

L'autografo del Presidente della Repubblica: un archetipo del discorso di fine anno prodotto mediante ADT

Lorenzo Bernardi ¹, Arjuna Tuzzi ²

¹ Dipartimento di Scienze Statistiche – Università di Padova – Italia

² Dipartimento di Sociologia – Università di Padova – Italia

Riassunto

In questo lavoro viene proposto un metodo per ottenere un modello ideale di discorso di fine anno del Presidente della Repubblica Italiana. La versione breve del discorso di fine anno vuole rappresentare in poche battute i tratti personali, le scelte comunicative e i contenuti caratteristici di un dato Presidente. Il corpus è costituito dai 60 discorsi di fine anno di tutti i 10 Presidenti della Repubblica italiana (1949-2008). Nell'ambito della ADT (analisi dei dati testuali) vengono utilizzati uno dopo l'altro alcuni strumenti di analisi tradizionali e disponibili nel software *TALTAC*²: la ricerca di poliformi di uso comune e di segmenti ripetuti, il metodo di riconoscimento delle specificità basato sul modello ipergeometrico e l'ordinamento di frammenti di testo attraverso misure *TFIDF*.

Abstract

A method to obtain an ideal model of the end-of-year address of the President of the Italian Republic is suggested in this study. The short version of the end-of-year address is aimed at mirroring in few words the personal traits, the communication choices and the distinctive contents of a given President. The corpus includes the 60 end-of-year addresses delivered by all the 10 Presidents of the Italian Republic (1949-2008). In the frame of the ATD (analysis of textual data) we employed step-by-step some traditional tools available in the software *TALTAC*²: the recognition of multi-words, compounds and repeated sequences of words, the "characteristic elements" method based on the hypergeometric model, and the ranking of text-chunks obtained by means of *TFIDF* measures.

Keywords: repeated sequences, characteristic elements, *TFIDF*, end-of-year address, institutional discourse

1. Introduzione

Nel 2007 un gruppo di ricerca coordinato da Michele A. Cortelazzo e Arjuna Tuzzi (2007) ha dato alle stampe il volume *Messaggi dal colle*, costituito da un insieme di contributi che si devono a ricercatori di varie discipline: linguisti, semiologi, sociologi, statistici, storici ed esperti di mezzi di comunicazione. Il loro impegno è dedicato a esaminare da angoli visuali e interpretativi diversi e, sperabilmente, complementari i 57 discorsi augurali di fine d'anno dei nove Presidenti della Repubblica succedutisi dal 1948 al 2006: Luigi Einaudi, Giovanni Gronchi, Antonio Segni, Giuseppe Saragat, Giovanni Leone, Sandro Pertini, Francesco Cossiga, Oscar Luigi Scalfaro, Carlo Azeglio Ciampi.

Le analisi compiute segnalano, per vari aspetti, tipicità espressive e contenutistiche pur all'interno di un tradizionale filo espositivo rituale. In larga misura si potrebbe affermare che, per quella preziosa occasione di rapporto con l'intera cittadinanza italiana, a ogni Presidente sembra corrispondere un modello comunicativo che implicitamente riflette la caratura umana, politica,

psicologica dell'autore. In sostanza quasi una compiuta manifestazione della propria personalità, in parte liberata da alcuni laccioli dell'ufficialità.

Tale affermazione può apparire a qualcuno, esaminando i singoli testi, in parte forzata in quanto ogni Presidente compone il proprio intervento ripercorrendo i fatti della politica, la gamma delle attese e degli impegni che la animano, le speranze e le difficoltà che la guidano e limitano; ovviamente i temi evocati, i problemi indicati e le soluzioni immaginate attengono strettamente al momento storico e alla sensibilità individuale; tuttavia se si guarda poi (anche visivamente, come rileva il contributo di Zotti Minici, 2007, che considera la resa televisiva – quando questa era possibile – dell'intervento) al tono espressivo, allo stile comunicativo, agli atteggiamenti assunti, non vi è dubbio che ci si trova di fronte, lungo la serie dei relativi interventi (si va da un minimo di 2 per Antonio Segni ai 7, corrispondenti alla durata del mandato, di tutti gli altri con l'eccezione di Einaudi che è stato colui che ha avviato l'iniziativa ma solo a partire dal secondo anno del suo mandato), a un *unicum* per ogni Presidente. Con riguardo a questi aspetti, in particolare alcuni degli autori dello studio, al fine di rendere più immediata la sintesi delle considerazioni tratte dall'analisi puntuale (statistica, linguistica, politologia, sociologica, storica ecc.) della pluralità dei testi, pervengono a definire ogni Presidente con una frase incisiva ma ampiamente illustrativa dell'uomo, dello stile, delle attenzioni (*Einaudi: il trionfo del noi; Gronchi: il padre educatore; Segni: il "noi" istituzionale; Saragat: la parcellizzazione dei cittadini; Leone: l'amministratore dello stato; Pertini: l'esplosione della convivialità; Cossiga: l'apologia del dobbiamo; Scalfaro: le diverse facce della familiarità; Ciampi: dalla prosa all'epica*) o all'estremo con un unico espressivo attributo: *austero, affettuoso, bonario, amichevole, ansioso, istituzionale*, anche se tutti appaiono guidati da uno spirito di fondo connotato dall'esigenza di risultare un *convinto e necessario educatore delle regole e dei valori* (Zijno, 2007; La Mendola and Sterchele, 2007).

In questo contributo ci si avventura in un'impresa quasi acrobatica, al limite dell'infedeltà verso gli autori e con ampio rischio di disconoscimento da parte dei colpiti. Ampliando per qualche via la definizione del profilo del Presidente, guardando alle parole e ai concetti usati, richiamando più i temi costantemente sollevati piuttosto che i fatti citati, tentando di adottare sintassi e forme della specifica retorica, si vuole costruire l'archetipo espositivo in qualche misura tipico di ogni Presidente. L'operazione è svolta estraendo dai testi di ciascuno (ovviamente con successo e prossimità agli originali fortemente dipendenti dal numero dei discorsi, dalla loro lunghezza, dalla loro similarità piuttosto che dalla loro varietà) espressioni emblematiche, argomenti verso cui si mostra peculiare sensibilità, costruzioni linguistiche e contenuti psicologici che dalle analisi statistiche risultano significativamente appartenenti a ogni Presidente. In qualche misura si va alla ricerca del discorso idealtipico, dell'ideale modello in quanto rappresentazione rarefatta della realtà, del messaggio profondo o latente che ciascuno avrebbe voluto fare se gli fosse stata concessa una sola brevissima decisiva occasione.

