *token*.

¹⁵ <https://github.com/openai/whisper>

Il materiale audio, diviso in segmenti della durata di 30 secondi ciascuno, deve essere inizialmente convertito sulla scala Mel, che fa uso dei Mel Frequency Cepstral Coefficients (MFCC). La scala Mel rispecchia più fedelmente il modo in cui i suoni del parlato sono percepiti dall'orecchio umano, e dunque in maniera non lineare, con una minore capacità di discernere il segnale a frequenze alte: mentre tra 0 e 1000 Hz gli intervalli tra le bande si mantengono regolari, dopo i 1000 Hz essi si allargano. Per estrarre i coefficienti MFCC, il suono viene dapprima segmentato in finestre di 25 millisecondi sulle quali viene operata una trasformata di Fourier¹⁶ a corto termine, che è in grado di dividerle in bande di frequenza. Avendo così ottenuto le bande di frequenza, attraverso funzioni, è possibile mapparle sui filtri Mel dei quali viene calcolato il logaritmo. Infine, ai logaritmi delle bande è possibile applicare una funzione, chiamata *Discrete Cosine Transform*, la quale rende i coefficienti cepstrali (Sahidullah & Saha 2012). I coefficienti cepstrali possono essere in numero eccessivo, o non fornire informazioni rilevanti sul segnale: esistono tecniche per la loro riduzione, per ridurre la dimensione dello spazio vettoriale richiesto per la rappresentazione del segnale generata dall'*encoder*. L'input così elaborato e convertito sulla scala log-Mel viene trasmesso a un *encoder*, a sua volta una rete neurale che genera, a partire dall'input, una sua rappresentazione entro uno spazio vettoriale. La componente di decodifica, il *decoder*, elabora un *output* a partire da queste rappresentazioni, cercando di prevedere la trascrizione corretta del segnale.

Le trascrizioni ottenute mediante *whisper.ai* sono state poi corrette manualmente, ricorrendo ove necessario all'ascolto dei *file* audio in questione; i numeri sono stati trascritti per esteso. Nonostante l'accuratezza elevata, alcuni elementi si sono dimostrati critici per il *software*, tra i quali:

- La grafia di parole quali *amministrazione* e derivati, spesso trascritte come *amministrazione*; *onorata* per *onorata*; *prejudizio* per *pregiudizio*; *elettorale*, trascritto come *electorale*, *ritemmetica* per *aritmetica*; *spreciudicato* per *spregiudicato*, *sottrata* per *sottratta*.
- I confini di parola, con universioni o scansioni errate: *su le* per *sulle*; *piuttorlerabile* per *più tollerabile*; *si è ad ricercare* per *sia da ricercare*; *familiario* per *familiare o*; *la dirittura* per *addirittura*; *d'annoso* per *dannoso*, *in propriamente* per *impropriamente*, *la veste* per *l'aveste*.

¹⁶ Una trasformata di Fourier è un procedimento matematico che consente di scomporre segnali acustici complessi nelle loro componenti semplici, le cui frequenze sono chiamate armoniche. Di queste armoniche, la prima è chiamata frequenza fondamentale.

- Intere parole non comprese: *li ha mutiata per nell'ammucchiata, bomgo verde per buongoverno.*
- Gli italiani regionali, infine, si sono dimostrati particolarmente problematici, e ancora di più la presenza di inserti dialettali, come nel discorso pronunciato da Sebastiano (Nello) Musumeci: *comu finisci si cunta, su pagate per fare chistu.*

Sempre dal *corpus* è stato possibile ricavare anche gli impliciti contenuti nei discorsi selezionati (Tabella 1); si noti che è possibile che allo stesso segmento sia stata attribuita più di un'etichetta, e dunque il totale non rispecchia il numero di segmenti annotati nei discorsi presi in considerazione.

Implicature	Presupposizioni	Topicalizzazioni	Vaghezza
97	160	37	66

Tabella 1. Impliciti annotati nei discorsi selezionati.

Per ogni parlante sono stati isolati 3 enunciati contenenti presupposizioni e 3 enunciati che ne erano privi, come *dataset* di controllo, dei quali sono stati estratti *speech rate* e *articulation rate* con le modalità descritte successivamente.

4.4. Metodi per il calcolo dello *speech rate*

Lo *speech rate* è definito propriamente come il rapporto tra il numero di sillabe contenute in un segmento di parlato e la durata del segmento stesso in secondi; sono possibili altre misure, che calcolano il numero di foni, di fonemi, o di parole al secondo. Una variante di questa misura è il cosiddetto *articulation rate*, che, mantenendo la stessa formula, sottrae alla durata del segmento in secondi ogni pausa rilevata, rendendo dunque la velocità di eloquio unicamente quando la produzione è realmente in atto. Derivata da queste misure è l'*average syllable duration*, definita come la durata media di una sillaba in secondi.

Nel presente elaborato il calcolo dello *speech rate* è stato effettuato sfruttando programmi ampiamente diffusi e utilizzati specificamente per analisi fonetiche. Esistono metodi, come lo script di De Jong & Wempe (2009) per il programma Praat, che non richiedono trascrizioni dei *file* audio: il calcolo dello *speech rate* si basa sul numero di sillabe, determinato mediante l'identificazione automatica dei nuclei sillabici. L'estrazione dell'intensità, per finestre di brevissima durata, permette di identificare ogni picco come un potenziale nucleo sillabico. Viene poi calcolata la media dell'intensità per la durata del segmento, per escludere picchi

estremi, e sono considerati nuclei sillabici solo i picchi preceduti da un calo di intensità e che presentano sonorità. La velocità d'eloquio viene quindi calcolata in sillabe al secondo. In una versione successiva dello stesso *script*, sono considerate anche le informazioni legate ai silenzi, identificati mediante l'impostazione di una soglia, in dB, al di sotto della quale il programma è in grado di identificare le pause. Questo permette l'estrazione dell'*articulation rate*, ottenuto sottraendo i millisecondi di silenzio alla durata totale della produzione. La differenza tra le due misure può rivelarsi significativa, per esempio, nei casi di parlato lento e che presenta molte pause.

Altre metodologie richiedono una trascrizione del materiale audio di cui si intende ottenere la velocità d'eloquio: in particolare, lo *script* utilizzato per il presente elaborato richiede una tipologia di *file*, chiamata TextGrid, normalmente utilizzata dal programma Praat (Boersma, 2001; Boersma & Weenink, 2022)¹⁷. A partire dal materiale audio e dalle trascrizioni ottenute con *whisper.ai*, opportunamente rivedute e corrette, sono stati creati dei *file* TextGrid attraverso la piattaforma WebMaus (Schiel, 1999; Kisler & Reichel, 2017)¹⁸. Si tratta di una piattaforma gratuita messa a disposizione e gestita dall'Università Ludwig-Maximilian di Monaco di Baviera, nell'ambito del *Bavarian Archive for Speech Signals*, ospitato dall'*Institut für Phonetik und Sprachverarbeitung*.

Il programma è in grado di segmentare i file audio e allineare ogni segmento alla trascrizione corrispondente, qualora essa venga fornita dall'utente. I file prevedono diversi livelli di annotazione, a seconda delle esigenze: è in grado di rendere un *tier* di annotazione che associa a ogni parola, nella sua trascrizione ortografica, la trascrizione fonetica corrispondente mediante IPA o mediante l'alfabeto X-SAMPA (*Speech Assessment Methods Phonetic Alphabet*), che utilizza solo caratteri appartenenti al sistema ASCII, e un *tier* che corrisponde al singolo fono, sempre trascritto in X-SAMPA. Si segnala che il programma è in grado anche, attraverso un riconoscitore automatico, di generare la stessa trascrizione a partire dal solo materiale audio.

L'*output* di questo processo è un *file* di tipo TextGrid, analizzabile con il programma Praat.

Per il calcolo dello *speech rate* in sillabe al secondo è stata sfruttata la *pipeline*, offerta da WebMaus, che effettua, in sequenza, i servizi G2P, Maus e Pho2Syl, che ha permesso di ottenere un *tier* che associa a ogni sillaba la relativa trascrizione.

¹⁷ <https://www.fon.hum.uva.nl/praat/>

¹⁸ <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface>

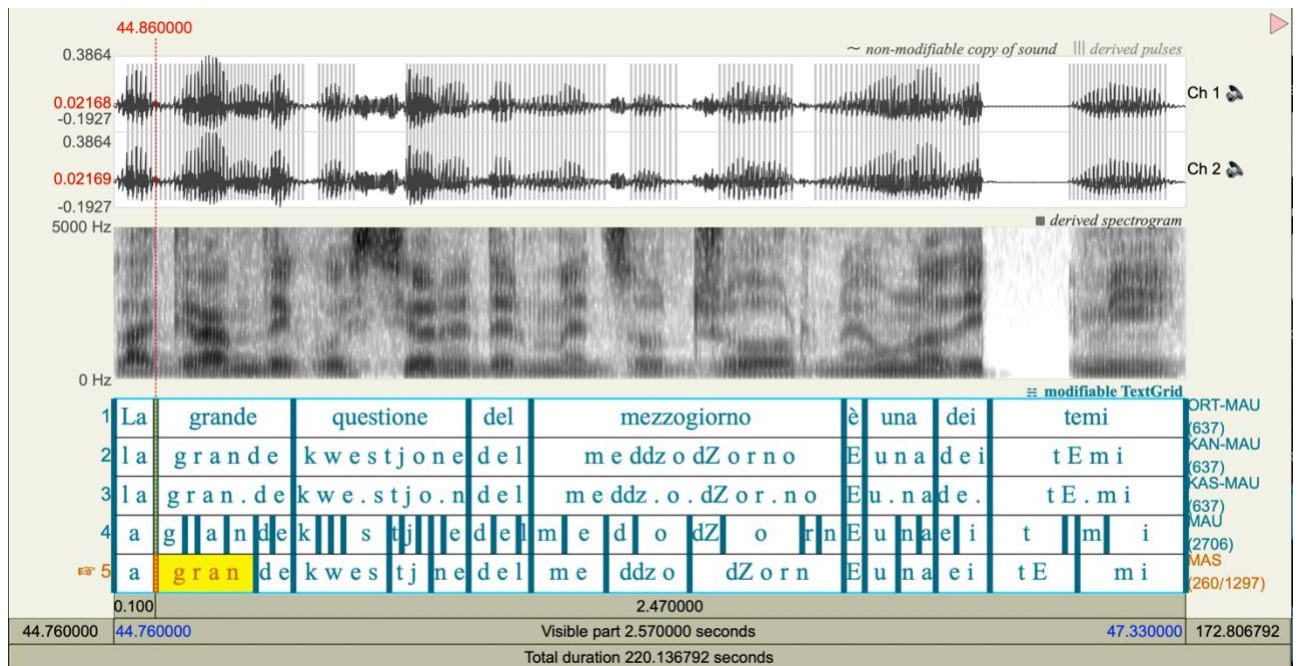


Figura 2. Una schermata del programma Praat, nella quale si possono vedere il sonogramma del segnale audio e i cinque livelli di annotazione del file TextGrid ottenuto mediante la pipeline fornita da WebMaus.

Infine, partendo dai *file* audio e dai TextGrid così ottenuti, i dati relativi alle caratteristiche temporali del parlato sono stati calcolati mediante lo *script* tratto da Combei & Didoni (2024) in grado di estrarre *speech rate*, *articulation rate* e altre informazioni a partire dagli intervalli annotati e dai relativi tempi di inizio e di fine. Gli intervalli di silenzio sono contati e considerati come pause ai fini del calcolo di queste misure: in particolare, la loro durata viene sottratta alla durata totale del *file* audio per il calcolo dell'*articulation rate*, che considera unicamente il tempo effettivo di fonazione.

Per questo motivo, è possibile che i valori di *speech rate* e *articulation rate* siano abbastanza simili tra di loro, ma anche che differiscano in maniera significativa. Nessun discorso presente nel dataset è privo di pause: la durata dei silenzi varia da 4,47 secondi a ben 202,23 secondi. Nel primo caso, lo *speech rate* è di 6,03 sillabe al secondo mentre l'*articulation rate* è di poco superiore (6,11 sillabe al secondo). Nel secondo caso la differenza è molto più marcata: da uno *speech rate* di 4,22 sillabe al secondo a un *articulation rate* di 6,00 sillabe al secondo.

Il codice sfrutta la libreria *praatio*, che fornisce la possibilità di aprire e leggere *file* TextGrid e dà accesso ad altre funzionalità utili per lavorare con *file* audio annotati. Si è scelto di considerare *speech rate* e *articulation rate* in sillabe al secondo.

Per quanto riguarda gli enunciati contenenti presupposizioni, a partire dai *file* così processati è stato possibile isolare segmenti di audio e di *file* TextGrid tramite il programma Praat; di

questi segmenti sono stati calcolati *speech rate* e *articulation rate* secondo lo stesso procedimento descritto in precedenza.

4.5. Descrizione statistica dei dati

A partire dai dati ottenuti è stata svolta un'analisi statistica utilizzando il programma RStudio, che permette di realizzare *script* nel linguaggio di programmazione R, creato *ad hoc* per effettuare analisi descrittive, modelli lineari e rappresentazioni grafiche dei dati. Inizialmente vengono fornite alcune statistiche descrittive dei risultati ottenuti, quali la media, la mediana e la deviazione standard. Successivamente è stato creato un modello di regressione lineare in R attraverso l'ambiente di sviluppo RStudio e i pacchetti contenuti in *tidyverse* (Wickham et al. 2019) per R, come *ggplot2* (Wickham et al. 2016).

Per quanto riguarda le statistiche descrittive utilizzate, si intende con minimo e massimo,

```
1 dat <- metrics_ar
2 str(dat)
3 min(dat$articulation_rate)
4 max(dat$articulation_rate)
5 range(dat$articulation_rate)
6 max(dat$articulation_rate) - min(dat$articulation_rate)
7 mean(dat$articulation_rate)
8 median(dat$articulation_rate)
9 sd(dat$articulation_rate)
10 aggregate(articulation_rate~gender, data=dat, summary)
11 aggregate(articulation_rate~age_group, data=dat, summary)
12 aggregate(articulation_rate~orientamento, data=dat, summary)
13 aggregate(articulation_rate~provenienza, data=dat, summary)
```

Figura 3. Alcune delle funzioni utilizzate per l'estrazione delle statistiche descrittive.

rispettivamente, il valore minimo e il valore massimo di una variabile (in questo caso, *speech rate* e *articulation rate*), la cui differenza rende il *range* nel quale sono collocate tutte le misure rilevate. Questi valori sono ottenibili con le funzioni *max()* e *min()* in R, o, contemporaneamente, con la funzione *range()* (Winter 2020).

La media, ottenuta attraverso la funzione *mean()*, è la media aritmetica dei valori. La mediana, ottenuta con la funzione *median()*, è il valore centrale della distribuzione. Si ricava ordinando tutti i valori interessati in ordine crescente: la mediana è il valore centrale, se le osservazioni sono in numero dispari, o la media aritmetica dei due valori centrali se sono in numero pari. La mediana, rispetto alla media, è meno soggetta a variazioni consistenti dovute a valori estremi (Winter 2020).

È possibile visualizzare questi valori isolatamente per diverse categorie (come genere, fasce d'età, diverse regioni di provenienza) mediante la funzione *aggregate()* di R.

Infine, la deviazione standard, calcolata tramite la funzione *sd()*, è un valore che rende conto della variabilità dei dati. Si ottiene dalla radice quadrata dello scarto quadratico medio, cioè della sommatoria dei quadrati delle differenze tra ciascun valore (x_i) e la media aritmetica di tutti i valori (\bar{x}), divisa per il numero di osservazioni (N) a cui viene sottratto 1 quando si tratta di un campione, e non dell'intero universo.

La formula per il calcolo della deviazione standard è la seguente:

$$SD = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N - 1}}$$

Tutte le funzioni utilizzate sono presenti in R.

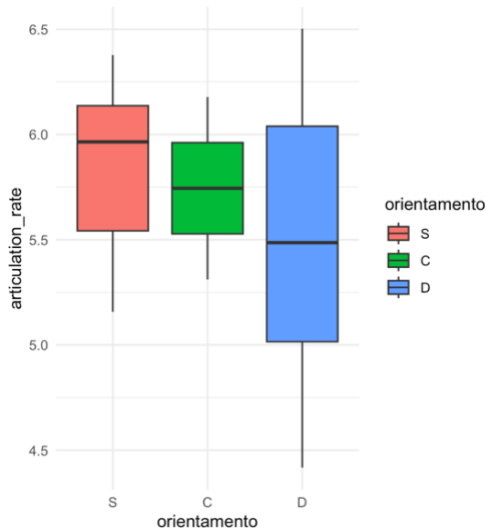


Figura 4. Esempio di boxplot che rappresenta i valori di articulation rate per orientamento politico.

Di queste misure viene fornita una rappresentazione grafica attraverso *boxplot*, ottenuti tramite la libreria *ggplot2* per R.

Lo stesso procedimento che consente di trovare la mediana permette di calcolare i quartili, che corrispondono ai valori che si trovano al 25%, 50%, 75% e 100% dei dati disposti in ordine crescente (dunque, metà dei dati si trovano al di sotto del secondo quartile, e metà dei dati si trovano al di sopra di esso). Il rettangolo che costituisce il *box* del *boxplot* rappresenta i dati contenuti tra il primo e il terzo quartile, mentre la linea nera al suo interno rappresenta il secondo quartile, che corrisponde alla mediana. Le linee al di sopra e al di sotto del *box*, chiamate anche *whiskers*, hanno una lunghezza pari a 1,5

volte la differenza dei valori del terzo e del primo quartile (Winter 2020). Se i valori così ottenuti sono superiori o inferiori rispetto al massimo e al minimo compresi nel *dataset*, la lunghezza dei *whiskers* si arresta al massimo o al minimo dei valori presenti.

Ad esempio, la Figura 4 rappresenta l'*articulation rate* dei discorsi analizzati nello studio suddivisi per appartenenza politica del parlante. Nel caso dei parlanti riconducibili a

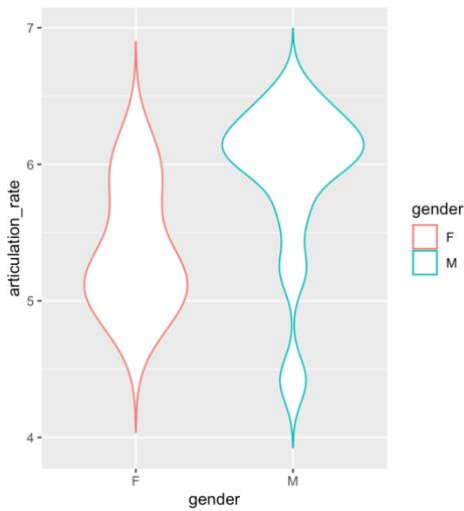


Figura 5. Esempio di violin plot che rappresenta la distribuzione dei valori di *articulation rate* divisi per genere.

schieramenti di destra o centro-destra (in blu), il valore del primo quartile corrisponde a 5,02 sillabe al secondo, il valore del terzo quartile corrisponde a 6,04 sillabe al secondo, mentre la mediana, ovvero il secondo quartile, corrisponde a 5,49 sillabe al secondo. La differenza tra terzo e primo quartile è di 1,02 sillabe al secondo, e la lunghezza dei *whiskers* sarebbe dunque di 1,53 sillabe al secondo, ma il *dataset* non presenta valori di *articulation rate* così bassi o così alti: i *whiskers* si fermano dunque al minimo e al massimo dei dati presenti, e cioè, rispettivamente, 4,42 e 6,50 sillabe al secondo.

