

Trattamento Automatico della Lingua per la comunicazione della Pubblica Amministrazione

Dominique Brunato, Andrea Cimino, Felice Dell’Orletta, Simonetta Montemagni, Giulia Venturi

Istituto di Linguistica Computazionale "A. Zampolli" (ILC-CNR), Pisa
ItaliaNLP Lab (www.italianlp.it)
italianlp@ilc.cnr.it

Abstract

Il contributo illustra le attività portate avanti dal Laboratorio ItaliaNLP Lab finalizzate al miglioramento delle strategie comunicative della Pubblica Amministrazione (PA), impiegando strumenti di Trattamento Automatico della Lingua (TAL) sia per la valutazione dell’accessibilità linguistica dei testi rivolti al cittadino sia per favorire l’organizzazione strutturata dei contenuti prodotti dalle PA.

1 Introduzione

La fase di rinnovamento che sta attraversando la Pubblica Amministrazione è legata a due principali fattori: da un lato, una presa di coscienza interna, che sta portando progressivamente a ridisegnare i termini stessi del rapporto tra l’amministrazione e il cittadino, dall’altro, la richiesta di adeguarsi alle dinamiche sempre più pervasive della comunicazione in rete che coinvolge un ampio ed eterogeneo gruppo di cittadini. Nell’era della digitalizzazione e degli Open Data la promozione di metodi di monitoraggio e valutazione della qualità della comunicazione pubblica sta diventando perciò elemento cardine di una società realmente inclusiva. È quanto ricordato anche dalle Web Content Accessibility Guidelines proposte dalla *Web Accessibility Initiative* e dalla *Europe 2020: Europe’s growth strategy*¹, con la quale l’Unione europea ha posto tra i suoi obiettivi più ambiziosi quello di provvedere ad una “smart, sustainable and inclusive growth”.

In quanto segue, si presentano alcune attività avviate dal Laboratorio ItaliaNLP e finalizzate a sviluppare metodi e strumenti di Natural Language Processing (NLP) per la valutazione della comunicazione pubblica, sia rispetto alla forma linguistica usata sia rispetto ai contenuti veicolati.

2 Valutazione della comunicazione sanitaria

Il diritto alla salute, tutelato dal Servizio Sanitario Nazionale, si fonda su principi di universalità e di equità di accesso all’informazione per “la popolazione senza distinzione di condizioni individuali o sociali e secondo modalità che assicurino l’eguaglianza dei cittadini nei confronti del servizio”².

Allo scopo di fornire un supporto alla Pubblica Amministrazione in questa direzione, l’ItaliaNLP Lab ha attive una serie di collaborazioni con soggetti pubblici che si propongono di definire strategie di monitoraggio della qualità della comunicazione sanitaria. Sebbene tali collaborazioni riguardino modalità di comunicazione diverse, esse sono accomunate dalla specializzazione di metodologie di valutazione automatica della leggibilità di un testo. A questo scopo è usato READ-IT [Dell’Orletta *et al.*, 2011], il primo e al momento unico strumento di valutazione della leggibilità esistente per l’italiano basato su strumenti di NLP e ispirato alla letteratura linguistica più recente sulla complessità linguistica. READ-IT si basa su un’analisi sofisticata delle strutture linguistiche sottostanti al testo condotta con strumenti di NLP allo stato dell’arte per la lingua italiana. Oltre a fornire una valutazione affidabile della leggibilità, READ-IT è anche in grado di identificare i luoghi di complessità del testo (lessicale o sintattica), fornendo così un valido supporto per un eventuale processo di semplificazione testuale.