Lo abbiamo chiamato l'autografo del Presidente, la speranza è che nessuno lo giudichi apocrifio.

2. Corpus: i discorsi di fine anno dei Presidenti della Repubblica Italiana

I discorsi di fine anno di questo studio si riferiscono a dieci diversi Presidenti in un arco temporale che va dal 1949 al 2008. I Presidenti rappresentano un criterio di ripartizione del corpus in 10 subcorpora, uno per Presidente (Tab. 1). Rispetto alla ricerca citata nel precedente paragrafo (Cortelazzo and Tuzzi, 2007), sono stati presi in considerazione anche i primi tre discorsi dell'attuale Presidente Giorgio Napolitano; quindi, sono complessivamente 60, uno per anno, sette per Presidente con tre eccezioni: Einaudi (6), Segni (2) e Napolitano (3). Mediamente i

discorsi hanno una lunghezza di 1.600 occorrenze ma nel corso degli anni sono diventati via via sempre più ampi: il discorso più breve appartiene a Einaudi (150 occorrenze) e il più lungo a Scalfaro (5.012).

Ai fini di questa nuova analisi il corpus è stato organizzato anche in frammenti più piccoli. In questo lavoro con frammento si intende una porzione di corpus costituita da uno o più periodi nella quale si apre e si chiude un pensiero compiuto e che può essere estrapolata dal contesto e gestita come parte autonoma del discorso. La frammentazione del corpus è stata eseguita manualmente e ha generato 1.176 frammenti. Ogni frammento è identificato attraverso un codice univoco che permette di risalire al discorso (anno), al Presidente (nome) e alla posizione del frammento nel discorso (numero progressivo). La lunghezza dei frammenti è variabile perché dipende dalla struttura del discorso, dall'importanza del passaggio e dalla complessità dell'argomentazione. Si va da un minimo di 8 o 9 occorrenze (*lasciatemi dare per un momento spazio ai ricordi*, Scalfaro 1993; *se consideriamo l'anno trascorso troviamo motivi di fiducia*, Cossiga 1986) a un massimo di 436 (Leone 1973) e 446 occorrenze (Gronchi 1959). La lunghezza media dei frammenti è pari a 84 parole, la mediana è 69 parole, il 50% dei frammenti ha una lunghezza compresa tra 42 e 109 parole (primo e terzo quartile).

<i>Presidente</i>	<i>man-</i> <i>dato</i>	<i>num.</i> <i>discorsi</i>	<i>N</i> <i>occor-</i> <i>renze</i>	<i>V</i> <i>voca-</i> <i>bolario</i>	<i>hapax</i> <i>min</i>	<i>N</i> <i>max</i>	<i>N</i> <i>medio</i>	<i>N</i>	<i>num.</i> <i>fram-</i> <i>menti</i>
Luigi Einaudi	1948-55	6	1.203	558	381	150	260	201	13
Giovanni Gronchi	1955-62	7	5.829	1.794	1.166	388	1.252	833	44
Antonio Segni	1962-64	2	1.795	781	557	738	1057	898	13
Giuseppe Saragat	1964-71	7	8.476	2.238	1.338	465	1.929	1.211	104
Giovanni Leone	1971-78	7	7.379	2.041	1.248	262	1.604	1.054	83
Sandro Pertini	1978-85	7	15.592	2.743	1.443	1.340	3.746	2.227	213
Francesco Cossiga	1985-92	7	13.890	3.001	1.750	418	3.345	1.984	138
Oscar Luigi Scalfaro	1992-99	7	24.675	4.133	2.254	2.085	5.012	3.525	303
Carlo Azeglio Ciampi	1999-06	7	12.597	2.919	1.723	1.193	2.129	1.800	175
Giorgio Napolitano	in carica	3	5.706	1.757	1.130	1.713	2.204	1.902	90
<i>Corpus</i>	<i>1948-08</i>	<i>60</i>	<i>97.142</i>	<i>10.142</i>	<i>4.946</i>	<i>150</i>	<i>5.012</i>	<i>1.619</i>	<i>1.176</i>

Tabella 1: Articolazione e misure del corpus

3. Metodo

In letteratura sono disponibili alcuni metodi che permettono di estrarre da un corpus organizzato in subcorpora un numero limitato di testi in grado di rappresentare al meglio le caratteristiche tipiche di ciascun subcorpus. Per esempio, Lebart et al. (1998) propongono un primo metodo che sfrutta le specificità e un secondo basato sul calcolo delle distanze chiquadrato. Questo studio affronta un percorso in parte alternativo che sfrutta in cascata le specificità e gli strumenti di *text mining* (TFIDF). La procedura si articola in cinque fasi:

1. La prima fase ha come obiettivo quello di aumentare la portata informativa delle forme grafiche attraverso il riconoscimento di poliformi di uso comune (*popolo italiano, forze dell'ordine, Presidente della Repubblica*) e di segmenti ripetuti identici nel corpus (*buon anno, amici miei, care italiane cari italiani*). I poliformi di uso comune possono essere identificati facendo riferimento a liste. I segmenti ripetuti, invece, vengono generati mediante procedure automatiche in grado di elencare tutte le sequenze di almeno due parole ripetute almeno due volte nel corpus. Allo scopo di selezionare i segmenti pieni e completi di maggior interesse, l'elenco può essere ordinato utilizzando gli indici *IS* di Morrone (1996):

$$IS = \sum_{i=1}^q \frac{f_{segm}}{f_i} \cdot P \quad IS_{rel} = \frac{IS}{q^2}$$

dove f_{segm} è la frequenza del segmento di lunghezza q , f_i è la frequenza della i -esima forma del segmento e P è il numero di forme grafiche piene presenti nel segmento.

Il riconoscimento nel corpus dei poliformi di uso comune e dei segmenti più rilevanti (lessicalizzazione) agevola il passaggio da un vocabolario per forme a uno misto, che contiene forme e espressioni.