Un diagramma a violino, come ad esempio quello illustrato nella Figura 5 è così chiamato per la sua forma, che ricorda quella della cassa di risonanza dello strumento musicale, e rappresenta graficamente la distribuzione di probabilità dei dati. Rispetto ai *boxplot* fornisce una quantità maggiore di informazioni: l'ampiezza del violino in ciascuna sua parte riflette la frequenza dei dati nella regione corrispondente.

4.6. Modello di regressione lineare

La regressione lineare è una tecnica statistica che ricerca una relazione di tipo lineare tra una variabile cosiddetta dipendente (y) e una o più variabili indipendenti (x), considerate come predittori della variabile dipendente.

$$y = b_0 + b_1x$$

Una retta è descritta matematicamente attraverso la sua pendenza e la sua intercetta: la

```
R 4.4.1 ~-|
> ar_model <- lm(articulation_rate ~ gender + age_group + orientamento + p
provenienza, data = metrics_ar)
> summary(ar_model)

Call:
lm(formula = articulation_rate ~ gender + age_group + orientamento +
provenienza, data = metrics_ar)

Residuals:
    Min       1Q   Median       3Q      Max
-0.74319 -0.16503 -0.01515  0.16321  0.82100

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.21255    0.41595  12.532 2.98e-08 ***
genderM      0.77437    0.25609   3.024  0.0106 *
age_groupadulto -0.47797    0.34727  -1.376  0.1939
age_groupanziano -0.88921    0.47535  -1.871  0.0860 .
orientamentoC  0.77331    0.49011   1.578  0.1406
orientamentoD  0.06256    0.36355   0.172  0.8662
provenienzacentro  0.52170    0.31006   1.683  0.1183
provenienza nord  0.59969    0.36495   1.643  0.1263
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4316 on 12 degrees of freedom
Multiple R-squared:  0.6601, Adjusted R-squared:  0.4618
F-statistic: 3.329 on 7 and 12 DF, p-value: 0.03273
```

Figura 6. Interfaccia di RStudio e output della regressione lineare.

pendenza è data dal mutamento del valore di y al cambiare del valore di x; l'intercetta è il valore previsto per y quando x = 0.

A partire dall'equazione fondamentale della retta, il procedimento di regressione mira a stimare il migliore coefficiente per l'intercetta (o *bias*, qui b_0) e per le variabili indipendenti (b_1 ed eventuali altre), che rendano il modello il più accurato possibile a partire dai dati presenti nel *dataset* (Winter 2020).

Nel presente elaborato, è stata effettuata una regressione lineare multipla, dal momento che le variabili indipendenti sono più di una (genere, fascia d'età, orientamento politico, provenienza geografica). Tutte queste variabili sono di natura categoriale: si tratta di etichette e non valori numerici.

Nella regressione lineare multipla, viene analizzato l'effetto di tutte le variabili indipendenti contemporaneamente sulla variabile dipendente.

In R, è possibile effettuare una regressione lineare con la formula $lm(y \sim x_1 + x_2 + \dots + x_n)$, dove y rappresenta la variabile dipendente e $x_1 \dots x_n$ rappresentano le variabili indipendenti, o predittori. I valori di riferimento impostati sono i seguenti:

- genere: femminile;
- fascia d'età: giovane (40 anni o meno);
- orientamento politico: sinistra o centro-sinistra;
- provenienza geografica: Sud Italia o isole.

Un modello di regressione lineare non ha potere predittivo al di fuori del *range* di dati presenti nel *dataset* e di essi non fornisce una descrizione perfettamente accurata: per questo è importante il calcolo dei residui, ovvero le distanze verticali tra i dati presenti nel *dataset* e la loro rappresentazione lineare. I residui forniscono dunque una buona approssimazione della capacità del modello di spiegare la variazione entro i dati e il *pattern* che essa segue. A partire dai residui è quindi possibile ottenere due misure, R^2 e *adjusted* R^2 , che esprimono la bontà dell'adattamento, ovvero quanto il modello ottenuto è in grado di spiegare la varianza dei dati (Winter 2020).

R^2 viene calcolato a partire dalla somma dei residui del modello di regressione lineare, elevati al quadrato (SSE, *sum of squared errors*), e dalla somma dei quadrati dei residui di un modello chiamato *null model*, che prevede solo l'intercetta, senza coefficienti per la pendenza. R^2 è quindi calcolato come segue:

$$R^2 = 1 - \frac{SSE_{model}}{SSE_{null}}$$

In questo modo, se la somma dei residui del modello è molto alta, l'indice ottenuto è inferiore. I valori di R^2 e *adjusted* R^2 sono compresi tra 0 e 1; possono anche essere considerati indici della forza dell'associazione tra due variabili (Winter 2020).

Nel caso della regressione multipla, l'indice più accurato è *adjusted* R^2 , che è sempre inferiore a R^2 semplice in quanto include un termine che riduce l'indice per evitare il fenomeno chiamato *overfitting*, che si verifica quando un modello si avvicina troppo ai *pattern* di un *dataset* (Winter 2020). Questo rischio è particolarmente importante da considerare in presenza di più di una variabile indipendente.

La formula per il calcolo di *adjusted* R^2 è la seguente, dove N è il numero di istanze presenti nel *dataset*, e k è il numero dei parametri presenti nel modello:

$$R_{adj}^2 = \frac{(1 - R^2)(N - 1)}{N - k - 1}$$

Infine, la funzione *lm()* rende anche i valori di *p-value*, derivanti da un procedimento definito *t test*. Uno dei vantaggi di questo test statistico è che tiene in considerazione la quantità di istanze presenti nel *dataset* (*sample size*). Questa misura viene utilizzata per calcolare l'errore standard (*standard error*, SE), che a sua volta figura nella formula con la quale viene calcolato *t*.

Il *t test* è in grado di rivelare se una variabile è statisticamente significativa, ovvero, nel caso presente, se ha un forte legame con la variabile dipendente e apporta un contributo significativo all'accuratezza del modello. Il test paragona i coefficienti stimati dal modello per le variabili indipendenti a quelli che si verificherebbero qualora fosse vera l'ipotesi nulla, e cioè, in questo caso, che non vi sia alcun legame tra la variabile dipendente e i suoi predittori.

A partire dal valore di *t* è possibile calcolare un valore, chiamato *p-value*, che esprime la probabilità di ottenere i dati a nostra disposizione se l'ipotesi nulla fosse quella vera (e cioè, nel nostro caso, se non esistesse alcun legame tra la velocità di eloquio e di articolazione e le variabili sociolinguistiche considerate).

È comunemente accettata una soglia di 0,05 come *p-value* per rifiutare l'ipotesi nulla: una soglia del genere significa che solo nel 5% dei casi, se l'ipotesi nulla fosse giusta, otterremmo una distribuzione come quella presente nel *dataset*.

```
11 model_summary <- summary(model)
12 coefficients <- model_summary$coefficients
13 summary_table <- data.frame(
14   term = rownames(coefficients),
15   estimate = coefficients[, "Estimate"],
16   std.error = coefficients[, "Std. Error"]
17 )
18 summary_table <- summary_table[summary_table$term != "(Intercept)", ]
19 coefficients_plot <- ggplot(summary_table, aes(x = term, y = estimate)) +
20   geom_col(fill = "skyblue", color = "black") +
21   geom_errorbar(aes(ymin = estimate - std.error, ymax = estimate + std.error), width = 0.2, color = "black") +
22   coord_flip() +
23   labs(x = "Predictor", y = "Coefficient Estimate",
24        title = "Coefficients for the Model",
25        subtitle = "Speech rate",
26        caption = "Error bars represent standard errors") +
27   theme_minimal()
```

Figura 7. Ambiente di lavoro di RStudio e codice che rende il grafico per la stima dei valori dei coefficienti, con i relativi errori standard.

Infine, grazie a uno *script* nell'ambiente di lavoro RStudio (Figura 7), è stato generato un grafico che rappresenta la stima dei coefficienti, con i relativi errori standard per ciascun coefficiente. Questa rappresentazione fornisce un'approssimazione di come ciascun valore delle variabili influenza la variabile dipendente, anche se non mostra informazioni relative al *p-value* legato a ciascun valore e, dunque, alla bontà del modello e al contributo apportato dalle singole variabili.

4.7. Analisi dei dati: enunciati contenenti impliciti

Sono stati isolati 33 segmenti di parlato contenenti presupposizioni e 33 segmenti che ne sono privi (*dataset* di controllo), tratti dai discorsi analizzati in precedenza (in particolare, sono stati isolati sei enunciati per parlante, tre contenenti presupposizioni e tre privi di impliciti). Questi dati sono stati preparati secondo il procedimento descritto nei paragrafi precedenti; le trascrizioni degli enunciati sono riportate in Appendice.

In questo caso, come variabile dipendente è stato considerato unicamente l'*articulation rate*, che non tiene conto di eventuali pause presenti nella fonazione. Esse possono invece apportare un contributo significativo nel calcolo dello *speech rate* semplice, che è comunque correlato alla velocità di articolazione.

Come variabili indipendenti, o predittori, sono stati utilizzati i seguenti:

1. presenza o assenza di impliciti;

2. genere;
3. fascia d'età;
4. orientamento politico.

È possibile visualizzare alcune statistiche descrittive per gli incroci di due o più variabili tramite la funzione *aggregate()* di R: otteniamo così minimo, massimo, media, mediana e quartili, per esempio, delle produzioni contenenti presupposizioni pronunciate da parlanti di genere femminile, isolatamente dalle altre categorie.

A partire da questi dati, e con queste variabili, è stato eseguito un nuovo modello di regressione lineare.

5. Risultati e discussione

Nel presente capitolo sono riportati i risultati delle analisi statistiche effettuate, come descritto nel Capitolo 4, relativamente a *speech rate* (velocità di eloquio) e *articulation rate* (velocità di articolazione), rispetto alle variabili sociolinguistiche considerate (genere, orientamento politico, provenienza geografica, fascia d'età), tramite l'ambiente di sviluppo RStudio.

Per la velocità di eloquio e di articolazione, vengono dapprima forniti i risultati delle analisi statistiche descrittive relativi all'intero *dataset* e alle sue sottosezioni, accompagnate da informazioni relative al numero e alla durata delle pause presenti nei discorsi; successivamente, vengono illustrati i risultati dei modelli di regressione lineare.

Infine, sono forniti questi stessi valori per la velocità di eloquio e di articolazione degli enunciati contenenti presupposizioni e di quelli che ne sono privi, seguiti dai risultati di alcuni modelli di regressione che possano spiegare l'effetto della presenza o assenza di impliciti su *speech rate* e *articulation rate*, con particolare attenzione rivolta alle pause. Si ricorda che, come indicato nel Capitolo 2, dedicato alla velocità di eloquio, il calcolo dello *speech rate* considera le pause che possono essere influenzate da fattori esterni, quindi indipendenti dalla volontà del parlante. Al contrario, l'*articulation rate* offre una rappresentazione più accurata dei tempi di fonazione effettivi e delle caratteristiche temporali intrinseche della produzione orale.

5.1. Lunghezza e distribuzione delle pause

Il numero di pause segmentate all'interno di ciascun *file* varia tra 26 e 496 unità; la durata totale delle pause entro la medesima registrazione varia tra 4,47 secondi e 202,23 secondi. Questi numeri da soli non forniscono informazioni utili: è importante comunque ricordare che le registrazioni avevano diverse lunghezze (da un minimo di 107,4 secondi, meno di un minuto e mezzo, a un massimo di 1201,4 secondi, poco più di 20 minuti).

La durata totale delle pause e il loro numero saranno dunque influenzati dalla lunghezza totale del discorso. Per questo motivo i dati sono stati normalizzati, calcolando il rapporto tra la durata totale di ciascun *file* audio e la durata totale delle pause in esso contenute, ottenuta grazie alla segmentazione e al *forced alignment* con lo strumento WebMaus (Kisler et al., 2017); di questi valori sono riportati nella Tabella 2 minimo, massimo, media, deviazione standard e mediana, divisi per ciascun livello delle variabili sociolinguistiche considerate.

Rapporto tra la durata totale del discorso e delle pause	Minimo	Massimo	Media	Deviazione standard	Mediana
Genere: femminile	0,01	0,17	0,11	0,53	0,10
Genere: maschile	0,08	0,30	0,15	0,06	0,14
Fascia d'età: < 41 anni	0,01	0,11	0,08	0,04	0,08
Fascia d'età tra: 41 e 60 anni	0,06	0,30	0,16	0,07	0,15
Fascia d'età: > 60 anni	0,08	0,19	0,14	0,04	0,16
Orientamento politico: sinistra e centro-sinistra	0,01	0,17	0,09	0,05	0,09
Orientamento politico: centro	0,10	0,19	0,15	0,06	0,15
Orientamento politico: destra e centro-destra	0,08	0,30	0,16	0,06	0,16
Provenienza geografica: Nord	0,01	0,30	0,15	0,10	0,16
Provenienza geografica: Centro	0,06	0,17	0,12	0,04	0,10
Provenienza geografica: Sud e isole	0,08	0,19	0,13	0,04	0,14

Tabella 2. Minimo, massimo, media, deviazione standard e mediana del rapporto tra la durata delle pause e la durata totale del discorso, per categoria.

Le pause sono quindi particolarmente presenti nei discorsi dei politici di destra o centro-destra (con un rapporto medio di pause sulla durata totale di 0,16, con deviazione standard 0,06), provenienti dal Nord Italia (rapporto medio di 0,15, con deviazione standard di 0,10), di genere maschile (rapporto medio di 0,15, con deviazione standard di 0,06). Nel caso delle parlanti di genere femminile, il rapporto medio non è particolarmente alto (0,11), ma la deviazione standard è notevolmente maggiore rispetto agli altri gruppi (0,53): è dunque presente un'ampia variabilità all'interno dei discorsi pronunciati dalle donne. I giovani, con un rapporto

medio di 0,08 (con deviazione standard di 0,04), fanno meno uso di pause rispetto a entrambe le fasce d'età successive: tra i 41 e i 60 anni il rapporto medio è di 0,16 (con deviazione standard di 0,07), e sopra i 60 il rapporto medio è di 0,14 (con deviazione standard di 0,04). Non vi sono differenze significative in termini di pause tra i parlanti provenienti dal Centro (rapporto medio di 0,12, con deviazione standard di 0,04) e dal Sud Italia (rapporto medio di 0,13, con deviazione standard di 0,04).

Per le categorie che presentano un rapporto elevato di pause rispetto al totale della durata della registrazione, è attesa dunque una differenza importante tra il valore di *speech rate* e di *articulation rate*.

5.2. Velocità di eloquio

Lo *speech rate* minimo contenuto nel dataset è 4,02 sillabe al secondo; il massimo è 6,03 sillabe al secondo, con un range di 2,01 sillabe al secondo. La media è di 4,91 sillabe al secondo. La mediana dell'intero dataset è di 4,88 sillabe al secondo; la deviazione standard è 0,60.

Questi risultati sono leggermente inferiori rispetto a quelli ottenuti da studi precedenti, che miravano a ottenere dati medi di *speech rate* per l'italiano. Si veda per esempio Rodero (2012), che, per il parlato radiofonico italiano (estratto a partire da programmi andati in onda sulle reti RAI), riscontra dei valori medi di 6,41 sillabe al secondo; Masini et al. (2024) riportano invece 6,02 sillabe al secondo. Questa discrepanza potrebbe essere dovuta alla presenza massiccia di pause, almeno per alcune categorie di parlanti, come riportato nella Tabella 2. È anche vero che i contenuti di questi discorsi presentano una complessità maggiore rispetto a buona parte del parlato spontaneo e radiofonico, e a differenza di quest'ultimo il discorso politico coinvolge più canali, incluso quello visivo: si vedano per esempio Maricchiolo et al. (2014) per un'analisi del ruolo dei movimenti corporei nel parlato politico, e Poggi et al. (2010) per uno studio sulla gestualità e sui ruoli rivestiti dall'atto dell'annuire. Tutti questi fattori – le pause e i loro possibili usi, la presenza di elementi paralinguistici in contemporanea alla fonazione – possono spiegare perché il campione di discorsi utilizzato nel presente lavoro si rivela globalmente più lento rispetto ai risultati riportati in letteratura per l'italiano.

La Tabella 3 riporta minimo, massimo, media e mediana dello *speech rate* suddiviso per ogni livello delle variabili considerate (genere, fascia d'età, orientamento politico, provenienza geografica).

Speech rate (sillabe al secondo, con pause)	Minimo	Massimo	Media	Deviazione standard	Mediana
Genere: femminile	4,02	6,03	4,81	0,67	4,78
Genere: maschile	4,09	5,84	4,99	0,56	4,99
Fascia d'età: < 41 anni	5,43	6,03	5,66	0,26	5,51
Fascia d'età tra: 41 e 60 anni	4,22	5,20	4,74	0,38	4,86
Fascia d'età: > 60 anni	4,02	5,41	4,57	0,52	4,50
Orientamento politico: sinistra e centro-sinistra	4,29	6,03	5,30	0,57	5,47
Orientamento politico: centro	4,78	4,99	4,88	0,15	4,88
Orientamento politico: destra e centro-destra	4,02	5,41	4,60	0,51	4,47
Provenienza geografica: Nord	4,02	6,03	4,80	0,82	4,53
Provenienza geografica: Centro	4,29	5,84	5,05	0,50	5,15
Provenienza geografica: Sud e isole	4,09	5,51	4,79	0,54	4,87

Tabella 3. Minimo, massimo, media, deviazione standard e mediana di velocità di eloquio per categoria.

I valori per la velocità di eloquio degli individui appartenenti alla fascia d'età più giovane (fino a 40 anni: media 5,66 sillabe al secondo, deviazione standard 0,26) sembrano essere superiori rispetto alle altre fasce (da 41 a 60 anni, si registra una media di 4,74 sillabe al secondo con una deviazione standard di 0,38; dai 61 anni, una media di 4,57 sillabe al secondo, con una deviazione standard di 0,52). Questo è compatibile con la letteratura esistente sull'argomento. Infatti, è dimostrata la tendenza delle persone più anziane a parlare più lentamente rispetto a quelle più giovani: Yuan et al. (2007) rileva velocità inferiori per i parlanti più anziani di entrambi i generi, sia in cinese che in inglese: questa differenza può essere attribuita a fattori di natura fisiologica, per il ridotto controllo neuromuscolare, soprattutto negli uomini, che porta anche a una maggiore durata di sillabe e foni (Schötz 2007). Già Stewart & Ryan (1982), nel

loro esperimento percettivo, dimostrano che l'età attribuita dai partecipanti allo studio alle voci a cui sono sottoposti aumenta al rallentare del parlato; Skoog Waller et al. (2015) dimostrano che una manipolazione della velocità di eloquio porta i soggetti dell'esperimento a valutare diversamente l'età dei parlanti.