La valutazione della qualità della comunicazione medico-paziente è obiettivo della collaborazione con il Centro gestione rischio clinico e sicurezza del paziente della Regione Toscana (GRC). Come discusso nella letteratura sulla sicurezza, il basso livello di leggibilità delle informative di consenso è origine di contenziosi da parte dei pazienti in caso di trattamenti che determinano un danno attribuibile alle cure. Allo scopo di monitorare la situazione nella regione Toscana, l’ItaliaNLP ha condotto uno studio del livello di leggibilità delle informative di consenso impiegate nelle Aziende sanitarie toscane [Venturi *et al.*, 2015]. Lo studio si è basato su un corpus di documenti attualmente in uso nelle 16 Aziende ospedaliere del Servizio sanitario della Toscana che coprono diverse specialità mediche, con il duplice obiettivo di: i) fornire evidenza dei possibili ostacoli linguistici alla piena comprensione del consenso informato e ii) definire future linee guida per la redazione di informative chiare e semplici.

Una seconda collaborazione è in atto con il Laboratorio MeS - Istituto di Management della Scuola Superiore Sant’Anna di Pisa, all’interno della quale l’ItaliaNLP Lab sta adottando metodi e strumenti di valutazione automatica della leggibilità per monitorare le modalità di comunicazione esterna delle aziende sanitarie. Lo studio prevede l’esame dei siti web istituzionali delle aziende presenti sull’intero territorio nazionale allo scopo di individuare possibili elementi linguistici

¹http://ec.europa.eu/europe2020/pdf/europe_2020_explained.pdf

²Legge n. 833 del 23 dicembre 1978.

stici che possono rappresentare un ostacolo all'accesso all'informazione per un pubblico di cittadini ampio ed eterogeneo rispetto alle competenze e abilità linguistiche.

3 Valutazione della comunicazione istituzionale

Incentivare nelle pubbliche amministrazioni uno stile comunicativo improntato sui principi di chiarezza e semplicità linguistica diventa ancora più improrogabile se consideriamo il caso della comunicazione istituzionale. Per sua natura, la comunicazione istituzionale non si rivolge a specialisti bensì al cittadino comune, che raramente è un esperto di materie amministrative. I metodi sviluppati dall'ItaliaNLP Lab per migliorare la comunicazione istituzioni-cittadino sono sperimentati nella collaborazione con l'Accademia della Crusca, l'Istituto di Teorie e Tecniche dell'Informazione Giuridica (ITTIG) del CNR di Firenze e l'Associazione per la Qualità degli Atti Amministrativi (AQuAA). Il contributo del laboratorio è rivolto in particolare alla specializzazione di READ-IT come supporto operativo al drafting amministrativo e normativo [Brunato e Venturi, 2016], anche implementando alcune delle regole di buona scrittura incluse nella *Guida alla Redazione degli Atti Amministrativi. Regole e Suggerimenti*, redatta dall'Accademia della Crusca e dall'ITTIG-CNR. L'attività di alta formazione professionale svolta dal laboratorio rappresenta una seconda linea di attività. Dalla sua prima edizione, siamo parte del collegio docente del master "La lingua del diritto", organizzato dall'Università di Pavia e dal Senato della Repubblica. Il nostro contributo consiste nel formare esperti di scrittura giuridica consapevoli delle potenzialità d'uso dei metodi di valutazione automatica della leggibilità per la scrittura e revisione di atti chiari e precisi.

4 Valutazione della comunicazione del sistema scolastico

Anche la scuola intesa come sistema sta vivendo un momento di profondo rinnovamento, che coinvolge anche i processi di comunicazione. In particolare, mettere a punto adeguate strategie di comunicazione con i soggetti esterni sta diventando sempre più centrale per la gestione stessa dell'organizzazione scolastica. In questo contesto, l'ItaliaNLP Lab ha avviato una collaborazione con l'Istituto Nazionale di Documentazione, Innovazione e Ricerca Educativa (Indire) per esaminare come la scuola comunica all'esterno le proprie politiche riguardo a tematiche strategiche quali lo sviluppo di competenze di base o specifiche, l'innovazione dei percorsi didattici, la dispersione scolastica. A questo scopo, metodi e strumenti di Information Extraction, Knowledge Management e analisi stilistica automatica sono impiegati nell'analisi della sempre più numerosa documentazione prodotta dalla scuola. L'obiettivo è fornire evidenze per la valutazione del modo in cui la scuola comunica la propria offerta formativa ed educativa.