2. La seconda fase estrae dal nuovo vocabolario i tratti caratteristici di un Presidente utilizzando il tradizionale modello ipergeometrico (Lafon, 1980; Lebart et al., 1998; Bolasco, 1999; Tuzzi, 2003; Pauli and Tuzzi, 2009).

$$\Pr\{X = x \mid N, n_i, n_j\} = \frac{\binom{n_i}{x} \binom{N - n_i}{n_j - x}}{\binom{N}{n_j}} \quad (1)$$

dove X è la variabile aleatoria che rappresenta la frequenza della i -esima forma/espressione nel subcorpus del j -esimo Presidente e i parametri N, n_i, n_j sono, rispettivamente, le occorrenze del corpus, della i -esima forma/espressione, del subcorpus del j -esimo Presidente. Il calcolo del p -value:

$$\sum_{n_{ij} \leq x \leq n_i} \Pr\{X = x \mid N, n_i, n_j\} < \alpha, \quad (2)$$

dove n_{ij} è la frequenza osservata i -esima forma/espressione nel subcorpus del j -esimo Presidente, permette di identificare le forme/espressioni che risultano significativamente sovrautilizzate dal Presidente rispetto all'intero corpus. Vengono riconosciute come specifiche di un Presidente quelle con p -value inferiore alla soglia fissata convenzionalmente in 2,5%.

Il subcorpus di ogni Presidente è costituito da diversi discorsi (sette nella maggior parte dei casi) e, quindi, l'occorrenza di una forma/espressione nel suo subcorpus è il risultato della somma delle occorrenze in ciascuno dei suoi discorsi. Il calcolo delle specificità permette solo di riconoscere forme/espressioni che risultano sovrautilizzate nel subcorpus del Presidente rispetto a tutti gli altri ma, se una forma/espressione è ripetuta molte volte solo in uno dei suoi discorsi, può essere erroneamente riconosciuta come significativa e, quindi, come specifica del Presidente. Utilizzare una misura della dispersione all'interno del subcorpus (Baayen, 2001; Tuzzi and Tweedie, 2000; Lebart, 2004; Pauli and Tuzzi, 2009) è importante per capire se le parole/espressioni riconosciute dal modello come specifiche per un Presidente sono concentrate in uno o pochi discorsi. Da un punto di vista tecnico la questione non è diversa dalla ricerca delle specificità e si può applicare nuovamente il modello (1). In questo caso, però, il subcorpus del Presidente assume il ruolo di corpus e i discorsi quello di subcorpora: nella (2) i parametri N, n_i, n_j sono, rispettivamente, le occorrenze del subcorpus del Presidente, della i -esima forma/espressione, del discorso j -esimo e n_{ij} è la frequenza osservata della i -esima forma/espressione nel discorso j -esimo del Presidente. Fissata nuovamente per il p -value una soglia del 2,5%, la logica va rovesciata: prese in considerazione solo le forme/espressioni già riconosciute come specifiche del Presidente, vengono a questo punto rigettate quelle che risultano specifiche per qualcuno dei suoi discorsi.

Il risultato di questo doppio controllo permette la costruzione di una lista L di forme/espressioni che rappresentano bene lo stile e le scelte personali del Presidente perché specifiche del Presidente rispetto agli altri ma non specifiche di qualcuno dei suoi discorsi.

3. La terza fase muove dalla lista L di forme/espressioni ed estrae dal corpus i frammenti più pertinenti per il Presidente, cioè le porzioni di corpus che con maggiore densità ne contengono le forme/espressioni specifiche. Per l'estrazione dei frammenti si utilizza un indice della famiglia $TFIDF$ che è costituito da una parte TF (*term frequency*, occorrenze della forma/espressione) e una parte IDF (*inverse document frequency*, logaritmo del rapporto tra il numero M di frammenti del corpus e il numero m di frammenti che contengono quella forma/espressione). Tra le numerose varianti disponibili in letteratura (Salton, 1989; Salton and Buckley, 1988), per questa applicazione si è scelta una versione normalizzata che annulla l'effetto della diversa lunghezza dei frammenti. Il valore del $TFIDF_i$ di un frammento con riferimento a una forma/espressione i si ottiene dividendo il $TFIDF$ per la radice della somma dei quadrati dei $TFIDF_j$ di tutte le forme/espressioni j contenute nel frammento:

$$TFIDF_i = \frac{tf_i \cdot \log \frac{M}{m_i}}{\sqrt{\sum_j \left(tf_j \cdot \log \frac{M}{n_j} \right)^2}}$$

Il $TFIDF$ di un frammento con riferimento a una lista L di forme/espressioni si ottiene come somma dei $TFIDF_i$ di tutte le forme/espressioni contenute nella lista

$$TFIDF = \sum_{i \in L} TFIDF_i$$

e fornisce una misura di quanto il frammento è pertinente rispetto alla lista o, in altri termini, di quanto il frammento è rappresentativo dello stile e delle scelte personali del Presidente.

4. Nella quarta fase si ricompono *ex post* un discorso del Presidente a partire dai frammenti riconosciuti come più pertinenti. Sulla base del valore assunto dal $TFIDF$ è possibile, infatti, ordinare i frammenti del corpus in ordine di rilevanza. Inoltre, dato che nella creazione dei frammenti è stato mantenuto un codice che tiene conto dell'ordine, tra i frammenti più pertinenti è possibile utilizzare il criterio di ordinamento per ricostruire la struttura originale del discorso. Questo accorgimento permette, per esempio, di ritrovare le espressioni tipiche per l'apertura e la conclusione del discorso. Per ottenere una versione condensata del discorso di un Presidente è sufficiente fissare una lunghezza massima (in questo caso è stato scelto un discorso breve di 700-800 parole) ed eliminare eventuali frammenti ridondanti.

4. Risultati

Tutti i risultati descritti in questo paragrafo sono stati ottenuti attraverso procedure disponibili nel software *TaLTaC*² (Bolasco et al., 2008).

Attraverso la lista dei poliformi di uso comune e la selezione dei segmenti ripetuti più interessanti (un migliaio tra quelli con *IS* più elevato), sono entrate a far parte del vocabolario 1.093 espressioni per un totale di 4.339 occorrenze. Dopo questa operazione il vocabolario risulta costituito da 11.045 entrate. Il calcolo delle specificità ha portato alla creazione di 10 liste, una per

ogni Presidente, contenente l'elenco di forme/espressioni particolarmente utili a rappresentarne lo stile e le scelte personali. Oltre ai due passaggi già descritti (selezione delle forme/espressioni specifiche del subcorpus del Presidente ed esclusione delle forme specifiche di qualche discorso) in questa fase è stata operata un'ulteriore selezione per ridurre la lista. I problemi di riduzione di queste liste saranno discussi nel prossimo paragrafo. Di seguito sono riportate le liste di Pertini e Ciampi (rispettivamente, 67 e 71 elementi) con tra parentesi le occorrenze nel subcorpus/corpus e il relativo *p-value*.