Allo stesso modo, mostrano valori maggiori i parlanti e le parlanti provenienti dal Centro Italia (media 5,05 sillabe al secondo, deviazione standard 0,50), e appartenenti a schieramenti di sinistra o centro-sinistra (media 5,30 sillabe al secondo con una deviazione standard di 0,57).

Le differenze di genere, in termini di velocità di eloquio, sembrano essere contenute, anche se emerge la tendenza, già osservata in letteratura, a registrare velocità più elevate nelle produzioni dei soggetti di genere maschile: si veda, per esempio Pépiot (2014), nel capitolo dedicato alla letteratura sull'argomento della velocità di eloquio. Anche Yuan et al. (2007) confermano questa differenza, sia per l'inglese sia per il cinese: sebbene quantitativamente lieve, essa si rivela altamente significativa.

Per la significatività di queste differenze, si rimanda al paragrafo dedicato ai risultati della regressione lineare.

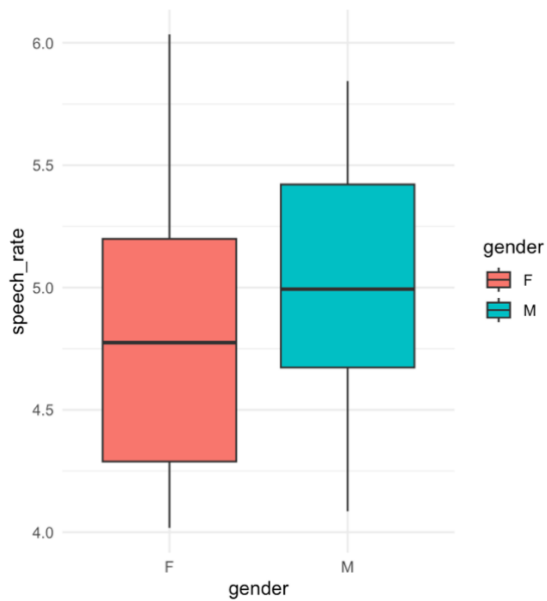


Figura 8. Boxplot dei valori di velocità di eloquio per genere.

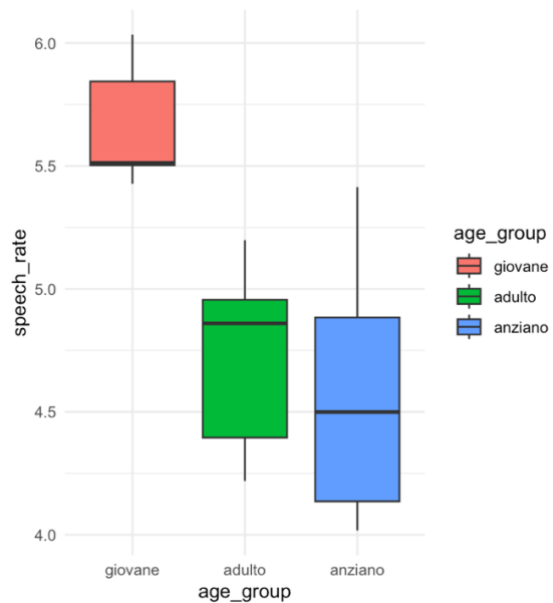


Figura 9. Boxplot dei valori di velocità di eloquio per fascia d'età.

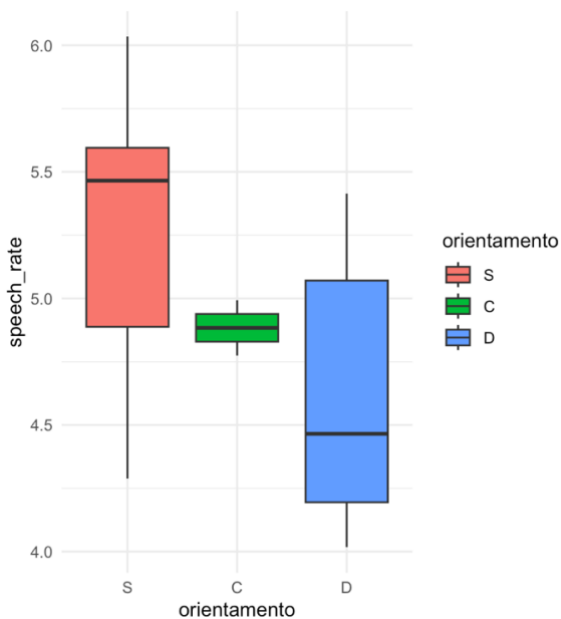


Figura 10. Boxplot dei valori di velocità di eloquio per orientamento politico.

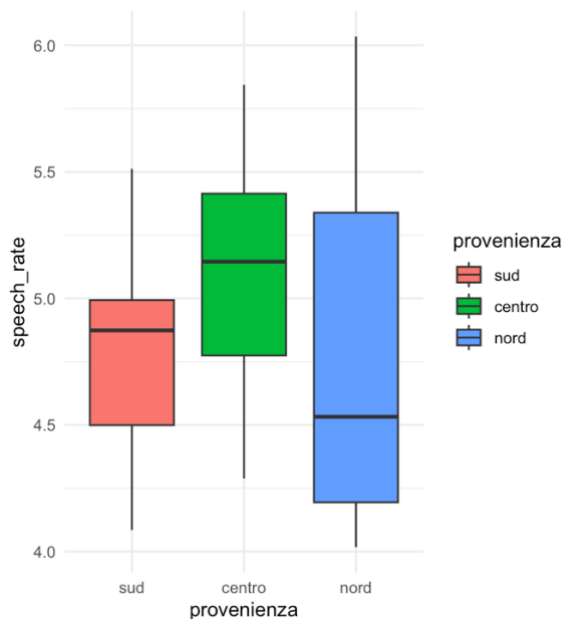


Figura 11. Boxplot dei valori di velocità di eloquio per provenienza geografica.

Nella Figura 8, possiamo vedere che i valori della velocità di eloquio per i parlanti di genere maschile (a destra) sono generalmente più elevati rispetto a quelli delle parlanti di genere femminile (a sinistra), ma che il valore massimo registrato dalle parlanti è comunque superiore rispetto al valore massimo registrato dai parlanti, come si può osservare dai *whiskers* del *boxplot*. Si ricorda che il *box* rettangolare del *boxplot* rappresenta i dati compresi tra primo e

terzo quartile (ossia la metà dei dati collocati a metà della distribuzione); la linea orizzontale interna al *box* rappresenta la mediana, che è il secondo quartile; i *whiskers*, ovvero le linee che escono dal *box* sia sopra che sotto, hanno una lunghezza pari a 1,5 volte la differenza tra il primo e il terzo quartile (tranne nei casi in cui questi valori siano fuori dal *range* ricoperto dal *dataset*: i *whiskers* si arrestano quindi al massimo o al minimo). I valori di *speech rate* delle parlanti di genere femminile presentano comunque una distribuzione più ampia rispetto ai parlanti uomini, soprattutto per quanto riguarda i valori più elevati.

Nella Figura 9 sono riportati gli stessi dati, divisi per fascia di età. È immediatamente visibile la differenza tra il gruppo di parlanti di età inferiore a 41 anni (a sinistra) e tutti gli altri: non solo essi presentano valori di gran lunga superiori, ma questi valori sono anche distribuiti nella parte superiore del *box*, con una mediana molto vicina al margine inferiore del *box*. Vi sono dunque molti valori concentrati nella zona immediatamente inferiore alla mediana (ricordiamo che la mediana è definita come il valore centrale della distribuzione, o la media dei due valori centrali se sono in numero pari), mentre l'altra metà è distribuita in maniera meno fitta, e raggiunge valori massimi di gran lunga superiori rispetto a quelli degli altri due gruppi di parlanti più anziani.

Nella Figura 10 sono riportate le distribuzioni dei valori di velocità di eloquio divisi per appartenenza politica del parlante: è subito evidente che i valori della velocità di eloquio degli individui appartenenti a schieramenti di centro (in verde) hanno una distribuzione molto più ridotta rispetto ai parlanti di sinistra/centro-sinistra e di destra/centro-destra. Anche la loro mediana è intermedia rispetto alle altre due. I più veloci sono i parlanti di orientamenti di sinistra e centro-sinistra, con una distribuzione dei valori globalmente rivolta verso l'alto, e valori massimi di gran lunga più elevati; tuttavia, la distribuzione si allunga notevolmente anche verso il basso, con valori minimi inferiori rispetto a quelli dei parlanti di centro. Nel caso dei parlanti appartenenti a schieramenti di destra o centro-destra, la mediana dei valori è inferiore rispetto a quella dei parlanti di centro, ma i valori compresi tra il primo e il terzo quartile, rappresentati dal *box*, hanno una distribuzione che comprende i valori dei parlanti di centro nella loro interezza, con minimi e massimi rispettivamente ad essi inferiori e superiori.

Infine, la Figura 11 riporta le distribuzioni dei valori di *speech rate* per provenienza geografica. È immediatamente visibile un'ampia variazione interna al gruppo di parlanti provenienti dal Nord Italia, con un *box* di dimensioni maggiori e valori minimi e massimi rispettivamente inferiori, e superiori, rispetto a quelli delle altre categorie. I parlanti provenienti dalle regioni centrali, comunque (in verde), registrano la mediana più elevata, con una

variazione interna contenuta; ancora più contenuta è la variazione interna ai valori della velocità di eloquio dei parlanti provenienti dal Sud del Paese e dalle isole maggiori.

Tuttavia, queste misure da sole non forniscono informazioni particolarmente accurate sul ruolo e l'importanza rivestiti da tali variabili: per questo motivo è necessario procedere nell'analisi attraverso la costruzione di un modello di regressione lineare.

5.3. Velocità di articolazione

Il calcolo della velocità di articolazione è definito come il rapporto tra il numero di sillabe prodotte dal parlante e il tempo totale di fonazione, che differisce dal tempo totale utilizzato per il calcolo dello *speech rate* in quanto sono stati sottratti i secondi di pausa (cfr. Capitolo 4.4).

Le interruzioni e le pause nella produzione possono dipendere da diversi fattori: da esitazioni e difficoltà nella selezione delle parole o nella gestione del discorso, ma anche, soprattutto quando si tratta di discorsi tenuti pubblicamente, o in ambienti non silenziosi, da cause esterne indipendenti dal parlante.

Per l'intero *dataset*, il valore minimo è 4,42 sillabe al secondo, il valore massimo è di 6,50 sillabe al secondo, con un *range* di 2,08; la media è di 5,66 sillabe al secondo e la mediana è di 5,85 sillabe al secondo, con una deviazione standard di 0,59. I valori per ciascuna variabile sono riportati nella Tabella 4.

Di nuovo, si tratta di risultati leggermente inferiori rispetto a quanto riportato in letteratura: Pellegrino et al. (2011) indicano, come valore medio per il parlato spontaneo italiano, 6,99 sillabe al secondo (avendo escluso dagli stimoli utilizzati per il calcolo ogni pausa di durata superiore a 150 ms). Masini et al. (2024), invece, segnalano un valore medio di 6,20 sillabe al secondo.

Articulation rate (sillabe al secondo, senza pause)	Minimo	Massimo	Media	Deviazione standard	Mediana
Genere: femminile	4,82	6,11	5,37	0,45	5,21
Genere: maschile	4,42	6,50	5,89	0,60	6,05
Fascia d'età: < 41 anni	5,95	6,38	6,13	0,18	6,11

Fascia d'età tra: tra 41 e 60 anni	5,06	6,22	5,64	0,45	5,70
Fascia d'età: > 60 anni	4,42	6,50	5,35	0,74	5,23
Orientamento politico: sinistra e centro-sinistra	5,16	6,38	5,83	0,45	5,97
Orientamento politico: centro	5,31	6,18	5,75	0,61	5,75
Orientamento politico destra e centro-destra	4,42	6,50	5,50	0,69	5,49
Provenienza geografica: Nord	4,82	6,11	5,66	0,58	5,98
Provenienza geografica: Centro	5,06	6,50	5,73	0,56	5,74
Provenienza geografica: Sud e isole	4,42	6,21	5,54	0,74	5,65

Tabella 4. Minimo, massimo, media, deviazione standard e mediana di velocità di articolazione per categoria.

Sono nuovamente gli individui di genere maschile (media di 5,89 sillabe al secondo, deviazione standard 0,60), di età inferiore ai 41 anni (media 6,13 sillabe al secondo, deviazione standard 0,18), appartenenti a schieramenti di sinistra o centro-sinistra (media di 5,83 sillabe al secondo, deviazione standard di 0,45) e provenienti dal Centro Italia (media di 5,73 sillabe al secondo, deviazione standard di 0,56) a presentare valori più elevati, anche se, al netto delle pause, la differenza di velocità tra le personalità appartenenti a schieramenti differenti viene ridimensionata.

Rispetto allo *speech rate*, osserviamo valori superiori per ogni categoria, dal momento che le pause sono state sottratte alla durata totale dei segnali audio.

Compatibilmente con le informazioni sulle pause riportate nella Tabella 2, le categorie che presentavano un rapporto di pause sul totale maggiore riportano differenze significative nei

valori di *speech rate* e *articulation rate*: si tratta, per esempio, degli appartenenti a schieramenti di destra e centro-destra (con un rapporto medio di pause sul totale di 0,16, con deviazione standard di 0,06), che hanno valori medi di *speech rate* di 4,60 (deviazione standard 0,51) e di *articulation rate* di 5,50 (deviazione standard 0,69), con una differenza di 0,90 sillabe al secondo.

Viceversa, gli individui di età inferiore ai 41 anni (che presentano un rapporto medio di pause sul totale di 0,08 (deviazione standard 0,04) passano da valori medi di *speech rate* di 5,66 (deviazione standard 0,26) a un *articulation rate* medio di 6,13 (deviazione standard di 0,18), con una differenza di 0,47 sillabe al secondo.

Per questo motivo esso è considerato una misura più affidabile rispetto allo *speech rate* per descrivere le proprietà temporali del parlato.

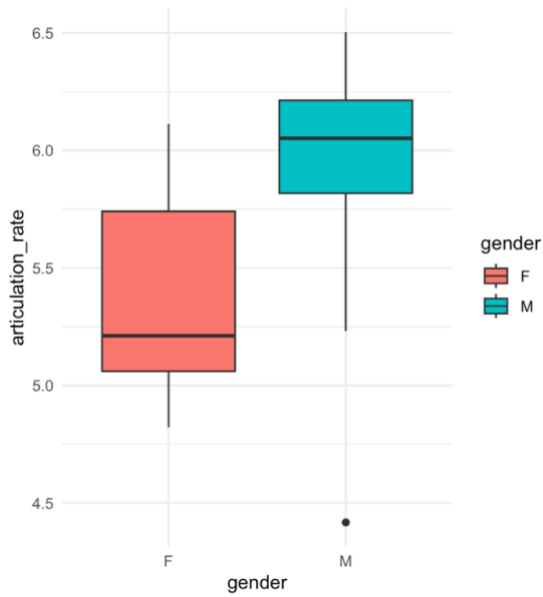


Figura 12. Boxplot dei valori di velocità di articolazione per genere.

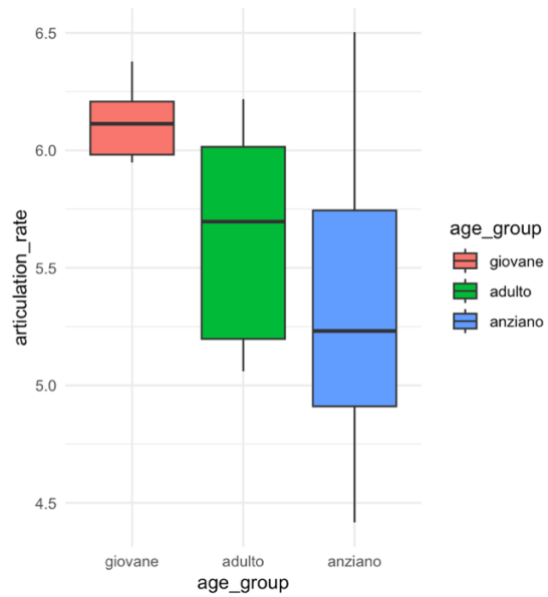


Figura 13. Boxplot dei valori di velocità di articolazione per fascia d'età.

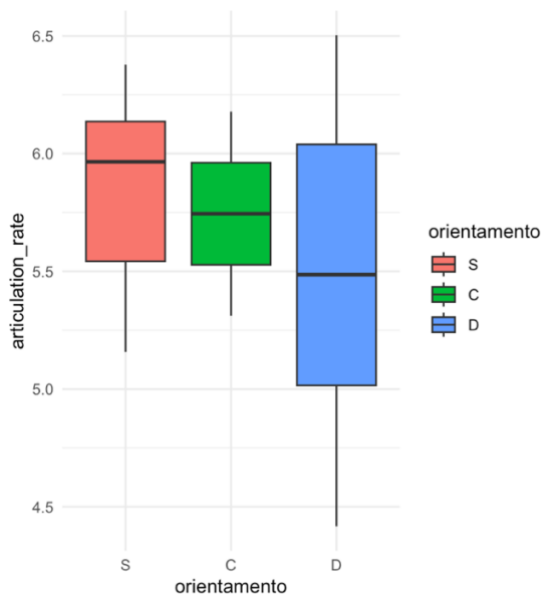


Figura 14. Boxplot dei valori di velocità di articolazione per orientamento politico.

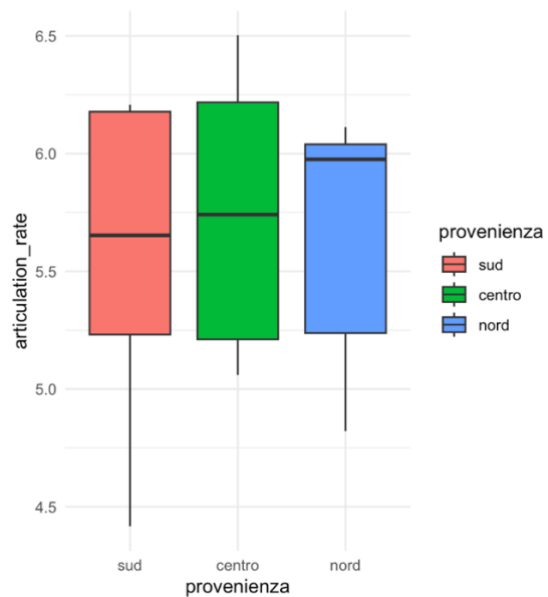


Figura 15. Boxplot dei valori di velocità di articolazione per provenienza geografica.

