

LA SEMIOTICA PUÒ MIGLIORARE L'APPRENDIMENTO SUPERVISIONATO DELLE RETI NEURALI?

IL CASO DI STUDIO DEL TWEET DI PAPA FRANCESCO^[1]

FRANCESCO GALOFARO, MAGDALENA MARIA KUBAS,
IULM e Università degli Studi di Torino^[2]

ABSTRACT: *The paper focuses on a case–study: a corpus of 1234 Italian tweets in reply to Pope Francis’ ecological tweets related to the encyclical letter Laudato si’ has been collected and labelled by the research team of the Semiotic and big data lab at the University of Turin using semiotic categories to substitute the vague notion of “subjectivity” in use in sentiment analysis. A simple neural network has been trained on the corpus to classify messages into “history” and “discourse” with a final accuracy score of 97%. The paper explores the practical and social implications of the feasibility study as well as its limitations and suggests further transdisciplinary research using semiotics as a standard vocabulary to increase cooperation between social and computer sciences. The analysis is accompanied by a historical premise and a brief analysis of the concordance between Saint Francis of Assisi’s Canticle of the Creatures and Pope Francis’ encyclical letter Laudato si’.*

KEYWORDS: *Actor network theory, natural language processing, sentiment analysis, subjectivity, machine learning.*

(1) Questo progetto ha ricevuto finanziamenti dall’European Research Council (ERC) nell’ambito del programma di ricerca e innovazione Horizon 2020 dell’Unione europea (convenzione di sovvenzione n. 757314).

(2) La dr.ssa Magdalena Maria Kubas ha steso i paragrafi 2.1 e 2.2. Il resto dell’articolo è stato scritto dal dr. Francesco Galofaro. Si ringraziano gli studenti del Laboratorio di Semiotica e Big Data (UniTo) Linda Pigureddu, André Molina Carmona, Stefania Molinaro e Gianluca Bufo per il loro contributo alla realizzazione della ricerca.

1. Scopo della ricerca

Sebbene la *sentiment analysis* sia sempre più utilizzata per confrontare la comunicazione dei leader politici (Grubbs 2020) e d'impresa, le sottostanti nozioni tecniche di *polarità* (*polarity*) e *soggettività* (*subjectivity*) sono definite in modo vago e spesso incoerente. Ad esempio:

Sentiment polarity for an element defines the orientation of the expressed sentiment, i.e., it determines if the text expresses the positive, negative or neutral sentiment of the user about the entity in consideration (Kumar and Gupta 2021).

Questo tipo di definizione non specifica in che modo il testo esprima il suddetto sentimento e il suo “orientamento”. Se utilizzate per fornire istruzioni agli umani che etichettano i *corpora*, questo genere di definizioni vaghe contribuisce ad aumentare l'opacità degli algoritmi, con effetti sociali non prevedibili (Burrell 2015). Anche la classificazione della “soggettività” è problematica:

Subjectivity classification [...] distinguishes sentences (called objective sentences) that express factual information from sentences (called subjective sentences) that express subjective views and opinions. However, we should note that subjectivity is not equivalent to sentiment as many objective sentences can imply opinions, e.g., «We bought the car last month and the windshield wiper has fallen off» (Bing Liu 2021:11).

La distinzione tra fatti e opinioni è dibattuta in diverse tradizioni filosofiche come la filosofia della mente (Putnam 2002) e il post-strutturalismo. Come ha scritto Bruno (Latour 2017, p. 34), nonostante la vasta letteratura sull'indispensabile chiasmo tra fatti e valori, è evidente che la definizione dei primi incide necessariamente e in maniera decisiva sui secondi.

L'uso estensivo di reti neurali per classificare grandi corpora di messaggi porta alla necessità di istruzioni più precise agli esseri umani (spesso studenti universitari) responsabili dell'etichettatura dei corpora utilizzati per addestrarle. In particolare, i tratti linguistici impiegati per

esprimere “oggettività” e “soggettività” come effetti di senso sono stati studiati da Benveniste (1966–1974). Nonostante il trascorrere degli anni, la sua analisi è ancora discussa in diversi ambiti disciplinari quali l’antropologia (si veda Viveiros de Castro 2017, ma anche le critiche espresse a quest’ultimo in Lucatti 2022, pp. 20–25).