Pertini: *io* (176/297, 0.000); *ebbene* (46/56, 0.000); *terrorismo* (47/61, 0.000); *popolo italiano* (52/95, 0.000); *mila* (18/18, 0.000); *giovani* (61/140, 0.000); *questi giovani* (16/16, 0.000); *terroristi* (16/19, 0.000); *italiane e italiani* (13/14, 0.000); *guerra* (31/64, 0.000); *amici miei* (11/11, 0.000); *giovani che mi ascoltate* (9/9, 0.000); *studenti* (14/21, 0.000); *vittime* (16/27, 0.000); *miliardi* (8/8, 0.000); *diritti civili ed umani* (8/8, 0.000); *per dannata ipotesi* (8/8, 0.000); *nostra gioventù* (8/8, 0.000); *disoccupazione* (23/50, 0.000); *due superpotenze* (9/10, 0.000); *fame* (12/17, 0.000); *ricevo* (7/7, 0.000); *ricevuti* (7/7, 0.000); *italiani e italiane* (7/7, 0.000); *barriera contro il terrorismo* (7/7, 0.000); *fine dell'umanità* (7/7, 0.000); *ascoltate* (8/9, 0.000); *ordigni* (6/6, 0.000); *spendono* (6/6, 0.000); *tutte le mattine* (6/6, 0.000); *movimento operaio* (6/6, 0.000); *al vostro fianco* (6/6, 0.000); *gioinezza* (7/8, 0.000); *anziani* (12/23, 0.000); *scolaresche* (5/5, 0.000); *denutrizione* (5/5, 0.000); *come fossimo antichi amici* (5/5, 0.000); *nostro pianeta* (5/5, 0.000); *nostra esperienza* (5/5, 0.000); *mio animo* (8/12, 0.000); *armate* (4/4, 0.001); *atomica* (4/4, 0.001); *tonnellate* (4/4, 0.001); *disarmo totale e controllato* (4/4, 0.001); *sarebbe la fine dell'umanità intera* (4/4, 0.001); *creature umane* (4/4, 0.001); *questo terrorismo* (4/4, 0.001); *mio saluto* (5/7, 0.002); *indifferenti* (4/5, 0.003); *tutti i popoli della terra* (4/5, 0.003); *attentato* (3/3, 0.004); *angosciato* (3/3, 0.004); *morendo* (3/3, 0.004); *fame nel mondo* (3/3, 0.004); *per costruire ordigni di morte* (3/3, 0.004); *non armate la vostra mano* (3/3, 0.004); *io credo nella nostra gioventù* (3/3, 0.004); *morendo di fame* (3/3, 0.004); *io credo nel popolo italiano* (3/3, 0.004); *muoiono* (4/6, 0.007); *tutti gli uomini* (4/7, 0.014); *nostri giovani* (4/7, 0.014); *noi anziani* (3/4, 0.015); *credo* (12/41, 0.020); *milioni* (7/19, 0.020); *droga* (6/15, 0.021).

Ciampi: *Italia* (80/306, 0.000); *Unione Europea* (15/21, 0.000); *identità* (10/12, 0.000); *istituzioni* (27/75, 0.000); *Euro* (9/11, 0.000); *europei* (12/19, 0.000); *Europa* (39/139, 0.000); *care italiane cari italiani* (6/6, 0.000); *Risorgimento* (6/7, 0.000); *buongoverno* (5/5, 0.000); *mia moglie* (5/5, 0.000); *generazione* (7/10, 0.000); *nostra storia* (7/10, 0.000); *fiducia* (26/90, 0.000); *unità* (12/27, 0.000); *confronto* (10/20, 0.000); *prestigio* (7/11, 0.000); *spirito* (16/47, 0.000); *istituzione* (5/6, 0.000); *provincia* (5/6, 0.000); *mercati* (4/4, 0.000); *globalizzazione* (4/4, 0.000); *amata* (4/4, 0.000); *mantenimento della pace* (4/4, 0.000); *comunale* (4/4, 0.000); *ideali* (9/19, 0.000); *orgoglio* (6/9, 0.000); *patria* (20/70, 0.000); *tutto il cuore* (5/7, 0.001); *futuro* (10/25, 0.001); *imprese* (7/14, 0.001); *conflitti* (7/14, 0.001); *famiglia* (12/35, 0.001); *orgogliosi* (4/5, 0.001); *amor di patria* (4/5, 0.001); *slancio* (5/8, 0.002); *accrescere* (3/3, 0.002); *straordinarie* (3/3, 0.002); *tricolore* (3/3, 0.002); *coesa* (3/3, 0.002); *intero mondo* (3/3, 0.002); *italiane ed europee* (3/4, 0.002); *provincia italiana* (3/3, 0.002); *padri* (6/12, 0.002); *principio* (5/9, 0.003); *nostra civiltà* (4/6, 0.004); *augurale* (4/6, 0.004); *favorire* (4/6, 0.004); *radici* (6/13, 0.004); *città* (9/27, 0.006); *sentimenti* (6/14, 0.006); *sogno* (4/7, 0.008); *sfide* (4/7, 0.008); *nuove generazioni* (4/7, 0.008); *costruendo* (3/4, 0.009); *prevalga* (3/4, 0.009); *capitale* (3/4, 0.009); *rispetto dei diritti* (3/4, 0.009); *Europa unita* (3/4, 0.009); *nostri confini* (3/4, 0.009); *Costituzione* (14/56, 0.011); *unione* (4/8, 0.013); *consolidato* (3/5, 0.019); *migliorare* (3/5, 0.019); *cresce* (3/5, 0.019); *felice* (3/5, 0.019); *unità nazionale* (5/13, 0.020); *rafforzare* (4/9, 0.022); *europea* (6/18, 0.023); *dialogo* (11/44, 0.023); *valori* (11/44, 0.023).

Calcolando il *TFIDF* basato sulle liste *L* di specificità, si ottiene un *ranking* dei frammenti del corpus per ciascun Presidente. Tab. 2 riporta i valori dei primi 30 frammenti per le liste di Pertini, Ciampi e Napolitano.