Nella Figura 12 possiamo osservare il *boxplot* dei valori di *articulation rate* divisi per genere: gli uomini (a destra) registrano valori nettamente superiori, per massimo, minimo e mediana. È interessante comunque osservare la presenza di un valore *outlier*, che si discosta, cioè, in maniera significativa dal resto della distribuzione, rappresentato dal punto in fondo al grafico.

I valori di *articulation rate* suddivisi per fascia d'età (Figura 13) esibiscono un comportamento simile a quelli di *speech rate*, soprattutto per quanto riguarda i parlanti più giovani, che producono valori concentrati tra di loro verso l'alto, con una variabilità interna minore rispetto agli altri due gruppi. La fascia d'età intermedia, però, si colloca al di sopra di quella più anziana, come conseguenza della maggiore proporzione di pause contenute nei discorsi dei parlanti tra i 41 e i 60 anni (Tabella 2).

Per quanto riguarda l'appartenenza politica (Figura 14), i valori relativi a parlanti di schieramenti di destra o centro-destra presentano un'ampia variazione interna (e massimi superiori) rispetto agli altri due gruppi, che pure presentano entrambi mediane superiori. Anche in questo caso, i valori degli individui appartenenti a schieramenti di centro sono concentrati nella parte centrale.

Infine, i valori centrali di *articulation rate* degli individui divisi per provenienza geografica (Figura 15) occupano all'incirca la stessa zona, anche se con alcune differenze fondamentali: la mediana è molto più elevata per i parlanti provenienti dal Nord Italia, che presentano molti valori concentrati immediatamente sopra la mediana; i parlanti provenienti dal Centro raggiungono i valori più alti, mentre quelli provenienti dal Sud e dalle isole raggiungono invece i risultati più bassi.

5.4. Regressione lineare: *speech rate*

È stata effettuata una prima regressione lineare a partire dai risultati ottenuti relativamente alla velocità di eloquio (*speech rate*; per l'*articulation rate*, si veda il paragrafo successivo). Sono stati utilizzati, come riferimento per l'intercetta, i seguenti valori:

- Genere: femminile;
- Fascia d'età: meno di 40 anni;
- Orientamento politico: sinistra o centro-sinistra;

- Provenienza geografica: Sud Italia e isole maggiori.

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.41548    0.40984  13.214 1.64e-08 ***
genderM         0.19793    0.25234   0.784  0.4480
age_groupadulto -0.99553    0.34218  -2.909  0.0131 *
age_groupanziano -1.20356    0.46837  -2.570  0.0246 *
orientamentoC   0.42371    0.48292   0.877  0.3975
orientamentoD   0.03973    0.35822   0.111  0.9135
provenienzacentro 0.29868    0.30551   0.978  0.3476
provenienzanord  0.02629    0.35959   0.073  0.9429
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4253 on 12 degrees of freedom
Multiple R-squared:  0.681,    Adjusted R-squared:  0.4949
F-statistic: 3.659 on 7 and 12 DF,  p-value: 0.02378

```

Figura 16. Interfaccia di RStudio e risultati della regressione lineare per lo *speech rate*.

La Figura 16 riporta i risultati della regressione lineare. R^2 è di 0,681 e *adjusted* R^2 di 0,4949 (valore, questo, più adatto a stimare la bontà di un modello di regressione multipla): sarebbe dunque in grado di spiegare quasi la metà della variazione presente nel *dataset*, mentre l'altra metà sarebbe dovuta al caso o a fenomeni idiosincratici. Non si tratta di un cattivo risultato per la tipologia di dati utilizzati, che sono di natura linguistica e per questo motivo soggetti a una variabilità interna molto maggiore rispetto ad altri dati appartenenti, per esempio, alle scienze naturali: «La lingua è complessa e gli esseri umani sono disordinati, quindi i nostri modelli raramente spiegano una tale varianza»¹⁹ (Winter 2020: 77).

Il modello presenta un *p-value* complessivo di 0,024, al di sotto della soglia di 0,05 per considerare un risultato statisticamente significativo: significa che un modello con questi predittori, considerati complessivamente, ha una buona capacità di spiegare il comportamento della variabile dipendente. Non significa però che ogni singolo predittore è ugualmente utile, o significativo.

La fascia d'età si rivela essere il migliore predittore per la velocità di eloquio: la fascia intermedia (dai 41 ai 60 anni) è associata con un abbassamento dello *speech rate* (*p-value*, ottenuto tramite *t test*, di 0,0131) rispetto al valore del gruppo dei giovani (meno di 40 anni), considerato come riferimento. La fascia più elevata (a partire dai 60 anni) è associata con un abbassamento dello *speech rate* (*p-value*, ottenuto tramite *t test*, di 0,0246) rispetto al valore del gruppo dei giovani (meno di 40 anni), considerato come riferimento.

¹⁹ «Language is complex and humans are messy, so our models rarely account for such variance».

Gli altri parametri non danno risultati significativi: genere maschile ($p\text{-value}$ 0,4480), orientamenti politici di centro ($p\text{-value}$ 0,3975) e di destra o centro-destra ($p\text{-value}$ 0,9135), e la provenienza dal Centro ($p\text{-value}$ 0,3476) e dal Nord Italia ($p\text{-value}$ 0,9429) presentano dati di $p\text{-value}$ di gran lunga superiori, a conferma del loro scarso valore predittivo.

La Figura 17 rappresenta i coefficienti del modello, le loro stime e i loro errori standard.

I *box* azzurri rappresentano le stime dei coefficienti (visibili nella prima colonna della tabella in Figura 16), a partire dallo zero: la loro lunghezza rappresenta il valore dell'intercetta per un determinato valore di ciascuna variabile, e dunque quanto quest'ultima influenza il modello. Le linee nere all'interno dei *box* rappresentano gli errori standard per ciascun coefficiente (visibili nella seconda colonna della tabella in Figura 16).

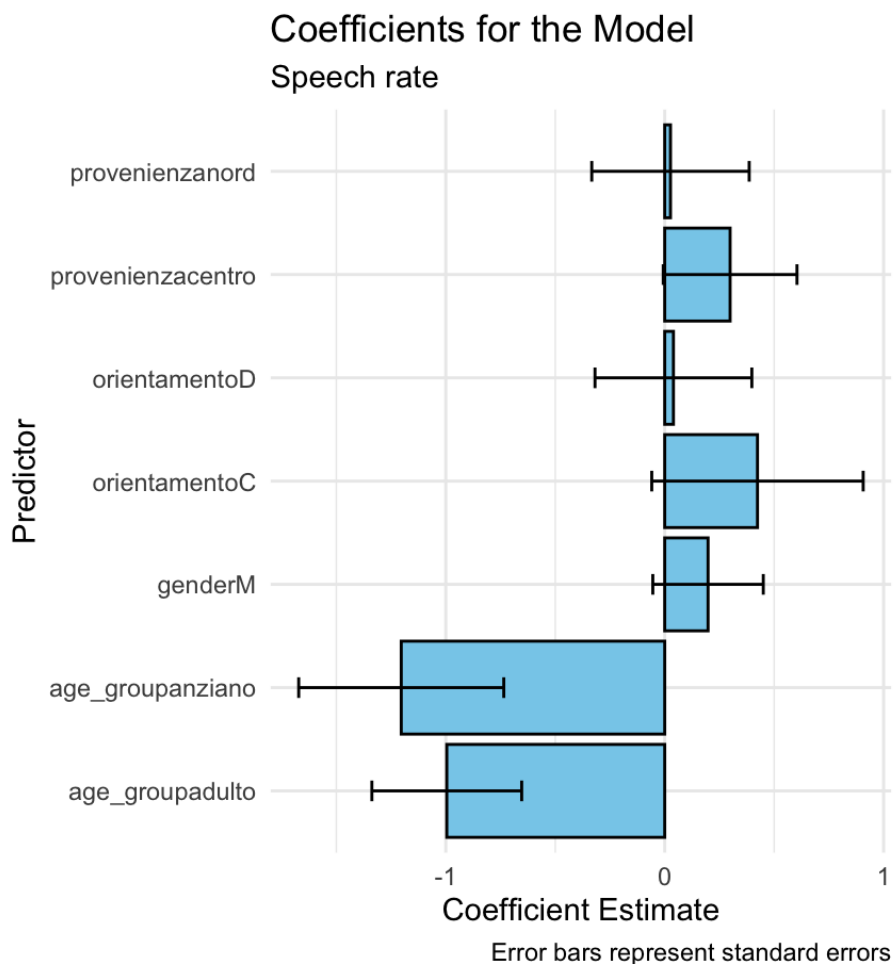


Figura 17. Coefficienti del modello di regressione lineare per lo speech rate.

Rispetto dunque ai valori di riferimento, la provenienza dal Nord Italia e l'orientamento di destra o centro-destra non sembrano influenzare notevolmente la velocità di eloquio. La provenienza dal Centro Italia e l'orientamento politico di centro mostrano una tendenza

contenuta a registrare valori maggiori di *speech rate*, mentre l'appartenenza al genere maschile genera un lieve innalzamento. La fascia d'età tra i 41 e i 60 anni e, soprattutto, quella oltre i 60 anni sono fortemente associate a una diminuzione della velocità di eloquio.

I risultati finora presentati includono comunque le pause nel computo dello *speech rate*, per cui è atteso che ci siano differenze rilevanti tra *speech rate* e *articulation rate* almeno per le categorie che, nella Tabella 2, fanno uso di una quantità rilevante di pause, come i parlanti settentrionali, o quelli appartenenti a schieramenti di destra o centro-destra.

5.5. Regressione lineare: articulation rate

È stato costruito un modello di regressione lineare anche a partire dai valori ottenuti relativamente alla velocità di articolazione (*articulation rate*). Sono stati utilizzati i medesimi valori di riferimento del paragrafo precedente (genere femminile, età inferiore a 41 anni, appartenenza politica di sinistra o centro-sinistra, provenienza meridionale o delle isole maggiori). I risultati sono riportati nella Figura 18.

La regressione lineare fornisce un coefficiente R^2 di 0,66 e un *adjusted* R^2 di 0,46: quest'ultimo valore è maggiormente attendibile, dal momento che si tratta di una regressione lineare che prevede più di una variabile indipendente.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.21255	0.41595	12.532	2.98e-08 ***
genderM	0.77437	0.25609	3.024	0.0106 *
age_groupadulto	-0.47797	0.34727	-1.376	0.1939
age_groupanziano	-0.88921	0.47535	-1.871	0.0860 .
orientamentoC	0.77331	0.49011	1.578	0.1406
orientamentoD	0.06256	0.36355	0.172	0.8662
provenienzacentro	0.52170	0.31006	1.683	0.1183
provenienzanord	0.59969	0.36495	1.643	0.1263

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4316 on 12 degrees of freedom
 Multiple R-squared: 0.6601, Adjusted R-squared: 0.4618
 F-statistic: 3.329 on 7 and 12 DF, p-value: 0.03273

Figura 18. Interfaccia di RStudio e risultati della regressione lineare per l'*articulation rate*.

Ciò significa che il modello ottenuto a partire da queste variabili e da questi valori di riferimento è in grado di spiegare circa il 46% della variazione dei dati contenuti nel *dataset*. Il 54% della variazione sarebbe invece dovuta al caso o a fenomeni idiosincratichi. Come si diceva

sopra, si tratta di un buon risultato per il tipo di dati utilizzati, che sono di natura linguistica e molto variegati (Winter 2020).

Il *p-value* complessivo del modello è di 0,03273, entro la soglia considerata convenzionalmente significativa di 0,05, Il modello è dunque in grado, complessivamente, di rendere conto dell'influenza dei predittori sulla variabile dipendente, anche se non tutti nella stessa misura.

In questo caso, è il genere ad avere l'effetto più forte sulla velocità di articolazione: in particolare, il genere maschile è associato a un innalzamento di *articulation rate* (*p-value* 0,0106). Il genere è seguito dall'appartenenza alla fascia d'età più elevata (sopra i 60 anni), che sarebbe però associata a una diminuzione della velocità di articolazione (*p-value* 0,0860). Quest'ultimo valore di *p-value* è leggermente al di sopra della soglia convenzionalmente accettata di 0,05; non è invece significativa l'appartenenza alla fascia d'età intermedia (dai 41 ai 60 anni, *p-value* 0,1939).

Rispetto ai valori utilizzati come riferimento, non risultano particolarmente significativi neppure l'appartenenza a schieramenti di centro (*p-value* 0,1406) o di destra o centro-destra (*p-value* 0,8662), e la provenienza dal Centro Italia (*p-value* 0,1183) o dal Nord del Paese (*p-value* 0,1263).

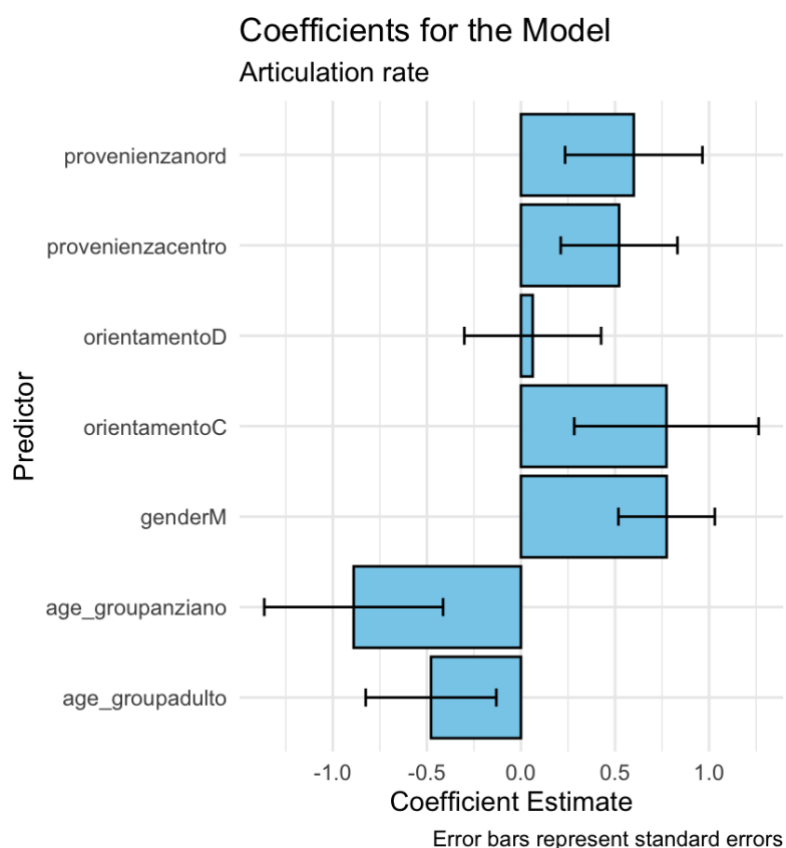


Figura 19. Coefficienti del modello di regressione lineare per l'articulation rate.

Ancora, la Figura 19 rappresenta le stime dei coefficienti del modello (prima colonna della tabella in Figura 18) e i loro errori standard (seconda colonna della tabella in Figura 18).

Rispetto ai valori di riferimento, dunque, l'appartenenza a schieramenti di destra o centro-destra non sembra apportare modifiche sostanziali ai valori di *articulation rate*. La provenienza dal Nord o dal Centro Italia, invece, è associata a un leggero aumento della velocità di eloquio, così come anche l'appartenenza a schieramenti di centro e al genere maschile.

Di nuovo, anche se in modo meno marcato rispetto al caso dello *speech rate*, l'appartenenza alla fascia d'età intermedia (41-60 anni) e, soprattutto, più anziana (più di 61 anni) è associata a una riduzione dei valori di velocità di articolazione.

5.6. Enunciati contenenti impliciti

Il *dataset* utilizzato per questa sezione dello studio è costituito da 66 enunciati isolati a partire dai *file* contenenti i discorsi politici considerati; per ogni parlante sono stati isolati tre enunciati contenenti presupposizioni e tre privi di qualsiasi tipologia di implicito. Per la presenza o assenza di impliciti, si è fatto riferimento all'annotazione presente nel *corpus* IMPAQTS; gli

enunciati selezionati sono riportati in Appendice. I segnali audio hanno una durata minima di 2,60 secondi e una durata massima di 14,55 secondi.

Non tutti gli enunciati così isolati presentano pause; il numero massimo di pause presenti in un singolo enunciato è 7.