Questi metodi sono stati applicati all'analisi di tre diverse tipologie di documenti: *i*) i Piani di Miglioramento (PdM) nei quali la scuola pianifica gli obiettivi e le azioni del percorso di miglioramento per il raggiungimento dei traguardi indicati nel Rapporto di Auto-Valutazione (RAV), *ii*) le risposte alla rilevazione telematica condotta da Indire nella quale

si chiedeva alle scuole di descrivere le esperienze dei progetti di Alternanza Scuola-Lavoro, *iii*) il portfolio professionale contenuto nella piattaforma Neoassunti che raccoglie le descrizioni e riflessioni dei docenti neoassunti sulle proprie esperienze didattiche. Le prime due tipologie di documenti sono state analizzate con la piattaforma di estrazione e organizzazione dell'informazione da corpora di dominio *Text-to-Knowledge, T2K* [Dell'Orletta *et al.*, 2014]. All'interno dei documenti linguisticamente annotati in modo automatico, sono state ricercate le entità rilevanti in base alla loro realizzazione linguistica, ovvero identificando sequenze di categorie morfo-sintattiche rappresentative di diversi tipi di modificazione nominale quali ad esempio *classi parallele, corso base di diritto del lavoro*, ecc. o delle azioni nelle quali entità di tipo nominale sono inserite, come esempio *utilizzare criteri di valutazione omogenei, imparare ad imparare*, ecc. Ciò ha permesso di tracciare un profilo del sistema scolastico variegato rispetto alla posizione geografica, all'ordine e grado scolastico.

Metodi di analisi stilistica automatica del testo sono stati inoltre impiegati per analizzare le caratteristiche linguistiche dei report scritti dai docenti neoassunti. Dai testi, linguisticamente annotati in modo automatico, è stata estratta un'ampia gamma di informazioni relative al lessico, alle strutture morfosintattiche e sintattiche impiegate. Come descritto più in dettaglio in [Chiriatti *et al.*, 2018], un'attenzione particolare è stata dedicata all'individuazione delle caratteristiche tipiche della comunicazione riflessiva, che sono state usate in esperimenti di classificazione per predire quando un testo che descrive un'esperienza didattica è riflessivo.

Riferimenti bibliografici

- [Brunato e Venturi, 2016] Dominique Brunato e Giulia Venturi. Le tecnologie del linguaggio per la leggibilità della comunicazione istituzionale. In *Saulle Panizza (ed.), Profili attuali di qualità degli atti normativi e amministrativi*, Pisa University Press, 2016.
- [Chiriatti *et al.*, 2018] Giulia Chiriatti, Valentina Della Gala, Felice Dell'Orletta, Simonetta Montemagni, Maria Chiara Pettenati, Maria Teresa Sagri, e Giulia Venturi. A NLP-based analysis of reflective writings by italian teachers. In *CLiC-it*, 2018.
- [Dell'Orletta *et al.*, 2011] Felice Dell'Orletta, Simonetta Montemagni, e Giulia Venturi. READ-IT: assessing readability of italian texts with a view to text simplification. In *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies*, 2011.
- [Dell'Orletta *et al.*, 2014] Felice Dell'Orletta, Giulia Venturi, Andrea Cimino, e Simonetta Montemagni. T2K: a system for automatically extracting and organizing knowledge from texts. In *LREC*, 2014.
- [Venturi *et al.*, 2015] Giulia Venturi, Tommaso Bellandi, Felice Dell'Orletta, e Simonetta Montemagni. NLP-based readability assessment of health-related texts: a case study on italian informed consent forms. In *Louhi - EMNLP*, 2015.