Fissata una lunghezza *a priori* per il discorso che si vuole ottenere, selezionando un insieme di frammenti con *TFIDF* elevato, utilizzando come guida l'ordine dei frammenti e, infine, eliminando eventuali frammenti ridondanti, è abbastanza facile ricostruire come un *collage* un vero e proprio condensato del Presidente. Si perviene, quindi, a un modello ideale.

I due esempi seguenti riguardano Pertini e Ciampi.

<i>Pertini</i>		<i>Ciampi</i>		<i>Napolitano</i>	
<i>Frammento</i>	<i>TFIDF</i>	<i>Frammento</i>	<i>TFIDF</i>	<i>Frammento</i>	<i>TFIDF</i>
0354 Pertini	3,552	0984 Ciampi	3,023	1130 Napolitano	2,883
0277 Pertini	3,534	0943 Ciampi	2,823	1105 Napolitano	2,841
0375 Pertini	3,395	0935 Ciampi	2,792	1095 Napolitano	2,772
0303 Pertini	3,366	0920 Ciampi	2,766	1099 Napolitano	2,647
0405 Pertini	3,149	0940 Ciampi	2,750	1121 Napolitano	2,608
0309 Pertini	3,035	1030 Ciampi	2,682	1114 Napolitano	2,604
0313 Pertini	2,971	1025 Ciampi	2,568	1073 Ciampi	2,557
0467 Pertini	2,863	1083 Ciampi	2,564	1135 Napolitano	2,545
0461 Pertini	2,796	1007 Ciampi	2,563	1129 Napolitano	2,527
0339 Pertini	2,701	0918 Ciampi	2,560	0210 Leone	2,526
0399 Pertini	2,614	1073 Ciampi	2,556	1156 Napolitano	2,522
0437 Pertini	2,548	1021 Ciampi	2,530	1169 Napolitano	2,468
0340 Pertini	2,547	1070 Ciampi	2,519	0133 Saragat	2,436
0415 Pertini	2,487	0921 Ciampi	2,430	1090 Napolitano	2,408
0338 Pertini	2,474	1089 Napolitano	2,406	1098 Napolitano	2,385
0374 Pertini	2,388	0951 Ciampi	2,389	0046 Gronchi	2,382
0412 Pertini	2,371	0985 Ciampi	2,338	0091 Saragatr	2,377
0370 Pertini	2,359	0916 Ciampi	2,337	0234 Leone	2,368
0287 Pertini	2,347	1066 Ciampi	2,306	0524 Cossiga	2,361
0383 Pertini	2,329	0050 Gronchi	2,190	1138 Napolitano	2,360
0310 Pertini	2,305	1013 Ciampi	2,181	1104 Napolitano	2,360
0439 Pertini	2,256	0564 Cossiga	2,179	0771 Scalfaro	2,359
0272 Pertini	2,225	0805 Scalfaro	2,154	1031 Ciampi	2,356
0424 Pertini	2,188	1012 Ciampi	2,141	1089 Napolitano	2,350
0326 Pertini	2,177	0944 Ciampi	2,141	1137 Napolitano	2,336
0294 Pertini	2,143	1035 Ciampi	2,137	0014 Gronchi	2,332
0260 Pertini	2,142	1054 Ciampi	2,111	1157 Napolitano	2,327
0410 Pertini	2,138	1031 Ciampi	2,098	0533 Cossiga	2,321
0292 Pertini	2,129	1022 Ciampi	2,081	1133 Napolitano	2,313
0468 Pertini	2,121	0945 Ciampi	2,065	1113 Napolitano	2,288

Tabella 2: Primi 30 frammenti in ordine di TFIDF per le liste di Pertini, Ciampi e Napolitano

4.1. Condensato Pertini

«Italiane e italiani, miei connazionali, vi confesso che sono titubante a parlarvi, perché non vorrei turbare la vostra festa in attesa del nuovo anno con note di tristezza.

Il mio pensiero si rivolge a coloro che nel mondo lottano contro la fame e contro la miseria. Vedete, italiani e italiane, mentre io vi parlo milioni e milioni di creature umane stanno morendo di fame. Questa strage di innocenti pesa come una condanna sulla coscienza di tutti gli uomini di Stato, e quindi anche sulla mia coscienza.

Si spendono miliardi per costruire ordigni di guerra, che se per dannata ipotesi fossero usati sarebbe la fine dell'umanità intera del nostro pianeta. E mentre si spendono miliardi per costruire ordigni di morte, vi sono migliaia e migliaia di creature umane che nel mondo stanno morendo di fame. Ebbene, le due superpotenze cerchino di trovare un accordo. Lo chiedono tutti i popoli della terra. Ansiosi noi attendiamo che si seggano intorno ad un tavolo i dirigenti di queste due nazioni per discutere. Discutano e cerchino di trovare un accordo. Sì, sembra un sogno il mio, quando io dico che bisognerebbe arrivare al disarmo totale e controllato. Con le guerre nulla si risolve. Ed il denaro che oggi si spende e si sperpera, per me, per costruire ordigni di morte che, se domani, per dannata ipotesi fossero usati scomparirebbe l'umanità dal nostro pianeta, e questi denari si usino invece, per sollevare dalla fame tanti esseri umani, per combattere la fame nel mondo.

Vi è questa preoccupazione quindi della guerra nucleare, della guerra atomica. Fanno bene i giovani a preoccuparsi. D'altra parte hanno ragione: si preoccupano del loro avvenire. Noi siamo al tramonto della nostra esistenza, loro hanno dinanzi a sé tutta una vita. Ed hanno ragione di voler vivere in pace la loro esistenza che non sia turbata

dalla guerra che sarebbe, ripeto, l'ultima guerra sarebbe, perché sarebbe la fine dell'umanità. Quindi io sono con questi giovani.

Credo nei giovani, lo vado sempre ripetendo, prendo contatto con i giovani. Centinaia di giovani hanno preso contatto con me quando ero Presidente della Camera dei Deputati. Io ho instaurato qui al Quirinale il metodo di ricevere tutte le mattine, quando l'anno scolastico è aperto, scolaresche di ogni parte d'Italia. Ebbene, da questi giovani io mi sono sempre sentito e mi sento porre delle domande molto serie. Non faccio loro dei discorsi, intreccio con loro una conversazione, un dialogo, come fossimo antichi amici, e ci riesco sempre. Io credo nella nostra gioventù, a differenza di molti anziani che in questo momento mi ascoltano. La stragrande maggioranza della nostra gioventù è politicamente e moralmente sana.