Gli enunciati contenenti presupposizioni prodotti da donne hanno in media 1,73 pause, con una durata media di 0,37 secondi, mentre quelli privi di impliciti ne hanno in media 2,47, con una durata totale media di 0,49 secondi. Gli enunciati contenenti presupposizioni prodotti da uomini hanno in media 1,61 pause, con una durata totale media di 0,59 secondi, mentre quelli privi di impliciti hanno in media 1,28 pause, con una durata totale media di 0,37 secondi.

Negli enunciati non contenenti impliciti, il numero di pause cresce con l'età: i giovani ne producono in media 1,11, della durata totale media di 0,37 secondi, i parlanti della fascia intermedia 1,42 con una durata totale media però leggermente inferiore (0,32 secondi), mentre gli anziani 2,75, con una durata totale media di 0,58 secondi. Questa tendenza cambia per gli enunciati contenenti presupposizioni, dove gli anziani producono meno pause degli adulti (1,75, con una durata media di 0,52 secondi, rispetto a 2,00, con una durata media di 0,70 secondi); i giovani ne producono comunque di meno, in media 1,11, con una durata media di 0,18 secondi.

I parlanti appartenenti a schieramenti di sinistra e centro-sinistra producono in media meno pause, sia negli enunciati contenenti presupposizioni (1,00 pause, che durano in totale in media 0,18 secondi) sia in quelli privi di impliciti (1,08 pause, che durano in totale in media 0,37 secondi), rispetto ai parlanti di centro e di destra o centro-destra. I parlanti di centro producono in media 1,83 pause negli enunciati contenenti presupposizioni (durata media 0,67 secondi) e 2,67 pause negli enunciati privi di impliciti (durata media 0,62 secondi), con durate medie però simili. I parlanti di destra o centro-destra producono in media 2,13 pause negli enunciati contenenti presupposizioni (durata media 0,66 secondi) e 2,07 pause in quelli privi di impliciti (durata media 0,40 secondi).

I parlanti del Nord producono più pause (in media 2,22, con una durata totale media di 0,83 secondi) rispetto a quelli di Centro (1,47, durata totale media 0,31 secondi) e del Sud (1,44, durata totale media 0,45 secondi) negli enunciati contenenti presupposizioni; questo vale anche per gli enunciati privi di impliciti, dove i parlanti settentrionali producono in media 2,11 pause (durata media 0,37 secondi), quelli di Centro 1,60 (durata media 0,36 secondi), quelli del Sud e delle isole maggiori 1,89 (durata media 0,59 secondi).

Questi dati sono importanti per meglio comprendere i risultati tra *speech rate* e *articulation rate*, e in particolare per rendere conto delle differenze tra queste due misure.

La Tabella 5 riporta massimo, minimo, media, mediana e deviazione standard di *speech rate* per ciascuna categoria, relativamente agli enunciati contenenti presupposizioni.

<i>Speech rate</i> (sillabe al secondo, con pause)	Minimo	Massimo	Media	Deviazione standard	Mediana
Enunciati con presupposizioni	3,82	7,13	5,57	0,82	5,52
Genere: femminile	4,41	6,73	5,41	0,72	5,41
Genere: maschile	3,82	7,13	5,70	0,90	5,68
Fascia d'età: < 41 anni	5,41	7,13	6,08	0,55	6,05
Fascia d'età: tra 41 e 60 anni	4,25	6,73	5,62	0,77	5,61
Fascia d'età: > 61 anni	3,82	6,68	5,13	0,85	5,02
Orientamento politico: sinistra e centro-sinistra	4,84	7,13	6,06	0,64	6,06
Orientamento politico: centro	3,82	6,68	5,14	1,17	4,71
Orientamento politico: destra e centro-destra	4,25	6,54	5,34	0,62	5,44
Provenienza: Nord	4,25	6,24	5,34	0,73	5,41
Provenienza: Centro	4,41	7,13	5,72	0,85	5,65
Provenienza: Sud e isole	3,82	6,68	5,55	0,89	5,52

Tabella 5. Minimo, massimo, media, deviazione standard e mediana di speech rate per gli enunciati contenenti presupposizioni, divisi per categoria.

Gli enunciati contenenti impliciti hanno uno *speech rate* medio di 5,57 sillabe al secondo (deviazione standard 0,82). Ancora una volta sono gli uomini a mostrare valori più elevati (5,70

sillabe al secondo, deviazione standard 0,90) rispetto alle donne (5,41 sillabe al secondo, deviazione standard 0,72). I giovani presentano anche qui *speech rate* superiori (6,08 di media, deviazione standard 0,55) rispetto alla fascia d'età intermedia (5,62 sillabe al secondo, deviazione standard 0,77) e agli anziani (5,13 sillabe al secondo, deviazione standard 0,85). I parlanti appartenenti a schieramenti di sinistra presentano velocità di eloquio maggiore (6,06 sillabe al secondo, deviazione standard 0,64) rispetto ai parlanti di destra (5,34 sillabe al secondo, deviazione standard 0,62) e, soprattutto, rispetto a quelli di centro (5,14 sillabe al secondo, deviazione standard 1,17), con quasi una intera sillaba al secondo di differenza. Per quanto riguarda la provenienza, i più veloci sono i parlanti del Centro Italia, con una media di 5,72 sillabe al secondo (deviazione standard 0,85) rispetto alle 5,34 (deviazione standard 0,73) dei parlanti del Nord e 5,55 sillabe al secondo (deviazione standard 0,89) dei parlanti del Sud Italia e delle isole maggiori. I risultati sono dunque compatibili con quelli ottenuti a partire dai discorsi nella loro interezza.

La Tabella 6 riporta gli stessi dati per gli enunciati privi di qualsiasi tipo di implicito.

<i>Speech rate</i> (sillabe al secondo, con pause)	Minimo	Massimo	Media	Deviazione standard	Mediana
Enunciati privi di impliciti	3,61	7,91	5,52	1,05	5,60
Genere: femminile	3,61	6,82	5,15	0,91	5,31
Genere: maschile	3,98	7,91	5,82	1,08	5,72
Fascia d'età: < 41 anni	4,79	7,16	6,03	0,73	5,94
Fascia d'età: tra 41 e 60 anni	3,61	7,91	5,62	1,32	5,47
Fascia d'età: > 61 anni	3,85	6,33	5,02	0,75	4,90
Orientamento politico: sinistra e centro-sinistra	3,61	7,12	5,70	0,98	5,85
Orientamento politico: centro	3,85	6,33	5,42	0,84	5,61

Orientamento politico: destra e centro-destra	3,83	7,91	5,41	1,21	4,98
Provenienza: Nord	4,67	6,89	5,75	0,91	5,87
Provenienza: Centro	3,61	7,91	5,49	1,24	5,61
Provenienza: Sud e isole	3,98	6,58	5,32	0,89	5,31

Tabella 6. *Minimo, massimo, media, deviazione standard e mediana di speech rate per gli enunciati privi di impliciti, divisi per categoria.*

Gli enunciati privi di impliciti presentano uno *speech rate* medio di 5,52 sillabe al secondo (deviazione standard 1,05). Compatibilmente con quanto riscontrato in precedenza, gli uomini hanno *speech rate* superiori rispetto alle donne (5,82 sillabe al secondo e 1,08 di deviazione standard, a fronte di 5,15 sillabe al secondo e 0,91 di deviazione standard); la fascia d'età più giovane (6,03 sillabe al secondo, con 0,73 di deviazione standard) è ancora una volta più veloce rispetto a quella intermedia (5,62 sillabe al secondo, deviazione standard 1,32) e, soprattutto, a quella più anziana (5,02 sillabe al secondo, deviazione standard). I politici di sinistra sono anche in questo caso più veloci, con una media di 5,70 sillabe al secondo (deviazione standard 0,98) rispetto agli appartenenti a schieramenti di centro (5,42 sillabe al secondo, deviazione standard 0,84) e di destra o centro-destra (5,41 sillabe al secondo, deviazione standard 1,21), che presentano medie molto simili tra di loro; questo nonostante il valore più basso in assoluto sia attestato entro questa categoria. Dal punto di vista della provenienza geografica, in questo caso i parlanti del Nord sono leggermente più veloci (5,75 sillabe al secondo di media, deviazione standard 0,91) rispetto ai parlanti provenienti dal resto d'Italia (Centro: 5,49 sillabe al secondo, con deviazione standard di 1,24; Sud Italia e isole maggiori: 5,32 sillabe al secondo, con deviazione standard di 0,89), diversamente da quanto accade negli enunciati contenenti presupposizioni, nei quali sono i parlanti di centro ad avere le medie superiori.

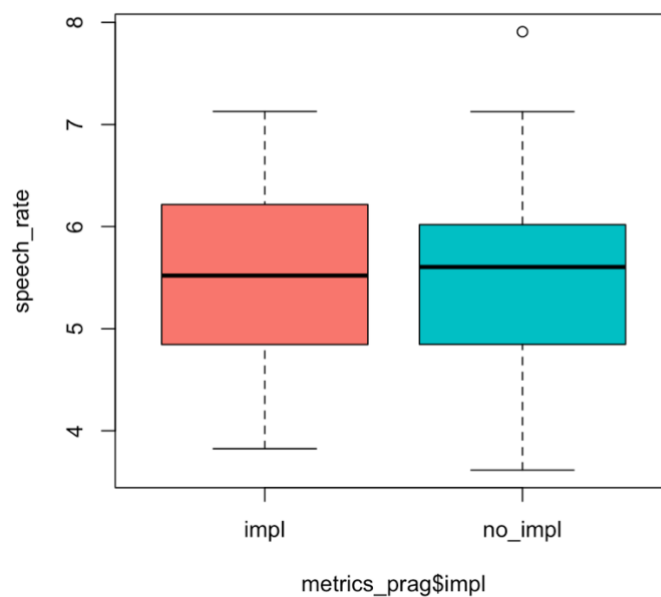


Figura 20. Valori di *speech rate* per enunciati contenenti presupposizioni ed enunciati privi di impliciti di qualsiasi tipologia.

La distribuzione dei valori di *speech rate* a seconda della presenza o assenza di presupposizioni è comunque molto simile (Figura 20); oltre a un valore *outlier* nella categoria degli enunciati privi di impliciti, indicato dal puntino al di sopra del grafico, la mediana per questi enunciati è solo lievemente superiore e i dati occupano all'incirca la stessa zona del grafico.

Vediamo ora queste stesse coppie di dati, suddivisi per le altre categorie: in rosso sono riportati i *boxplot* relativi alla velocità di eloquio negli enunciati contenenti presupposizioni, mentre gli stessi valori per gli enunciati privi di impliciti sono in blu.

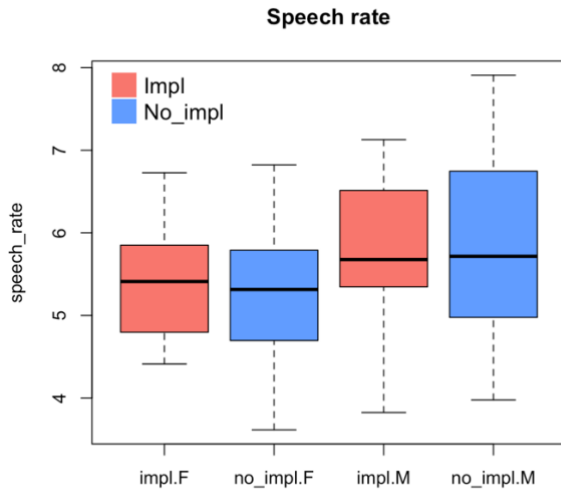


Figura 21. Valori di speech rate per genere, in presenza e in assenza di presupposizioni.

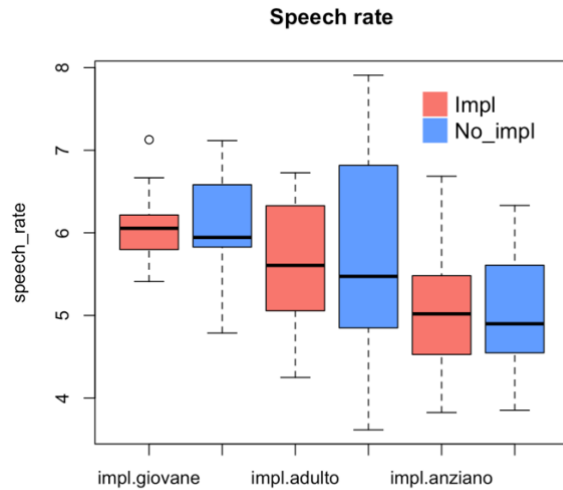


Figura 22. Valori di speech rate per età, in presenza e in assenza di presupposizioni.

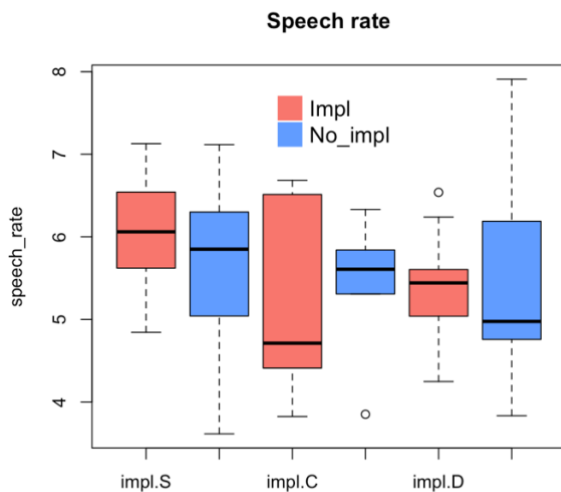


Figura 23. Valori di speech rate per orientamento politico, in presenza e in assenza di presupposizioni.

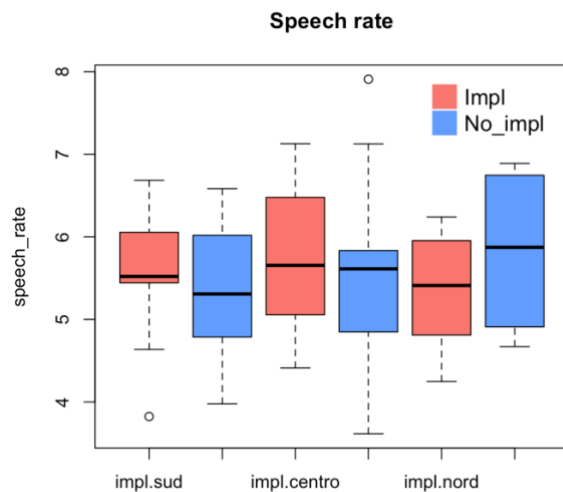


Figura 24. Valori di speech rate per provenienza geografica, in presenza e in assenza di presupposizioni.

È confermata la tendenza degli uomini a presentare velocità di eloquio maggiore, in entrambi i tipi di enunciati; gli enunciati privi di impliciti (in blu) presentano per entrambi i generi valori più sparsi, ma la presenza o assenza di presupposizioni non comporta conseguenze significative per le mediane (Figura 21).

In Figura 22 sono riportati i risultati per le tre fasce d'età: i giovani ancora una volta hanno mediane superiori agli altri gruppi, con distribuzioni più compatte e spostate verso l'alto (è presente un valore *outlier* verso l'alto nel caso degli enunciati contenenti presupposizioni pronunciati dai parlanti giovani). I parlanti della fascia d'età intermedia presentano la distribuzione più ampia, soprattutto per gli enunciati privi di impliciti, mentre gli anziani presentano le mediane più basse in assoluto.

La Figura 23 rappresenta mediane e distribuzioni, suddivise per orientamento politico: i parlanti di sinistra e centro-sinistra, e di destra e centro-destra, presentano valori della mediana più elevati negli enunciati contenenti presupposizioni (in rosso), mentre la tendenza è invertita per i parlanti appartenenti a schieramenti di centro. Comunque, gli enunciati con presupposizioni dei politici di centro presentano una distribuzione molto ampia, così come gli enunciati privi di messaggi impliciti dei politici di destra e centro-destra, che presentano valori massimi di molto superiori agli altri gruppi. Sono presenti due valori *outlier*: gli enunciati privi di impliciti dei politici di centro presentano un valore particolarmente basso, mentre gli enunciati con presupposizioni dei politici di destra presentano un valore particolarmente alto. Questi valori, che cadono al di fuori dei *whiskers* del *boxplot*, sono rappresentati mediante i cerchi sopra e sotto il grafico.

Infine, dal punto di vista della provenienza geografica, in Figura 24 sono registrate mediane simili per tutti i gruppi, ma lievemente superiori, in presenza di presupposizioni, nelle produzioni dei parlanti del Sud e del Centro del Paese; la tendenza si inverte per i parlanti del Nord. Gli enunciati con presupposizioni dei parlanti del Sud presentano valori *outlier* particolarmente bassi, mentre gli enunciati privi di impliciti dei parlanti del Centro Italia presentano un picco che sfiora le 8 sillabe al secondo.

È stato realizzato un modello di regressione lineare per questi valori di *speech rate* rispetto al riferimento (genere femminile, fascia d'età più giovane, orientamento politico di sinistra o centro-sinistra, provenienza dal sud Italia). I risultati sono riportati nella Figura 25.