In questa prospettiva, il presente contributo presenta i risultati di un esperimento condotto durante il *Laboratorio di semiotica e big data* dell’Università di Torino, a.a. 2021–2022. Gli studenti sono stati istruiti ad analizzare i tweet utilizzando le categorie di Benveniste per etichettare un corpus di 1234 risposte ai tweet di papa Francesco sul tema dell’ecologia con riferimento alla sua lettera enciclica *Laudato si’*. Si è scelto questo topic poiché il tema è molto divisivo e ha stimolato un’ampia gamma di risposte, dall’entusiasmo agli insulti e ai discorsi di odio.

2. L’approccio semiotico

Partendo dal caso studio, le caratteristiche innovative della comunicazione di papa Francesco hanno attirato l’attenzione degli studiosi di semiotica fin dall’inizio del suo pontificato (Lorusso e Peverini 2017). Anche un sociologo di ispirazione semiotica come Bruno Latour (2017, pp. 287–288) ha discusso la lettera enciclica *Laudato si’* e i suoi aspetti innovativi.

Venendo alla *sentiment analysis*, la semiotica è considerata pertinente dai professionisti del settore (Ceriani 2016) e utile a una migliore comprensione dei messaggi di Twitter (Santangelo 2016). Inoltre, l’approccio sociosemiotico è stato impiegato da un gruppo di ricerca transdisciplinare per studiare i pregiudizi che coinvolgono le tecniche di apprendimento automatico e i rischi per la democrazia (Beretta, Santangelo *et al.* 2019). In particolare,

in the era of Big Data, many of the data used today have not been generated using probabilistic sampling, but are rather selected through non probabilistic methods (very often acquired from third parties, or with opportunistic methods, thanks to the pervasiveness of digital technologies), which do not provide to each unit of the population

the same opportunity to be part of the sample; this means that some groups or individuals are more likely to be chosen, others less (Beretta, Santangelo et al. 2019, p. 3).

Anche nel nostro caso il corpus non può essere in nessun caso considerato come rappresentativo dell'opinione pubblica o simili. Il presente lavoro deve essere considerato solo come uno studio di fattibilità: lo sviluppo di un algoritmo semiotico pienamente funzionante in grado di classificare il grado di soggettività dei messaggi implica una ricerca più approfondita e una cooperazione transdisciplinare tra semiotica, statistica e teoria dell'informazione.

2.1. *Scienze umane e corpora: una premessa storica*

Nella parte iniziale del presente contributo è utile mettere a fuoco due questioni: la prima riguarda i *corpora*, la seconda le concordanze o elenchi di frequenza che si possono estrarre dai *corpus*. Si tratta di strumenti per l'analisi linguistica-statistica, e più specificamente lessicografica. Il secondo implica l'esistenza (cioè un'adeguata preparazione) del primo. La manualistica e le pubblicazioni scientifiche sull'argomento sono cospicue, soprattutto per quel che riguarda i *corpora*, le concordanze e gli usi di questi strumenti dell'epoca della digitalizzazione (Giuliano 2013). Per introdurre l'argomento confrontiamo brevemente due definizioni del *corpus*.

(1) Secondo l'Enciclopedia Treccani online si tratta di un termine

usato nella nomenclatura bibliografica per indicare la riunione di più opere, raccolte e pubblicate con lo scopo di fornire serie ordinate e complete degli scritti di uno o più autori o intorno a determinate materie.

(2) D'altra parte, secondo Baker (2006, p. 26):

the term *corpus* merely refers to a body of electronically encoded text, it is not the case that a corpus consists of any collection of texts, picked

at random. Instead, researchers have produced a range of recognizably different types of corpora, depending on the sorts of research goals they have in mind. [...] Corpora therefore tend towards having a more balanced, carefully thought-out collection of texts that are representative of a language variety or genre.