Altro motivo di preoccupazione, di questa mia tristezza, è il terrorismo, che non dà pace al popolo italiano, questo terrorismo che si scatena, che turba la vita del nostro paese. Un giorno sapremo chi è che manovra questi terroristi, chi è che vuole destabilizzare il regime democratico italiano, chi vuole distruggere questa nostra Repubblica democratica, la cui conquista molto è costata al popolo italiano. Io vi confesso, amici miei, che non ho più l'animo di recarmi a rendere omaggio ai caduti, poliziotti e carabinieri che sono caduti sotto i mitra dei terroristi. Le corone, i telegrammi di condoglianze non possono ridare la vita ai morti e non possono placare il dolore dei vivi. Ma noi dobbiamo pensare a queste forze dell'ordine così benemerite, e che sono in lotta, ripeto, per difendere noi, per difendere il popolo italiano e la patria dalla violenza e dal terrorismo. Ma il terrorismo è stato soprattutto vinto dalla volontà del popolo italiano. Cioè il popolo italiano ha fatto barriera contro il terrorismo.

Io ai giovani questo io dico, e martello su questo punto, costituiscono l'avvenire del popolo italiano: battetevi sempre per la libertà, per la pace e per la giustizia sociale. Vi ripeto forse quello che ho detto l'anno scorso, dovete scusarmi: non armate la vostra mano, armate invece il vostro animo di un grande ideale e di una grande e vigorosa fede. Ed allora continuate la vostra strada. Ed io che sono ormai al tramonto della mia vita resterò sempre al vostro fianco finché un anelito, finché un alito di vita mi animerà.»

4.2. Condensato Ciampi

«Care italiane, cari italiani, come tutti voi, festeggerò fra poco l'inizio del nuovo anno. Sarà un anno felice, se ognuno, nelle proprie responsabilità, lo affronterà con fiducia in se stesso e nei valori in cui crede, con fedeltà alla propria coscienza, agli ideali fondanti della vita democratica.

I punti cardinali del mio procedere sono la coscienza dell'identità italiana; l'impegno per la libertà e per la giustizia sociale; il senso dell'appartenenza all'Europa; l'impegno per la pace nel mondo. Penso che la memoria del passato sia il terreno su cui cresce l'unità della nazione.

Mi sono stati già nel passato affidati delicati incarichi nelle istituzioni italiane ed europee, ma sto ora verificando quanto sia più complessa e impegnativa la responsabilità che la nostra Costituzione attribuisce al Capo dello Stato. Interpretare ed esprimere, con passione civile e con assoluta imparzialità, sentimenti e valori condivisi, esigenze e bisogni che riflettono l'interesse generale del paese. E guardare sempre all'unità nazionale come bene primario da tutelare e consolidare.[Napolitano]

L'uomo di Governo deve reagire mirando a realizzare una più forte coesione fra tutti coloro che sanno come si costruisce la pace; rafforzando le istituzioni che abbiamo creato in applicazione coerente dei nostri valori: la nostra Repubblica, l'Unione Europea, le Nazioni Unite. Istituzioni garanti, con gli strumenti che i popoli loro affidano, della convivenza civile, del progresso e della dignità di tutti. Nell'impegno di svolgere al meglio il compito che mi è stato affidato dal Parlamento, il mio fine è di accrescere la fiducia nelle istituzioni; di stimolarne il buon funzionamento; di favorire la distensione tra le forze politiche; di garantire il rispetto della Costituzione, sulla quale ho giurato. Ho fiducia nell'Italia. Ho fiducia nel popolo italiano. Ho fiducia nelle istituzioni che ci siamo liberamente date. E ho fiducia nell'Europa, la nostra nuova patria più grande, che stiamo costruendo.

Stiamo scrivendo insieme con gli altri parlamenti nazionali e con il Parlamento europeo la carta dei diritti fondamentali dell'unione. Stiamo, insomma, creando l'Europa dei cittadini e delle istituzioni, dopo quella dei mercati. Di quest'Europa l'Italia ha voluto e ha saputo essere parte. L'Italia non è mai mancata a nessuno dei momenti fondamentali del processo di integrazione europea. Non è mancata all'appuntamento dell'Euro. La generazione a cui io appartengo, che combatté e soffrì l'ultima e più grande strage della storia moderna, ha realizzato un sogno. Abbiamo portato in tutta Europa pace e democrazia. Noi ne siamo fieri, quando vediamo i nostri giovani vivere con tanta naturalezza la libertà e l'amicizia fra tutti gli europei. Ma l'opera va continuata: sta a voi giovani portarla a compimento. Siete nati, vivete in un'Europa di pace, di libertà. Tenete alti, e diffondete nel mondo, i suoi ideali.

Toccherà a voi completarne e rafforzarne le istituzioni. Per tutti gli Europei non c'è un domani se non in un'Unione Europea più coesa.

Unione Europea, significa pace in Europa. Questo gli italiani lo sanno, lo sentono. Insieme con le istituzioni, e ancor più in fretta, cresce il sentimento di identificazione con l'Europa nell'animo dei nostri giovani: che viaggiano, e sono ormai moltitudini, da un paese all'altro; che studiano qui o in altri paesi europei continuando a sentirsi ovunque a casa loro. Diventando più europei non si sentono sicuramente meno italiani. Un'identità più complessa è anche un'identità più ricca e più forte.

Oggi ci sentiamo europei, ma anche orgogliosamente italiani. Da tempo non era così forte l'attaccamento dei cittadini, in ogni parte d'Italia, ai simboli della nostra nazione: il tricolore, l'inno risorgimentale di Mameli, la Costituzione. Ovunque, dalle Alpi alla Sicilia, sento presente e crescente un forte patriottismo. Ad esso ho dato espressione, e ho trovato immediata, spontanea rispondenza. Può apparire singolare che proprio ora che si manifesti più intenso il nostro patriottismo, si rafforzino anche altri sentimenti: una maggiore consapevolezza di appartenere alla più grande patria europea; e una più forte coscienza dell'identità regionale e comunale. Non c'è contraddizione alcuna fra amore della propria città e regione, amor di patria, amore d'Europa. Io amo, insieme, la mia Livorno, la Toscana, l'Italia, l'Europa. Siamo il paese delle cento città.