```

Call:
lm(formula = speech_rate ~ metrics_prag$impl + metrics_prag$gender +
    metrics_prag$age + metrics_prag$orientamento + metrics_prag$provenienza,
    data = metrics_prag)

Residuals:
    Min       1Q   Median       3Q      Max
-1.76513 -0.51929  0.07998  0.53796  2.04337

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.30796    0.47621  11.146 6.07e-16 ***
metrics_prag$impl  0.05196    0.21138   0.246  0.8067
metrics_prag$genderM  0.63110    0.31389   2.011  0.0491 *
metrics_prag$ageadulto -0.39406    0.46308  -0.851  0.3984
metrics_prag$ageanziano -0.94431    0.60234  -1.568  0.1225
metrics_prag$orientamentoC  0.37668    0.57477   0.655  0.5149
metrics_prag$orientamentoD -0.08379    0.47176  -0.178  0.8597
metrics_prag$provenienzacentro  0.40403    0.38669   1.045  0.3005
metrics_prag$provenienzanord  0.50058    0.40997   1.221  0.2271
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8586 on 57 degrees of freedom
Multiple R-squared:  0.2601,    Adjusted R-squared:  0.1562
F-statistic: 2.504 on 8 and 57 DF,  p-value: 0.02096

```

Figura 25. Interfaccia di RStudio e risultati della regressione lineare che considera la presenza di impliciti, oltre alle variabili già analizzate, per lo *speech rate*.

Il modello presenta complessivamente R^2 di 0,2601 e *adjusted* R^2 di 0,1562: è dunque in grado di spiegare solo meno del 16% della variabilità interna ai dati. Il *p-value* complessivo è di 0,02096, quindi significativo perché entro la soglia di 0,05.

L'unica variabile che mostra di influenzare lo *speech rate* nel *dataset* è il genere: l'appartenenza al genere maschile (*p-value* 0,0491) è associata a un aumento della velocità di eloquio.

L'appartenenza a entrambe le fasce d'età superiori è legata a un abbassamento della velocità, ma l'effetto non è significativo (*p-value* 0,3984 tra i 41 e i 60 anni, 0,1225 oltre i 60); l'orientamento politico non si rivela significativo (*p-value* 0,5149 per gli appartenenti a schieramenti di centro, 0,8597 per gli appartenenti a schieramenti di destra e centro-destra). La provenienza da regioni centrali o settentrionali, sebbene legata, in entrambi i casi, a un aumento della velocità di eloquio, non influisce in modo significativo sui risultati del modello (*p-value* rispettivamente di 0,3005 e 0,2271).

Per quanto riguarda l'*articulation rate*, il nuovo *dataset* presenta un minimo di 3,65 sillabe al secondo, un massimo di 7,91 sillabe al secondo (con un range di 4,26 sillabe al secondo), una media di 5,89 sillabe al secondo, una mediana di 5,95 sillabe al secondo e una deviazione standard di 0,58 sillabe al secondo. I risultati per gli enunciati selezionati per la seconda parte dello studio non sono significativamente diversi rispetto a quelli ottenuti a partire da i discorsi

nella loro interezza e dalle *features* sociolinguistiche, in termini di media e mediana, nonostante presentino valori più estremi per massimo e minimo. Gli enunciati selezionati sembrano dunque fornire una rappresentazione appropriata del primo *dataset* utilizzato.

La Tabella 7 riporta i dati ottenuti, per categoria, in riferimento agli enunciati contenenti impliciti.

Articulation rate (sillabe al secondo, senza pause)	Minimo	Massimo	Media	Deviazione standard	Mediana
Enunciati con presupposizioni	4,64	7,13	5,96	0,67	5,95
Genere: femminile	4,81	7,09	5,72	0,70	5,74
Genere: maschile	4,64	7,13	6,16	0,59	6,06
Fascia d'età: < 41 anni	5,73	7,13	6,28	0,45	6,22
Fascia d'età: tra 41 e 60 anni	4,96	7,09	6,09	0,65	5,95
Fascia d'età: > 61 anni	4,64	6,78	5,58	0,69	5,54
Orientamento politico: sinistra e centro-sinistra	4,96	7,13	6,28	0,62	6,26
Orientamento politico: centro	4,87	6,78	5,78	0,79	5,70
Orientamento politico: destra e centro-destra	4,64	6,83	5,78	0,60	5,81
Provenienza: Nord	4,81	6,53	5,77	0,50	5,83
Provenienza: Centro	4,87	7,13	6,02	0,78	6,04
Provenienza: Sud e isole	4,64	6,78	6,05	0,63	6,06

Tabella 7. Minimo, massimo, media, deviazione standard e mediana di articulation rate per gli enunciati contenenti presupposizioni, divisi per categoria.

Globalmente, gli enunciati contenenti presupposizioni sono pronunciati con una velocità di articolazione media di 5,96 sillabe al secondo (deviazione standard 0,67).

Compatibilmente con quanto rilevato a partire dagli interi discorsi e dall'analisi dello *speech rate* su questi enunciati, gli uomini (media 6,16 sillabe al secondo, deviazione standard 0,59) presentano valori più alti rispetto alle donne (media 5,72 sillabe al secondo, deviazione standard 0,70); la velocità media diminuisce con l'aumentare dell'età dei soggetti (meno di 41 anni: media 6,28 sillabe al secondo, deviazione standard 0,45; tra 41 e 60 anni: media 6,09 sillabe al secondo, deviazione standard 0,65; più di 61 anni: media 5,58 sillabe al secondo, deviazione standard 0,69).

I parlanti appartenenti a schieramenti di sinistra mostrano velocità di articolazione maggiore (media 6,28 sillabe al secondo, deviazione standard 0,62) rispetto a quelli di altri orientamenti (centro: media 5,78 sillabe al secondo, deviazione standard 0,79; destra e centro-destra: media 5,78 sillabe al secondo, deviazione standard 0,60). Questo è compatibile con quanto osservato a partire dai dati relativi allo *speech rate*.

La provenienza geografica non dà luogo a differenze significative: i parlanti del Nord sono leggermente più lenti con una media di 5,77 sillabe al secondo e una deviazione standard di 0,50, mentre i parlanti del Centro e del Sud Italia presentano medie molto simili (rispettivamente, 6,02 sillabe al secondo, con una deviazione standard di 0,78 e 6,05 sillabe al secondo, con una deviazione standard di 0,63). In questo senso, i risultati si discostano da quelli ottenuti per lo *speech rate*: negli enunciati privi di impliciti, erano proprio i parlanti del Nord a presentare valori medi di velocità più alti.

La Tabella 8 riporta i dati ottenuti, per categoria, in riferimento agli enunciati privi di impliciti.

<i>Articulation rate</i> (sillabe al secondo, senza pause)	Minimo	Massimo	Media	Deviazione standard	Mediana
Enunciati privi di impliciti	3,65	7,91	5,83	0,99	5,94
Genere: femminile	3,65	6,82	5,49	0,83	5,52
Genere: maschile	4,06	7,91	6,12	1,04	6,25

Fascia d'età: < 41 anni	5,30	7,18	6,28	0,62	6,09
Fascia d'età: tra 41 e 60 anni	3,65	7,91	5,89	1,17	5,83
Fascia d'età: > 61 anni	4,06	6,91	5,44	0,93	5,49
Orientamento politico: sinistra e centro-sinistra	3,65	7,18	5,95	0,94	5,99
Orientamento politico: centro	4,25	6,91	6,00	0,92	6,24
Orientamento politico: destra e centro-destra	4,06	7,91	5,67	1,09	5,47
Provenienza: Nord	5,01	6,89	5,96	0,71	5,94
Provenienza: Centro	3,65	7,91	5,79	1,12	5,92
Provenienza: Sud e isole	4,06	6,97	5,67	1,08	6,09

Tabella 8. Minimo, massimo, media, deviazione standard e mediana di articulation rate per gli enunciati privi di impliciti, divisi per categoria.

Complessivamente, questi enunciati, caratterizzati dall'assenza di messaggi impliciti di qualsiasi genere, presentano un valore medio di velocità di articolazione di 5,83 sillabe al secondo (deviazione standard 0,99), solo lievemente inferiore rispetto a quello riscontrato per gli enunciati che contenevano presupposizioni.

Di nuovo, i valori per i soggetti di genere maschile sono generalmente più alti rispetto a quelli delle donne, con una media di 6,12 sillabe al secondo (deviazione standard 1,04) rispetto a 5,49 sillabe al secondo (deviazione standard 0,83).

Anche in questo caso l'*articulation rate* diminuisce con l'avanzare dell'età: i soggetti di età inferiore a 41 anni presentano velocità di articolazione media di 6,28 sillabe al secondo (deviazione standard 0,62), i soggetti di età compresa tra i 41 e i 60 anni presentano velocità di articolazione media di 5,89 sillabe al secondo (deviazione standard 1,17), mentre i soggetti di

età superiore ai 61 presentano un valore medio di 5,44 sillabe al secondo (deviazione standard 0,93).

I parlanti appartenenti a schieramenti di sinistra o di centro-sinistra e di centro presentano medie molto simili: rispettivamente, valori medi di 5,95 sillabe al secondo (deviazione standard 0,94) e di 6,00 sillabe al secondo (deviazione standard 0,92). I parlanti appartenenti a schieramenti di destra o centro-destra sono globalmente più lenti, con una media di 5,67 sillabe al secondo (deviazione standard 1,09). In questo senso i risultati differiscono rispetto agli enunciati contenenti impliciti: in quel caso, politici di centro e di destra o centro-destra presentavano medie uguali, inferiori a quelle dei parlanti di sinistra o centro-sinistra.

Infine, i parlanti provenienti dal Nord Italia sono leggermente più veloci, con una media di 5,96 sillabe al secondo (deviazione standard 0,71) rispetto ai parlanti provenienti dalle regioni centrali (5,79 sillabe al secondo, deviazione standard 1,12) e dal Sud o dalle isole maggiori (5,67 sillabe al secondo, deviazione standard 1,08).

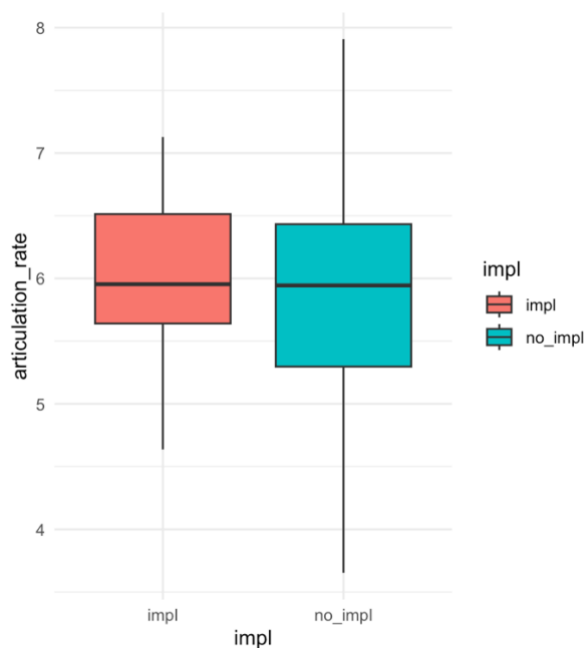


Figura 26. Valori di articulation rate per enunciati contenenti presupposizioni ed enunciati privi di impliciti.

Com'è osservabile dalla Figura 26, la mediana dei valori di velocità di articolazione degli enunciati contenenti presupposizioni e degli enunciati privi di qualsiasi implicito è molto simile, ma la distribuzione dei valori per gli enunciati che non contengono presupposizioni è molto più ampia, con valori minimi inferiori a 4 sillabe al secondo e picchi che sfiorano le 8

sillabe al secondo. Per quanto riguarda gli enunciati contenenti presupposizioni, i minimi sono di poco superiori a 4,5 sillabe al secondo e i massimi di poco superiori alle 7 sillabe al secondo.

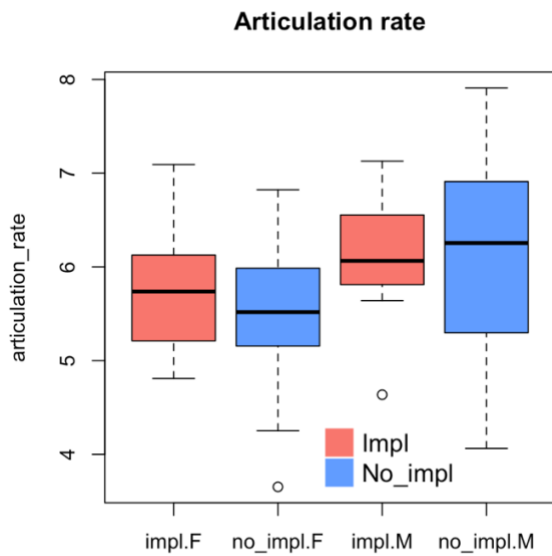


Figura 27. Valori di articulation rate per genere, in presenza e assenza di impliciti.

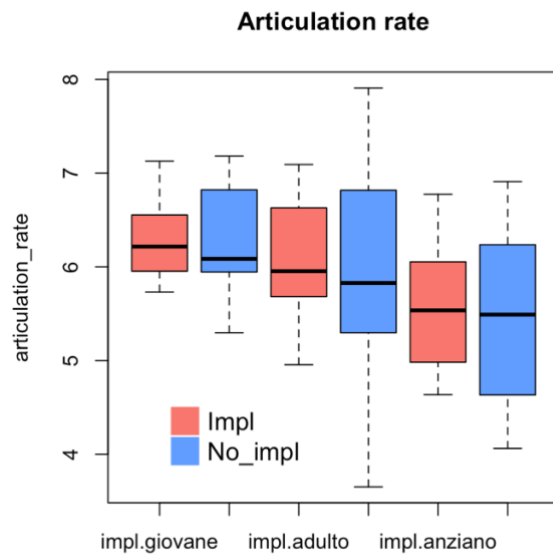


Figura 28. Valori di articulation rate per età, in presenza e assenza di impliciti.

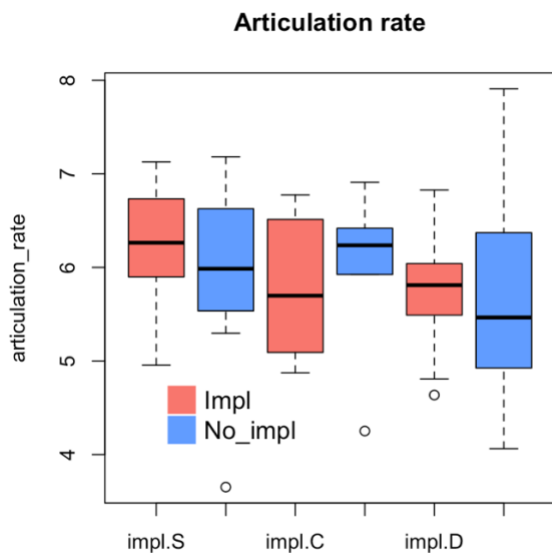


Figura 29. Valori di articulation rate per orientamento politico, in presenza e assenza di impliciti.

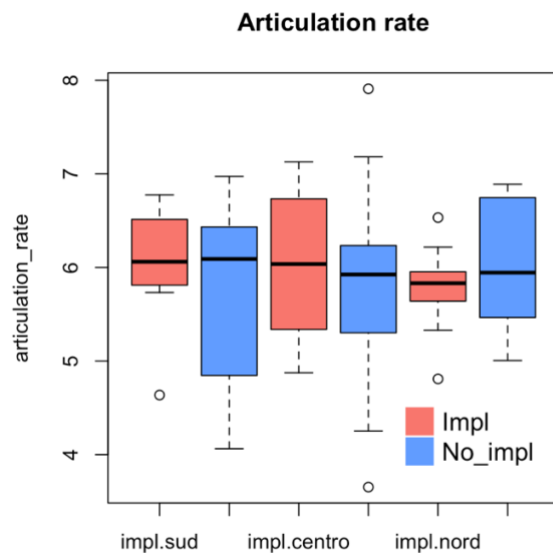


Figura 30. Valori di articulation rate per provenienza geografica, in presenza e assenza di impliciti.

La Figura 27 raffigura i valori medi di velocità di articolazione negli enunciati contenenti presupposizioni (in rosso) ed enunciati che ne sono privi (in blu), prima per gli individui di genere femminile (a sinistra), poi per gli individui di genere maschile (a sinistra).

Nel caso delle donne, entrambe le categorie di enunciati presentano una mediana molto simile, e una distribuzione dei valori paragonabile, anche se i valori sono globalmente lievemente superiori nel caso degli enunciati che contengono presupposizioni. Negli enunciati privi di impliciti pronunciati da parlanti donne, e in quelli contenenti presupposizioni pronunciati dagli uomini, sono presenti valori *outlier*, che si discostano sensibilmente dal resto della distribuzione (in entrambi questi casi, verso il basso): essi sono rappresentati dai punti al di sotto del *boxplot*. Gli uomini presentano invece valori molto più condensati intorno alla mediana negli enunciati contenenti presupposizioni; gli enunciati privi di impliciti presentano invece una distribuzione molto ampia di velocità di articolazione, con un valore minimo intorno alle 4 sillabe al secondo e un massimo che sfiora le 8 sillabe al secondo.

In Figura 28 sono riportate le stesse coppie di dati, riferiti a enunciati con presupposizioni (in rosso) e altri che sono privi di impliciti (in blu), divisi per fascia d'età: la prima coppia si riferisce ai parlanti di età inferiore ai 41 anni, la seconda coppia si riferisce a parlanti di età compresa tra i 41 e i 60 anni, la terza coppia si riferisce a parlanti di età superiore ai 60 anni. I giovani riportano in entrambe le categorie di enunciati le mediane superiori, con distribuzioni relativamente contenute; la fascia d'età intermedia riporta, a sua volta, valori della mediana intermedi, anche se nel caso degli enunciati privi di impliciti la distribuzione è decisamente più ampia rispetto alle altre, con un minimo inferiore a 4 sillabe al secondo e un massimo che si avvicina alle 8 sillabe al secondo. Infine, i parlanti più anziani presentano globalmente le mediane più basse, e distribuzioni contenute. In tutti i casi, le mediane relative agli enunciati contenenti presupposizioni sono solo leggermente più alte rispetto a quelle degli enunciati che non ne contengono.

La Figura 29 riporta queste stesse coppie di dati, divisi per orientamento politico: è interessante in questa divisione la presenza di ben tre valori definibili *outlier* rispetto alle distribuzioni (nel caso degli enunciati privi di impliciti pronunciati da parlanti di sinistra o centro-sinistra e di centro, e degli enunciati contenenti presupposizioni pronunciati da parlanti di destra o centro-destra). Per i parlanti di sinistra o centro-sinistra e di destra o centro-destra, le mediane riferite agli enunciati contenenti presupposizioni sono superiori rispetto a quelle degli enunciati privi di impliciti, mentre la tendenza si inverte per i parlanti di centro. Tra i due poli opposti, i parlanti di sinistra e centro-sinistra presentano valori generalmente più elevati,

anche se la distribuzione dei valori di *articulation rate* per gli enunciati privi di messaggi impliciti pronunciati dai parlanti di destra o centro-destra è molto ampia rispetto a tutte le altre.

Infine, la Figura 30 rappresenta queste coppie di dati divise per provenienza geografica. Per tutte le categorie, sia di provenienza che di tipologia di enunciato, le mediane si mantengono stabili intorno alle 6 sillabe al secondo; gli enunciati contenenti presupposizioni pronunciati da parlanti del Nord e, soprattutto, del Sud Italia presentano una distribuzione più concentrata rispetto agli altri gruppi, e in generale più concentrata rispetto a quelle degli enunciati privi di impliciti degli stessi parlanti. Gli enunciati privi di impliciti pronunciati da parlanti provenienti dalle regioni del Centro Italia presentano molti valori esterni ai due quartili centrali, rappresentati dal *box* del grafico. Infine, in questo caso ci sono ben cinque valori *outlier*: in entrambe le direzioni del *boxplot* nel caso degli enunciati privi di impliciti pronunciati da parlanti di Centro e nel caso degli enunciati contenenti presupposizioni pronunciati da parlanti del Nord, e verso la parte bassa del grafico nel caso degli enunciati contenenti presupposizioni pronunciati da parlanti del Sud.

È stato inoltre realizzato un modello di regressione lineare anche per la presenza o assenza di impliciti per l'*articulation rate* (Figura 31). Sono state considerate tutte le variabili utilizzate anche nelle regressioni precedenti, oltre alla presenza o assenza di impliciti. Come valori di riferimento sono stati utilizzati i medesimi delle regressioni precedenti (genere femminile, età inferiore a 41 anni, appartenenza a schieramenti di sinistra o centro-sinistra, provenienza meridionale), oltre all'assenza di impliciti.

```

Call:
lm(formula = articulation_rate ~ impl + gender + age + orientamento +
    provenienza, data = metrics_prag)

Residuals:
    Min       1Q   Median       3Q      Max
-1.90208 -0.45445  0.00761  0.50176  1.75984

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.69975    0.41110   13.865 <2e-16 ***
impl          0.12738    0.18248    0.698  0.4880
genderM       0.57082    0.27098    2.107  0.0396 *
ageadulto    -0.21862    0.39977   -0.547  0.5866
ageanziano   -1.09868    0.51999   -2.113  0.0390 *
orientamentoC  0.89880    0.49619    1.811  0.0753 .
orientamentoD  0.02290    0.40726    0.056  0.9554
provenienzacentro 0.07392    0.33383    0.221  0.8255
provenenzanord  0.33489    0.35392    0.946  0.3480
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7412 on 57 degrees of freedom
Multiple R-squared:  0.3157,    Adjusted R-squared:  0.2197
F-statistic: 3.287 on 8 and 57 DF,  p-value: 0.003757

```

Figura 31. Risultato del modello di regressione lineare che considera la presenza di impliciti, oltre alle variabili già analizzate, per l'articulation rate.

Il modello si rivela meno efficace dei precedenti (R^2 0,3157, *adjusted* R^2 0,2197): è in grado di spiegare meno del 22% della varianza presente nel nuovo *dataset*. Il *p-value* dell'intero modello è, in questo caso, 0,003757, quindi significativo perché entro la soglia di 0,05: il modello è dunque ragionevolmente in grado di rendere conto della variazione interna ai dati, a partire dalle variabili utilizzate. Rispetto ad esse emergono alcuni risultati significativi:

- risulta significativamente associata a un aumento della velocità di articolazione l'appartenenza al genere maschile (*p-value* 0,0396);
- è fortemente legata a una diminuzione della velocità di articolazione l'appartenenza alla fascia d'età più elevata (*p-value* 0,0390);
- sebbene di poco al di sotto della soglia convenzionalmente accettata di 0,05, anche l'appartenenza a schieramenti di centro è associata a un aumento di *articulation rate* (*p-value* 0,0753).

Tuttavia, la presenza o assenza di impliciti non sembra avere nessun valore predittivo rispetto alla velocità di articolazione (*p-value* 0,4880), così come anche l'appartenenza alla fascia d'età intermedia (*p-value* 0,5866), a schieramenti politici di destra o centro-destra (*p-value* 0,9554). Neanche la provenienza geografica si rivela in alcun modo significativa: rispetto ai valori di riferimento, non influisce la provenienza dal Centro Italia (*p-value* 0,8255) o dal Nord (*p-value* 0,3480).

Per valutare in modo isolato l'effetto della presenza di presupposizioni sulla velocità di articolazione, è stato costruito un secondo modello di regressione lineare semplice (Figura 32).