Vengo ai problemi economici. So bene che quest'anno molte famiglie hanno avuto difficoltà con il loro bilancio, hanno fatto fatica. Il troppo lungo ristagno dell'economia, in Italia e in Europa, ha colpito soprattutto i più deboli. Andiamo incontro al nuovo anno incoraggiati dai primi segni di ripresa economica. Questi segni dobbiamo ora saperli sostenere con l'azione di tutti: imprenditori, lavoratori, istituzioni di Governo centrali e locali. Una politica di aperto, leale confronto tra istituzioni, imprenditori, lavoratori, che rilanci una capacità d'intesa che non mancò neppure in anni di grandi scontri ideologici, ci aiuterà a realizzare quello scatto di orgoglio, quel risveglio della fiducia di cui l'Italia ha bisogno.

Care italiane cari italiani. Non viviamo tempi tranquilli. Ma proprio per questo dobbiamo rinsaldare la nostra coesione, la nostra unità. Vi rivolgo, e con me mia moglie, un forte, affettuoso augurio. Che nel nuovo anno possiate realizzare le vostre speranze, i vostri progetti, qualcuno almeno dei vostri sogni. E che Iddio protegga le nostre famiglie, e la nostra amata patria.»

5. Discussione e conclusioni

Nella procedura in cinque fasi proposta in questo studio restano molti spazi di discrezionalità e alcune questioni aperte. L'analisi delle specificità e il *ranking* basato sul *TFIDF* permettono l'identificazione dei frammenti più adatti per la composizione del condensato ma, al fine di pervenire a un prodotto coerente e ben strutturato, permangono problemi di selezione dei frammenti (sia in termini di numero di unità da tenere in considerazione sia di varietà dei contenuti che si desidera veicolare) e di assemblaggio delle varie parti (con particolare attenzione agli elementi connettivi e coesivi dei diversi frammenti).

A monte è necessario formare liste ben compatte di specificità e, quindi, effettuare un'ulteriore selezione delle centinaia di parole/espressioni che in un corpus di queste dimensioni possono potenzialmente risultare significative per ogni Presidente. Fare affidamento solo su soglie meno permissive per i *p-value* non è necessariamente l'unica strada da percorrere perché si rischia di perdere informazioni interessanti (soprattutto tra i poliformi e i segmenti che sono gli elementi più preziosi ma anche i più rarefatti). Oltre alla riduzione del numero di specificità, la lista pone anche un problema semantico che, in teoria, si potrebbe tradurre in una procedura di selezione delle sole parole/espressioni pertinenti a una rosa di temi decisi *a priori* o nella realizzazione di distinte liste a tema. In questo studio, date le difficoltà di effettuare una sistematica classificazione semantica di parole/espressioni isolate dal contesto d'uso, si è preferito lavorare con un'unica lista per Presidente, agendo con criterio statistico sulle soglie dei *p-value*.

A valle è necessario assemblare i diversi frammenti in un prodotto coerente e coeso. Grazie alla presenza nel codice identificativo di un numero progressivo che tiene memoria delle posizioni

reciproche dei frammenti nei discorsi originali, è stato relativamente semplice ottenere buone formule per l'apertura e la conclusione del condensato; più complesso è stato, viceversa, gestire il flusso dei frammenti e, quindi, dei diversi argomenti al suo interno. In questo caso si è scelto di assegnare un'etichetta semantica a tutti i frammenti con *TFIDF* elevato (indicativamente i primi cento nel *ranking*) al fine di identificare il tema trattato (pace, terrorismo, guerra fredda, Costituzione, Europa, patriottismo ecc.) e poi, all'interno di ogni tema, sono stati selezionati uno o due frammenti tra quelli con *TFIDF* più elevato. La selezione è stata operata discrezionalmente all'interno di una rosa di frammenti molto simili tra di loro nei contenuti e così ripetitivi da risultare ridondanti. A questo proposito bisogna ricordare che nei discorsi di fine anno di alcuni Presidenti della Repubblica, e in quelli di Pertini in particolare, si riscontrano spesso interi paragrafi ripetuti quasi identici in discorsi di anni diversi e, per effetto del *TFIDF*, questi "quasi doppi" finiscono molto vicini nella graduatoria ma devono essere esclusi dal condensato finale.

In Tab. 2 (colonna con il codice identificativo del frammento) si può facilmente osservare che in cima alla graduatoria dei frammenti più pertinenti di un Presidente possono essere inclusi frammenti di discorsi appartenenti ad altri Presidenti. Il numero e la posizione nella graduatoria di questi innesti si possono considerare una base per pervenire a una vera e propria misura della forza dello stile personale del Presidente o, da un altro punto di vista, della dipendenza tra discorsi di Presidenti diversi. Nelle prime 30 posizioni del *ranking* di Pertini si vedono solo frammenti estratti dai discorsi dello stesso Pertini, a testimonianza del fatto che i discorsi di Pertini sono unici nel loro genere. Scendendo ulteriormente lungo la scala di priorità offerta dal *TFIDF* di Pertini, a un certo punto entrano frammenti dei discorsi di Napolitano (sul tema dei giovani e della disoccupazione) e di Ciampi (sul tema del terrorismo). Nel condensato di Ciampi entra quasi subito in graduatoria un frammento di Napolitano (riportato nel condensato ed evidenziato con il corsivo) e, nonostante la prima parte della graduatoria sia occupata in prevalenza da frammenti dei discorsi dello stesso Ciampi, si vedono via via alcuni innesti provenienti dai discorsi di Gronchi, Cossiga, Scalfaro.

Napolitano rappresenta un caso più complesso perché dalle nostre analisi sembra non avere ancora un profilo ben definito. Siccome non ha ancora concluso il mandato, il materiale a disposizione riguarda solo i primi tre discorsi e risulta difficile tracciarne le specificità. In questi casi gli innesti risultano utili per ricostruire le parti mancanti, quasi come frammenti di un codice genetico non ancora disponibile nella sua interezza. Per esempio, quando Napolitano parla di coesione sociale, tema centrale dei suoi discorsi e tema caro al suo predecessore, si trova facilmente continuità tra frammenti dei suoi discorsi e quelli di Ciampi:

«Più coesione significa anche più equità, meno disparità dei redditi e delle condizioni di vita, più vicinanza e sostegno per le persone e le famiglie [...]. Più coesione significa inoltre uno sforzo maggiore per integrare nel sistema dei nostri principi e precetti costituzionali, senza discriminarli o tenerli ai margini, gli stranieri di cui l'Italia oggi ha certamente bisogno, e di cui è stato ed è giusto regolare l'ingresso legale nel nostro paese.» (Napolitano)

«L'uomo di Governo deve reagire mirando a realizzare una più forte coesione fra tutti coloro che sanno come si costruisce la pace; rafforzando le istituzioni che abbiamo creato in applicazione coerente dei nostri valori [...]. Questa è la via da seguire, tendendo la mano a tutte le civiltà, a tutti i popoli [...]. Questo l'Italia chiede a tutti i suoi cittadini, come a tutti gli stranieri che vivono in mezzo a noi e condividono i nostri diritti e i nostri doveri.» (Ciampi)

Nell'affrontare temi istituzionali e costituzionali, si rende evidente una vicinanza a Cossiga, oltre al già citato Ciampi (non a caso questo passaggio risulta tra i primi anche nel *ranking* del suo predecessore):

«Mi sono stati già nel passato affidati delicati incarichi nelle istituzioni italiane ed europee, ma sto ora verificando quanto sia più complessa e impegnativa la responsabilità che la nostra Costituzione attribuisce al Capo dello

Stato. Interpretare ed esprimere, con passione civile e con assoluta imparzialità, sentimenti e valori condivisi, esigenze e bisogni che riflettono l'interesse generale del paese, e guardare sempre all'unità nazionale come bene primario da tutelare e consolidare.» (Napolitano)

«L'essere chiamato a rappresentare l'Italia, a essere garante della sua Costituzione, l'ho vissuto non solo come un altissimo mandato, ma soprattutto come un dovere, una missione.» (Ciampi)

«Sono convinto che nessun obiettivo è arduo se noi cittadini, noi istituzioni della Repubblica, troveremo tensione morale, operosità e passione civile, unità nell'essenziale. [...] Da parte mia, nel doveroso rispetto dei limiti posti dalla Costituzione alle mie funzioni, ma nel pieno e responsabile esercizio dei poteri e delle prerogative che mi sono conferite dalla Costituzione, mai dimenticherò che la mia autorità deriva dall'essere stato eletto [...] in nome e per conto del popolo sovrano [...]. Quest'autorità, in democrazia non può essere un mio privilegio personale: è una autorità di servizio, di servizio alla comunità, alla patria italiana che amo, da cittadino eguale ad ogni altro cittadino, e che ho giurato solennemente di servire con lealtà.» (Cossiga)

L'esperienza di applicazione di questo metodo di lavoro ci sembra in conclusione promettente perché attraverso semplici e ben collaudati strumenti dell'ADT è possibile ottenere un condensato efficace dello stile e dei contenuti, cioè quello che abbiamo chiamato provocatoriamente l'autografo del Presidente. A livello di produzione si può procedere con la stessa logica nella stesura assistita di nuovi discorsi (con la possibilità di stabilirne *a priori* lunghezza, contenuti e finalità) oppure, in fase di analisi, si possono facilmente riconoscere ripetizioni e riprese.

Il *ghostwriter* è attrezzato (o anche avvisato?).

Riferimenti

- Baayen H.R. (2001). *Word Frequency Distributions. Exploring Quantitative Aspects of Lexical Structure*. Dordrecht: Kluwer Academic Pub.
- Bolasco S. (1999). *L'analisi multidimensionale dei dati*. Roma: Carocci.
- Bolasco S., Baiocchi F. and Morrone A. (2008). *TaLTaC²: Trattamento automatico Lessicale e Testuale per l'analisi del Contenuto di un Corpus (Ver.: 2.8.0.2)* [<http://www.taltac.it>].
- Cortelazzo M.A. and Tuzzi A. (editors) (2007). *Messaggi dal Colle. I discorsi di fine anno dei presidenti della Repubblica*. Venezia: Marsilio.
- La Mendola S. and Sterchele D. (2007). Costruire Noi. I presidenti della Repubblica italiana come cerimonieri di un rituale mediale. In Cortelazzo, M.A. and Tuzzi, A., editors, *Messaggi dal Colle*. Marsilio, Venezia, pp. 55-85.
- Lafon P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots*, vol. (1): 127-65.
- Lebart L. (2004). Validité des visualisations de données textuelles. In Purnelle, G., Fairon, C. and Dister, A., editors, *Le poids des mots. JADT2004*. UCL Presses Universitaires de Louvain, 2, pp. 708-715.
- Lebart L., Salem A. and Berry L. (1998). *Exploring textual data*. Dordrecht: Kluwer-Academic Pub.
- Morrone A. (1996). Temi generali e temi specifici dei programmi di governo attraverso le sequenze di discorso. In Villone, M. and Zuliani, A., editors, *L'attività dei governi della repubblica italiana (1947-1994)*, Bologna: il Mulino, pp. 351-369.
- Pauli F. and Tuzzi A. (2009). The End of Year Addresses of the Presidents of the Italian Republic (1948-2006): discursal similarities and differences. *Glottometrics*, vol. (18): 40-51.
- Salton G. (1989) *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer*. Boston (MA): Addison-Wesley Longman.
- Salton G. and Buckley C. (1988). Term-weighting approach in automatic text retrieval. *Information Processing & Management*, 24(5): 513-523.
- Tuzzi A. (2003). *L'analisi del contenuto*. Roma: Carocci.

- Tuzzi A. and Tweedie F.J. (2000). The Best of Both Worlds: Combining MOCAR and MCDISP. In Rajman, M. and Chappelier, J.C., editors, *JADT 2000*, EPFL, vol. 1, pp. 271-76.
- Zijno A. (2007). Io presidente, voi cittadini. Autorappresentazione del capo dello Stato e rappresentazione dei cittadini nei discorsi di fine anno. In Cortelazzo, M.A. and Tuzzi, A., editors, *Messaggi dal Colle*. Venezia: Marsilio, pp. 87-107.
- Zotti Minici A. (2007). Dove guardano i presidenti a fine anno. In Cortelazzo, M.A. and Tuzzi, A., editors, *Messaggi dal Colle*. Venezia: Marsilio, pp. 47-53.