```
Call:
lm(formula = articulation_rate ~ impl, data = metrics_prag)

Residuals:
    Min       1Q   Median       3Q      Max
-2.17788 -0.53177  0.08313  0.58502  2.07776

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.8308      0.1468  39.727  <2e-16 ***
implimpl     0.1274      0.2076   0.614   0.542
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8432 on 64 degrees of freedom
Multiple R-squared:  0.00585, Adjusted R-squared:  -0.009683
F-statistic: 0.3766 on 1 and 64 DF, p-value: 0.5416
```

Figura 32. Risultati della regressione lineare semplice, che indaga il rapporto tra la tipologia di enunciato e l'articulation rate.

Questo modello ha permesso di determinare se la presenza di presupposizioni nell'enunciato ha un effetto significativo sulla velocità di eloquio: l'effetto si rivela inesistente (R^2 0,00585, p -value 0,542).

Non sembra quindi esserci un legame tra la presenza o assenza di presupposizioni e l'*articulation rate* nel campione di enunciati estratti dai discorsi considerati nello studio.

Un'indagine ulteriore, con un campione più ampio, potrebbe rivelarsi più utile per investigare il legame tra la presenza di contenuti impliciti nel discorso, sia politico che appartenente ad altri ambiti, e le caratteristiche temporali del parlato.

6. Conclusione e sviluppi futuri

Nei primi capitoli di questo elaborato sono state fornite alcune basi teoriche relative alla velocità di eloquio e di articolazione (Capitolo 2) e al parlato politico (Capitolo 3), con particolare attenzione rivolta agli studi di natura fonetica e alla tematica della persuasione. Nel Capitolo 4 è stata approfondita la metodologia con la quale è stato condotto lo studio, che è diviso in due parti: dapprima si pone come obiettivo scoprire il legame tra velocità di eloquio e variabili di natura sociolinguistica, entro il dominio del parlato politico dell'Italia contemporanea, a partire da una selezione di discorsi tenuti da personalità politiche di primo piano. In secondo luogo, è stata effettuata una selezione interna al primo campione di discorsi, che ha isolato diversi enunciati contenenti presupposizioni e altri che ne sono privi, che sono serviti come gruppo di controllo; gli enunciati selezionati sono riportati in Appendice. A partire da questo secondo *dataset*, sono state ricercate differenze, in termini di pause e di velocità di eloquio, tra le due categorie di enunciati.

Lo studio si è avvalso di *corpora* multimediali, che presentano una componente fonetica, e *in primis* del *corpus* IMPAQTS, elaborato dalle università di Firenze e di Roma Tre. Da questa risorsa sono stati ottenuti i materiali grezzi utilizzati, 20 discorsi pronunciati da 11 diverse personalità politiche tra il 1994 e il 2021.

Di ogni discorso è stata svolta la trascrizione automatica mediante lo strumento di *Automatic Speech Recognition* (ASR), poi corretta manualmente, per preparare i dati prima di utilizzare i servizi messi a disposizione da WebMaus (Kisler et al. 2017). Questo *web service*, messo a disposizione dall'università di Monaco di Baviera, permette di creare file di tipo *.TextGrid* leggibili dal programma Praat. A partire dai *file* audio e *.TextGrid*, uno *script* Python (Combei & Didoni 2024) ha permesso di estrarre le misure relative alle caratteristiche temporali del parlato preso in analisi (pause, velocità di eloquio, velocità di articolazione). I risultati dall'analisi sono riportati all'interno del quinto capitolo, accompagnati da alcune statistiche descrittive e dai risultati delle regressioni lineari effettuate.

Il *dataset* composto dai discorsi selezionati presenta globalmente valori di velocità di eloquio inferiori rispetto a quanto riportato in letteratura come media per la lingua italiana (Rodero 2012; Masini et al. 2024); questo risultato è compatibile col fatto che si tratta di discorsi spesso pianificati, e in generale che presentano contenuti di maggiore complessità rispetto al parlato spontaneo e appartenente ad altre tipologie (Rodero 2012 e Masini et al. 2024 si concentrano sul parlato radiofonico). Inoltre, rispetto al parlato spontaneo, il parlato politico è pensato per un pubblico più ampio: vi è la necessità che esso sia comprensibile e accessibile

per tutti gli individui che costituiscono la base elettorale, e che inevitabilmente presentano età, livelli di istruzione e provenienza socioculturale molto diversi.

Sono confermate molte delle tendenze già viste in letteratura: come riscontrato da Pépiot (2014) per l'inglese e il francese, gli uomini presentano velocità sia di eloquio che di articolazione mediamente superiori rispetto alle donne. Soprattutto, c'è una relazione netta e statisticamente significativa tra l'età dei parlanti e la loro velocità, sia di eloquio che di articolazione: questo vale sia per il *dataset* contenente i discorsi che per i singoli enunciati che sono stati successivamente isolati (sia quelli contenenti presupposizioni, sia quelli che ne sono privi). Le tendenze ottenute, per esempio, da Pépiot (2014), o da Yuan et al. (2007) e da Schötz (2007), risultano dunque confermate nello studio presente: anche nel campione analizzato gli uomini parlano a velocità maggiore rispetto alle donne, e i soggetti più anziani parlano più lentamente rispetto ai soggetti giovani, in modo statisticamente significativo. L'appartenenza dei politici a schieramenti di sinistra o centro-sinistra dà luogo a valori medi di velocità di eloquio e di articolazione superiori rispetto ad altri schieramenti, ma l'appartenenza politica non si rivela statisticamente significativa entro un modello di regressione lineare ottenuto a partire dai dati di questo studio. Allo stesso modo, la provenienza dalle regioni centrali dà spesso luogo a valori medi superiori di velocità di eloquio e di articolazione, ma questa variabile non rappresenta un predittore statisticamente rilevante nella spiegazione della variabilità dei dati all'interno del modello di regressione.

Emergono alcune lacune nella letteratura sulla velocità di eloquio in lingua italiana: per esempio, non è stato possibile trovare valori di riferimento per provenienza geografica. Potrebbe essere utile un'indagine sulle tendenze delle diverse varietà di italiano regionale in termini di velocità di eloquio e di articolazione, e sulle loro differenze, qualora esistano, che non si limiti alla divisione a grana grossa in Nord, Centro e Sud. Comunque, la maggior parte degli studi sui parametri temporali del parlato si concentra su altre lingue, prime fra tutte l'inglese; potrebbe essere interessante approfondire questa tematica in relazione all'italiano.

Il secondo *dataset*, riportato in Appendice, è costituito da 66 enunciati isolati attraverso il programma Praat (Boersma, 2001; Boersma & Weenink, 2022) a partire dai discorsi già analizzati. Di questi enunciati, la metà presentava presupposizioni, così come sono state annotate all'interno del *corpus* IMPAQTS, mentre l'altra metà ha svolto il ruolo di gruppo di controllo ai fini dell'analisi e non conteneva alcun tipo di impliciti.

Il tema degli impliciti all'interno dei discorsi politici, e della lingua politica in generale, è di particolare attualità e rilevanza nell'epoca contemporanea, con l'avvento delle nuove tecnologie di comunicazione di massa, che hanno rivoluzionato l'informazione e le modalità

con cui le personalità politiche si interfacciano con l'elettorato. Sia in formato audiovisivo, come nelle registrazioni per la televisione o le dirette sui *social network*, che attraverso la parola scritta, come nel caso dei *tweet* o dei post sulle pagine Facebook ufficiali, i discorsi hanno la potenzialità di raggiungere una platea di spettatori assai ampia e variegata. L'utilizzo di presupposizioni, implicature, topicalizzazioni e istanze di vaghezza può rientrare in strategie più ampie di carattere manipolatorio, per la capacità di questi fenomeni pragmatici di distogliere l'attenzione del destinatario del discorso dal messaggio che viene trasmesso implicitamente, che viene dunque passivamente accettato; oppure, essi possono allontanare la responsabilità nei confronti del messaggio da chi lo produce (Lombardi Vallauri 2019).

Una seconda analisi quantitativa non dà luogo a risultati rilevanti per quanto riguarda l'influenza di questi fenomeni pragmatici sulla velocità di eloquio e di articolazione. Sono nuovamente confermate tutte le tendenze che erano già emerse a partire dai discorsi, per quanto riguarda l'influenza di genere ed età, per la velocità sia di eloquio che di articolazione.

Nonostante lo studio presente non riscontri differenze di velocità tra gli enunciati contenenti presupposizioni e quelli che sono privi di qualsiasi tipo di impliciti entro il parlato politico, il legame tra velocità di eloquio e contenuti impliciti potrebbe essere ulteriormente approfondito mediante l'adozione di un campione più ampio di materiali linguistici, che potrebbero presentare durata maggiore, e mediante l'annotazione manuale, ove possibile, della presenza di contenuti nascosti. L'analisi potrebbe essere allargata ad altre tipologie di impliciti, la cui influenza sul parlato e sulle sue caratteristiche temporali potrebbe variare.

Lo studio ha comunque utilizzato una quantità ridotta di dati (20 discorsi); risultati più promettenti potrebbero essere ottenuti allargando il campione considerevolmente, grazie anche alla relativa facilità nel reperire le registrazioni di discorsi pubblici contemporanei. Sarebbe anche possibile un'indagine di tipo diacronico, per la disponibilità di discorsi risalenti alla seconda metà del secolo scorso, che indaghi eventuali cambiamenti nel tempo delle modalità con cui i discorsi politici sono pronunciati, cambiamenti legati anche ai diversi canali di comunicazione e al diverso rapporto che le personalità politiche desiderano instaurare con il proprio pubblico, attraverso la costruzione di un'immagine di sé. Anche in questo senso giocano un ruolo rilevante i nuovi canali di comunicazione di massa rappresentati dai *social media* come Facebook, Instagram e X (precedentemente Twitter).

Nel presente studio sono stati volutamente esclusi i discorsi tenuti *online*, sfruttando questi nuovi canali di comunicazione, sia in diretta che preregistrati. Essi sono comunque presenti e annotati all'interno del *corpus* IMPAQTS. Un'indagine specifica che tenga conto di questa tipologia di dati potrebbe essere molto utile, soprattutto in relazione al tema del populismo e

delle strategie di costruzione dell'immagine dei politici appartenenti a formazioni populiste, che si esplicano attraverso specifici metodi persuasivi.

Appendice

Enunciati contenenti impliciti

1. Sosteniamo questo sistema che dovrà anche collegare l'università alle industrie e alle piccole medie imprese [ATAJ_p1]
2. Dobbiamo anche innalzare il livello di formazione [ATAJ_p2]
3. Cercando di cancellare quella valle della morte che oggi c'è [ATAJ_p3]
4. Voi sapete benissimo che fare questo tipo di calcolo in Italia costituisce delle sperequazioni incredibili [BLOR_p1]
5. È evidente che c'è un freno fortissimo da parte di un apparato della pubblica amministrazione ogni volta che noi presentiamo norme per la semplificazione amministrativa [BLOR_p2]
6. Un dibattito in questa operazione di trasformazione che sta avvenendo nel mondo sanitario [BLOR_p3]
7. Per liberare le donne dal carico di cura che grava sproporzionatamente sulle loro spalle e le frena nel lavoro [ESCH_p1]
8. A partire dal sostegno alla preziosa attività che svolgono i centri antiviolenza e le case rifugio [ESCH_p2]
9. Dove sanno benissimo che non ci sono le strutture adeguate per accoglierli [ESCH_p3]
10. Sulle potenziali degenerazioni connesse alla strutturazione delle correnti e in magistratura [MACA_p1]
11. Se queste fossero esclusivamente luoghi di confronto luoghi di dialogo di condivisione di scambio di esperienze [MACA_p2]
12. Creando quelle condizioni indispensabili per un equilibrio istituzionale improntato alla separazione dei poteri e al rispetto dei ruoli [MACA_p3]
13. Allude al tentativo di trasformare Pompei in una sorta di pacchiana Disneyland [MORF_p1]
14. Dobbiamo rifiutare l'idea che per assumere sia necessario rendere più facili i licenziamenti [MORF_p2]
15. Battervi contro il rischio che il partito democratico diventi una coalizione di oligarchie [MORF_p3]
16. Siccome quasi tutto quello che sta cadendo sulla testa degli italiani dalle banche all'agricoltura alla pesca al commercio alle pensioni purtroppo deriva da scelte folli suicide di Bruxelles e di Strasburgo [MSAL_p1]
17. Che l'Italia torni a essere Italia a contare decidere accogliere chi vuole accogliere espellere chi vuole espellere [MSAL_P2]

18. E lo dedico a tutti quegli analisti falliti per cui la scelta della lega di crescere unire e presentarsi ovunque avrebbe rappresentato un problema al nord [MSAL_p3]
19. Questo taglio dei parlamentari costruito sulla falsa proposizione di un risparmio economico [PBIN_p1]
20. Con una protesta fioristica che non dà ragione del senso profondo di questo di questa proposta che noi stiamo facendo [PBIN_p2]
21. No alla fiducia che in qualche modo ti obbliga a ragionare in termini digitali o sì o no [PBIN_p3]
22. Che noi dobbiamo distinguere i privilegi dei politici dal costo della politica [RBIN_p1]
23. Io forse penso che quando Berlusconi riceveva i capi di stato nelle sue ville private faceva anche risparmiare [RBIN_p2]
24. Di coloro che continuano impropriamente a definirsi popolo della libertà o del buongoverno [RBIN_p3]
25. Seconda osservazione la questione più bruciante le adozioni [RBUT_p1]
26. Quelli che invece possono essere commissariati all'estero per essere inseriti sono tantissimi [RBUT_p2]
27. Questa insistenza sul tema della genitorialità ha un valore simbolico [RBUT_p3]
28. Per una ripresa di forza di sviluppo di energia della parte meridionale dell'Italia [RSPE_p1]
29. Per provare a sbloccare una situazione che così com'è oggi rischia di non essere più sostenibile [RSPE_p2]
30. E ancora ce lo segnalano gli ultimi drammatici dati consegnatici nei mesi precedenti dallo svimez [RSPE_p3]
31. Chi si permette di dire che mio figlio è impresentabile solo perché è figlio di un padre condannato in primo grado [SMUS_p1]
32. Nella consapevolezza che in Sicilia c'è una nuova coscienza antimafia [SMUS_p2]
33. È anche vero che noi abbiamo consentito per oltre centocinquanta anni che la mafia potesse attecchire potesse mettere radici [SMUS_p3]

Enunciati privi di impliciti

1. Rappresentano rappresentate un punto di riferimento del nostro sistema economico [ATAJ_n1]
2. Il credito cooperativo oggi è chiamato ad affrontare grandi sfide, la globalizzazione [ATAJ_n2]
3. E mi auguro che anche dopo la riforma il gruppo bancario cooperativo possa continuare il suo lavoro di prossimità [ATAJ_n3]
4. Ognuno di noi almeno una volta nella vita speriamo quando siamo nati abbiamo fatto ricorso al nostro sistema sanitario [BLOR_n1]

5. Noi abbiamo voluto con questi stati generali della salute fare il punto sul futuro e il presente della sanità italiana [BLOR_n2]
6. Noi veniamo da anni in cui il sistema è stato cambiato e ricambiato [BLOR_n3]
7. E allora per affrontarla bisogna agire in modo sistemico più su fronti quello normativo preventivo culturale economico [ESCH_n1]
8. Quando una donna non denuncia per paura di non essere creduta è il fallimento dello stato dell'intera società [ESCH_n2]
9. Serve formazione specifica del personale pubblico che viene a contatto con le donne che subiscono violenza [ESCH_n3]
10. Un saluto che desidero estendere a tutti coloro che stanno seguendo in streaming la cerimonia di apertura dell'anno accademico di questa prestigiosa università [MACA_n1]
11. Forte il richiamo al senso civico e alla responsabilità di tutti nell'interesse del prossimo di chi è più debole e della collettività [MACA_n2]
12. Penso allo sviluppo delle piattaforme digitali di comunicazione che ci permettono oggi di vivere questa preziosa cerimonia in sicurezza [MACA_n3]
13. Il crollo di Pompei, un fatto gravissimo innanzitutto per il valore simbolico che Pompei ha in tutto il mondo [MORF_n1]
14. Sono suggerimenti giusti, condivisibili, io però mi sento di dire che è esattamente quello che avete fatto in questi anni [MORF_n2]
15. Nessun imprenditore chiedeva l'abolizione dell'articolo diciotto perché di questo si tratta oggi [MORF_n3]
16. Sono contento di avere passato lì la mia serata di san valentino ed è un percorso che comincia col sud [MSAL_n1]
17. E a Pisa abbiamo eletto sia il deputato che il senatore nel collegio uninominale [MSAL_n2]
18. Almeno cinque milioni di persone ci hanno detto vai e fai [MSAL_n3]
19. Più parlamento non vuol dire più parlamentari vuol dire più qualità al lavoro del parlamento [PBIN_n1]
20. Per ribadire che se fossimo di meno si lavorerebbe meglio se fossimo di meno costeremmo meno [PBIN_n2]
21. In qualche modo scalza il Paese scalza le persone che hanno responsabilità di scelta [PBIN_n3]
22. Mi sembrava interessata all'inizio della trasmissione sulle ragioni del nostro risultato elettorale [RBIN_n1]
23. Sono state sostanzialmente due operazioni una aritmetica e l'altra algebrica [RBIN_n2]
24. E questo ha reso sicuramente molto difficile la nostra proposta [RBIN_n3]

25. Questa legge apre all'utero in affitto vai all'estero compri l'utero anzi compri il bambino affitti l'utero e compri il bambino [RBUT_n1]
26. Con la legge la legislazione italiana di adesso tu lo puoi adottare come tuo figlio [RBUT_n2]
27. Rimane un vuoto fondamentale nel processo di costruzione della sua identità da parte del bambino [RBUT_n3]
28. Si è da poco concluso un consiglio dei ministri in cui abbiamo approvato un provvedimento importante che dispone una serie di interventi di natura sanitaria [RSPE_n1]
29. Il primo articolo di questo provvedimento voglio ricordare è stato approvato pochi minuti fa dal consiglio dei ministri all'unanimità [RSPE_n2]
30. Nello stesso tempo le nostre autorità sanitarie sono al lavoro in modo particolare grazie al parere espresso dall'AIFA [RSPE_n3]
31. Quando ti alzi la mattina e senti il fiato della gente sulla nuca [SMUS_n1]
32. Se poi lo fai ogni giorno correndo il rischio di essere chiamato dall'autorità nazionale anticorruzione o dalla procura della repubblica [SMUS_n2]
33. Io provengo da un'isola che proprio minore non è, è la più grande isola del Mediterraneo [SMUS_n3]

Bibliografia

- Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of personality and social psychology*, 37(5), 715.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott. Int.*, 5(9), 341-345.
- Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer (Version 6.2.14) [Computer software].
- Bond, R. N., Feldstein, S., Simpson, S. (1988). Relative and absolute judgments of speech rates from masked and content-standard stimuli: The influence of vocal frequency and intensity. *Human Communication Research*, 14, 548-567.
- Bouvier, G., & Way, L. C. (2021). Revealing the politics in “soft”, everyday uses of social media: the challenge for critical discourse studies. *Social Semiotics*, 31(3), 345-364.
- Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta psychologica*, 67(2), 135-144.
- Brown, B. L., Strong, W. J., & Rencher, A. C. (1974). Fifty-four voices from two: the effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *The Journal of the Acoustical Society of America*, 55(2), 313-318.
- Buller, D. B., LePoire, B. A., Aune, R. K., & Eloy, S. V. (1992). Social perceptions as mediators of the effect of speech rate similarity on compliance. *Human Communication Research*, 19, 286-286.
- Buller, D. B., & Aune, R. K. (1992). The effects of speech rate similarity on compliance: Application of communication accommodation theory. *Western Journal of Communication (includes Communication Reports)*, 56(1), 37-53.
- Buller, D. B., & Aune, R. K. (1988). The effects of vocalics and nonverbal sensitivity on compliance: A speech accommodation theory explanation. *Human Communication Research*, 14(3), 301-332.
- Buller, D. B., & Burgoon, J. K. (1986). The effects of vocalics and nonverbal sensitivity on compliance: A replication and extension. *Human Communication Research*, 13(1), 126-144.
- Büring, D., (2011). “Focus and intonation”. In *The Routledge Companion to the Philosophy of Language*, 103-115. London: Routledge.
- Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31(2), 149-180.

- Byrne, D. (1997). An overview (and underview) of research and theory within the attraction paradigm. *Journal of Social and Personal Relationships*, 14(3), 417-431.
- Calzà L., Gagliardi G., Rossini Favretti R., Tamburini, F. (2021). Linguistic features and automatic classifiers for identifying mild cognitive impairment and dementia. *Computer Speech & Language* 65, 101-113.
- Cargile, A. C., Giles, H., Ryan, E. B., & Bradac, J. J. (1994). Language attitudes as a social process: A conceptual model and new directions. *Language & Communication*, 14(3), 211-236.
- Chafe, W. (1987). Cognitive constraints on information flow. *Coherence and grounding in discourse*, 11, 21-51.
- Combei, C. R., & Reggi, V. (2023). *Appraisal, Sentiment and Emotion Analysis in Political Discourse: A Multimodal, Multi-method Approach*. Taylor & Francis.
- Combei, C. R. (2023). Speaking Italian with a Twist: A Corpus Study of Perceived Foreign Accent.
- Combei, C.R. & Didoni, E. (2024). Beyond “I Didn't Do It”: A Linguistic Analysis of Denial in US Legal Settings. *Languages* 9(7), 1-20.
- Cominetti, F., Gregori, L., Lombardi Vallauri, E., & Panunzi, A. (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, 151-164.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Harvard University Press.
- Coschignano, S., Minnema, G., & Zanchi, C. (2023). Explaining the distribution of implicit means of misrepresentation: A case study on Italian immigration discourse. *Journal of Pragmatics*, 213, 107-125.
- Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion*, 12(5), 1085.
- Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. *Proceedings of 11th ICPHS*, Tallinn 5, 447–450.
- De Jong & Wempe 2009 De Jong, N.H. & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods* 41 (2), 385–390.

- De Jong, N.H., Pacilly, J., & Heeren, W. (2021). PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically, *Assessment in Education: Principles, Policy & Practice* 28:4, 456–476.
- De Mauro, T. (1982). *Minisemantica dei linguaggi non verbali e delle lingue*. Roma–Bari: Laterza.
- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and cognitive processes*, 5(4), 313-349.
- Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. *International Congress of Phonetic Sciences*, 471–474.
- Dellwo, V., Aschenberner, B., Wagner, P., Dancovicova, J., & Steiner, I. (2004). Bonntempo-corpus and bonntempo-tools: a database for the study of speech rhythm and rate. In *Interspeech* (pp. 777-780).
- Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski & I. Szigeti (Eds.). *Language and language-processing* (pp. 231–241). Frankfurt am Main: Lang.
- Dietrich, B. J., Hayes, M., & O'Brien, D. Z. (2019). Pitch perfect: Vocal pitch and the emotional intensity of congressional speech. *American Political Science Review*, 113(4), 941-962.
- Fairclough, N. (1993). Critical discourse analysis and the marketization of public discourse: The universities. *Discourse & society*, 4(2), 133-168.
- Fairclough, N. (2000). *New labour, new language?* Routledge.
- Fairclough, N. (2013). *Critical discourse analysis: The critical study of language*. Routledge.
- Feldstein, S., Dohm, F., & Crown, C. L. (1993). Gender as a mediator in the perception of speech rate. *Bulletin of the Psychonomic Society* 31(6), 521–524.
- Feldstein, S., Dohm, F., & Crown, C. L. (2001). Gender and Speech Rate in the Perception of Competence and Social Attractiveness. *The Journal of Social Psychology* 141:6, 785–806.
- Fetzer, A. (2014). I think, I mean and I believe in political discourse: Collocates, functions and distribution. *Functions of language*, 21(1), 67-94.
- Fillmore, Charles, J. (1982). Frame semantics. In *Linguistics in the Morning Calm*, ed. by The Linguistic Society of Korea, 111-137. Seoul: Hanshin.
- Flege, J. E. (1984). The detection of French accent by American listeners. *The Journal of the Acoustical Society of America*, 76(3), 692-707.
- Gagliardi G., Tamburini, F. (2021). Linguistic biomarkers for the detection of Mild Cognitive Impairment. *Lingue e linguaggio* 20(1), 3–31.

- Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3), 474-496.
- Gallois, C., Ogay, T., & Giles, H. (2005). Communication accommodation theory. *Theorizing about intercultural communication*, 121-148.
- Garassino, D., Brocca, N., & Masia, V. (2022). Is implicit communication quantifiable? A corpus-based analysis of British and Italian political tweets. *Journal of Pragmatics*, 194, 9-22.
- Giles, H., & Johnson, P. (1987). Ethnolinguistic identity theory: A social psychological approach to language maintenance. *International Journal of the Sociology of Language*, 68, 69-99.
- Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition*, 107(3), 1155-1164.
- Gosztolya, G., Vincze, V., Tóth, L., Pákási, M., Kálmán, J., & Hoffmann, I. (2019). Identifying mild cognitive impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. *Computer Speech & Language*, 53, 181-197.
- Gosztolya, G., Tóth, L., Grósz, T., Vincze, V., Hoffmann, I., Szatlóczki, G., ... & Kálmán, J. (2016). Detecting mild cognitive impairment from spontaneous speech by correlation-based phonetic feature selection. *Interspeech 2016*, 107-111.
- Grabe, E., Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology 7*, 515-546.
- Graham, T., Jackson, D., & Broersma, M. (2016). New platform, old habits? Candidates' use of Twitter during the 2010 British and Dutch general election campaigns. *New media & society*, 18(5), 765-783.
- Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41-58). Brill.
- Grice, H. P. (1978). Further notes on logic and conversation. In *Pragmatics* (pp. 113-127). Brill.
- Grice, H. P. (1993). *Logica e conversazione: Saggi su intenzione, significato e comunicazione*. Il Mulino.
- Harnsberger, J. D., Shrivastav, R., Brown Jr, W. S., Rothman, H., & Hollien, H. (2008). Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of voice*, 22(1), 58-69.

- Hoffmann, I., Nemeth, D., Dye, C. D., Pákási, M., Irinyi, T., & Kálmán, J. (2010). Temporal parameters of spontaneous speech in Alzheimer's disease. *International journal of speech-language pathology, 12*(1), 29-34.
- Huang, L. F., & Gráf, T. (2020). Speech Rate and Pausing in English: Comparing Learners at Different Levels of Proficiency with Native Speakers. *Taiwan Journal of TESOL, 17*(1), 57-86.
- Jaderberg, M., Simonyan, K., & Zisserman, A. (2015). Spatial transformer networks. *Advances in neural information processing systems, 28*.
- Jefferson, G. (2004), "Glossary of transcript symbols with an introduction". In: Lerner, Gene H. (ed.), *Conversation Analysis: studies from the first generation*, Amsterdam, John Benjamins, 13-31.
- Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of experimental psychology: learning, Memory, and cognition, 20*(4), 824.
- Jescheniak, J. D., Schriefers, H., & Hantsch, A. (2003). Utterance format effects phonological priming in the picture-word task: Implications for models of phonological encoding in speech production. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 441.
- Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. *Probabilistic linguistics, 21*, 1-30.
- Kapatsinski, V., Easterday, S., & Bybee, J. (2020). Vowel reduction: a usage-based perspective. *Rivista di Linguistica, 32*(1), 19-42.
- Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language, 45*, 326-347.
- Lakoff, R. T. (2000). *The language war*. University of California Press.
- Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences, 22*(1), 1-38.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical society of America, 35*(11), 1773-1781.
- Lohmann, A., & Conwell, E. (2020). Phonetic effects of grammatical category: How category-specific prosodic phrasing and lexical frequency impact the duration of nouns and verbs. *Journal of Phonetics, 78*, 100939.

- Lombardi Vallauri, E., & Masia, V. (2014). Implicitness impact: measuring texts. *Journal of Pragmatics*, 61, 161-184.
- Lombardi Vallauri, E. (2019). *La lingua disonesta*. Il Mulino.
- Low, L. , Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and speech*, 43(4), 377-401.
- Luraghi, S., & Stringa, P. (2024). *Comunicare la crisi: Metafore e cornici concettuali fra pandemia, guerra, immigrazione*. Carocci.
- Macagno, F. (2022). Argumentation profiles and the manipulation of common ground. The arguments of populist leaders on Twitter. *Journal of Pragmatics*, 191, 67-82.
- Maricchiolo, F., Bonaiuto, M., & Gnisci, A. (2014). Body movements in political discourse. In *Body-Language-Communication. An International Handbook on Multimodality in Human Interaction. Volume 2*. (pp. 1400-1412) de Gruyter Mouton.
- Masini, F., Combei, C. R., Cicchirillo, R. (2024). The Prosody of List Constructions. In K. Nikiforidou & M. Fried (Eds.), *Multimodal communication from a Construction Grammar perspective*, John Benjamins: Amsterdam, pp. 1-31.
- Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & new media*, 22(2), 205-224.
- Mauri, C., Ballarè, S., Goria, E., Cerruti, M., & Suriano, F. (2019). KIParla corpus: a new resource for spoken Italian. In *CEUR WORKSHOP PROCEEDINGS* (pp. 1-7). SunSITE Central Europe.
- Miller, N., Maruyama, G., Beaber, R. J., Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality & Social Psychology* 34, 615-624.
- Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research*, 44(4), 778–792.
- Murray, H. A. (1943). *Thematic Apperception Test*. Cambridge, MA: Harvard University Press.
- Navarrete, E., Basagni, B., Alario, F. X., & Costa, A. (2006). Does word frequency affect lexical selection in speech production?. *Quarterly Journal of Experimental Psychology*, 59(10), 1681-1690.
- Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. In M. van Oostendorp, C.J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology*. 5 vols. Malden, MA and Oxford: Wiley-Blackwell.
- O’Connell, D. C., Kowal, S., Bartels, U., Mundt, H., & Van De Water, D. A. (1989). Allocation of time in reading aloud: Being fluent is not the same as being rhetorical. *Bulletin of the Psychonomic Society*, 27, 223-226.

- Ordin, M., Polyanskaya, L., & Ulbrich, C. (2011, August). Acquisition of Timing Patterns in Second Language. In *Interspeech*, 1129-1132.
- Ordin, M., & Polyanskaya, L. (2015a). Perception of speech rhythm in second language: The case of rhythmically similar L1 and L2. *Frontiers in Psychology*, 6, 126049.
- Ordin, M., & Polyanskaya, L. (2015b). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America*, 138(2), 533-544.
- Paschen, L., Fuchs, S., & Seifart, F. (2022). Final lengthening and vowel length in 25 languages. *Journal of Phonetics*, 94, 101179.
- Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, 539-558.
- Pépiot, E. (2014). Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. *Speech Prosody*, 7, 305–309.
- Poggi, I., D'Errico, F., & Vincze, L. (2010, May). Types of Nods. The Polysemy of a Social Signal. In *LREC*.
- Polyanskaya, L., Ordin, M., & Busa, M. G. (2017). Relative salience of speech rhythm and speech rate on perceived foreign accent in a second language. *Language and speech*, 60(3), 333-355.
- Polyanskaya, L., & Ordin, M. (2019). The effect of speech rhythm and speaking rate on assessment of pronunciation in a second language. *Applied Psycholinguistics*, 40(3), 795-819.
- Priva, U. C. (2017). Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition*, 160, 27–34.
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023, July). Robust speech recognition via large-scale weak supervision. In *International conference on machine learning* (pp. 28492-28518). PMLR.
- Ramus, F., Nespors, M., Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265–292.
- Reichel, U. D. (2012). PerMA and Balloon: Tools for string alignment and text processing. In *Proc. Interspeech*.
- Rodero, E. (2012). A comparative analysis of speech rate and perception in radio bulletins. *Text & Talk*, 32(3), 391-411.

- Roloff, M. E. (1987). Communication and conflict. In C. R. Berger & S. H. Chaffee (Eds.), *Handbook of communication science* (pp. 484–534). Sage Publications, Inc.
- Rosenthal, R. (Ed.). (1979). *Skill in nonverbal communication*. Cambridge, MA: Oelgeschlager, Gunn, & Hain.
- Rychlý, Pavel (2007), “Manatee/Bonito – A Modular Corpus Manager”. In: *1st Workshop on Recent Advances in Slavonic Natural Language Processing*, Brno, Masaryk University, 65-70.
- Sahidullah, M., & Saha, G. (2012). Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. *Speech communication*, 54(4), 543-565.
- Sbisà, M. (2009). Linguaggio ragione interazione. Per una pragmatica degli atti linguistici. In *EUT Edizioni Università di Trieste eBooks*.
- Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. In: Proc. ICPhS. San Francisco, pp. 607–610.
- Schötz, S. (2007). Acoustic analysis of adult speaker age. *Speaker classification I: Fundamentals, features, and methods*, 88-107.
- Seifart, F., Strunk, J., Danielsen, S., Hartmann, I., Pakendorf, B., Wichmann, S., Witzlack-Makarevich, A., De Jong, N. H., Bickel, B. (2018). Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences of the United States of America*, 115(22), 5720–5725.
- Seifart, F., Ludger Paschen & Matthew Stave (eds.). (2022). *Language Documentation Reference Corpus (DoReCo) 1.2*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).
- Skoog Waller S., Eriksson M., Sörqvist P. (2015). Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in Psychology*, 6.
- Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of speech rate on personality perception. *Language and speech*, 18(2), 145-152.
- Sommers, M. S., Spehar, B., Tye-Murray, N., Myerson, J., & Hale, S. (2020). Age differences in the effects of speaking rate on auditory, visual, and auditory-visual speech perception. *Ear and Hearing*, 41(3), 549-560.
- Stewart, M. A., & Ryan, E. B. (1982). Attitudes toward younger and older adult speakers: Effects of varying speech rates. *Journal of language and social psychology*, 1(2), 91-109.

- Street, R. L., & Brady, R. M. (1982). Speech rate acceptance ranges as a function of evaluative domain, listener speech rate, and communication context. *Communication Monographs*, 49(4), 290–308.
- Street Jr, R. L., Brady, R. M., & Putman, W. B. (1983). The influence of speech rate stereotypes and rate similarity on listeners' evaluations of speakers. *Journal of Language and Social Psychology*, 2(1), 37-56.
- Street Jr, R. L. (1982). Evaluation of noncontent speech accommodation. *Language & Communication*, 2(1), 13-31.
- Stuart, A., & Kalinowski, J. (2015). Effect of delayed auditory feedback, speech rate, and sex on speech production. *Perceptual and motor skills*, 120(3), 747-765.
- Szekely, A., Damico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Arevalo, A., Vargha, A., & Bates, E. (2005). Timed action and object naming. *Cortex*, 41(1), 7–25.
- Thakerar, J. N., Giles, H., & Cheshire, J. (1982). Psychological and linguistic parameters of speech accommodation theory. *Advances in the social psychology of language*, 205, 205-255.
- Trotta, D., & Guarasci, R. (2021). How are gestures used by politicians? A multimodal co-gesture analysis. *IJCoL. Italian Journal of Computational Linguistics*, 7(7-1, 2), 45-66.
- Van Bergem, D. (1995), Acoustic and lexical vowel reduction, IFOTT, Amsterdam.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Vincze, V., Gosztolya, G., Tóth, L., Hoffmann, I., Szatlóczki, G., Bánréti, Z., ... & Kálmán, J. (2016). Detecting mild cognitive impairment by exploiting linguistic information from transcripts. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 181-187.
- Voghera M. (2012), “Chitarra, violino, banjo e cose del genere”, in Thornton A. M., Voghera M. (a cura di), *Per Tullio De Mauro: studi offerti dalle allieve in occasione del suo 80° compleanno*, Aracne, Roma, pp. 341-364.
- Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & psychophysics*, 67(6), 939-950.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of open source software*, 4(43), 1686.

- Wickham, H., Chang, W., & Wickham, M. H. (2016). Package ‘ggplot2’. *Create elegant data visualisations using the grammar of graphics. Version, 2(1)*, 1-189.
- Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm?. *The Journal of the Acoustical Society of America*, *127(3)*, 1559-1569.
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91(3)*, 1707-1717.
- Winter, B. (2020). *Statistics for linguists: An introduction using R*. Routledge.
- Wodak, R. (2015a). Critical discourse analysis, discourse-historical approach. *The international encyclopedia of language and social interaction*, 1-14.
- Wodak, R. (2015b). *The politics of fear: What right-wing populist discourses mean*. Sage.
- Yuan, J., Liberman, M., & Cieri, C. (2007). Towards an integrated understanding of speech overlaps in conversation. *ICPhS XVI, Saarbrücken, Germany*.
- Zanchi, C. (2022). Per una sistematizzazione delle ingiustizie perpetrate attraverso il discorso: la rappresentazione delle donne e dei migranti nei media italiani. *Lingue, scritture, potere*, 89.
- Zellers, M., Schuppler, B., & Clayards, M. (2018). Introduction, or: why rethink reduction?. In *Rethinking Reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*. De Gruyter Berlin.
- Zhao, Y. (1997). The effects of listeners' control of speech rate on second language comprehension. *Applied linguistics*, *18(1)*, 49-68.