Le due definizioni testimoniano l'esistenza di stadi diversi nell'uso dei corpora, nella loro preparazione e negli strumenti che vengono adoperati in tutte le fasi di elaborazione e della fruizione: la prima definizione precede di 80 anni la redazione della seconda. Con ciò rimane saldo il principio dell'insieme testuale stabilito sulla base di uno o più criteri. Se i primi corpus furono raccolti nell'antichità — pensiamo ad esempio al *Corpus Hippocraticum* raccolto con il criterio autoriale-disciplinare⁽³⁾ — oggi sono numerosissimi i corpora digitali e accessibili online, tra cui l'OVI (la monumentale *Opera del Vocabolario Italiano* con le sue articolazioni; si tratta di un progetto inaugurato nel 1965) elaborato dall'Istituto del CNR presso l'Accademia della Crusca assieme a più software per interrogare il *corpus*⁽⁴⁾.

Per la ricerca sul francescanesimo dei primi secoli abbiamo il tradizionale *corpus* cartaceo delle *Fonti francescane* contenente le vite e gli scritti dei fondatori, cui si aggiungono corpora a tema liturgico, le fonti agiografiche dell'ordine, le fonti normative e altro ancora. Nella versione italiana le *Fonti* non hanno un'elaborazione digitale che permetta una consultazione, mentre le fonti antiche latine (francescane e clariane), assieme alla traduzione inglese, sono consultabili in formato digitale⁽⁵⁾.

La seconda questione riguarda le concordanze: si tratta di un elenco, ordinato alfabeticamente, delle parole contenute in un testo. Attraverso lo spoglio lessicale (oggi attuato con i mezzi informatici, ma fino a pochi decenni fa eseguito manualmente) si ottiene un elenco in cui i singoli termini sono «accompagnati da alcuni riferimenti che ne rendono

(3) Non esistendo un principio dell'autorialità, il nostro corpus conteneva opere (presumibilmente) di Ippocrate insieme a opere e commenti posteriori, di autori che oggi chiameremmo "della scuola": la formazione del corpus impiegò alcuni secoli dopo la morte di Ippocrate.

(4) Storia e strumenti di ricerca elettronica sono accessibili al link <http://tlio.ovi.cnr.it/TLIO/>.

(5) Cfr. il sito "Franciscan tradition" al link <https://franciscantradition.org/early-sources>.

possibile il ritrovamento nell'originale» (Gigliozzi 2003, p. 96). Come abbiamo detto, le concordanze sono uno strumento statistico. Ad esso possono collegarsi ad esempio indici delle parole (o di un tipo specifico di termini), elenchi di frequenze e altro. Tra le pratiche più antiche delle concordanze vi è la composizione delle concordanze bibliche, estratte manualmente: i frati Minori e i Predicatori si contendono alcuni nomi che potevano essere tra i primissimi compilatori, come Ugo da San Caro e Arlotto da Prato, figure attive tra la fine del XII e l'inizio del XIII secolo⁽⁶⁾; inoltre, si citano S. Antonio da Padova (ma non si tratterebbe di un'opera edita), e Ugo di S. Vittore. Il recente centenario dantesco⁽⁷⁾ permette di avvicinarci alla modernità: la *Divina Commedia* fu oggetto di numerose elaborazioni statistiche: nel secondo Ottocento vi fu un lavoro intenso sulle concordanze, non soltanto in Italia. Sono note le *Concordanze della Divina commedia di Dante Alighieri* di Giovanni Andrea Scartazzini (pubblicato per Brockhaus a Lipsia nel 1901) e *Concordance of the Divina Commedia* di Edward Allen Fay pubblicata nel 1888. Nel 1965 — un altro centenario dantesco — l'IBM Italia, per inaugurare il Centro nazionale universitario di calcolo elettronico a Pisa estrasse le concordanze della *Divina commedia* trascrivendo prima l'intero testo, suddiviso in versi, su 14.233 schede perforate. L'analisi del calcolatore durò quasi 19 ore. La concordanza è pubblicata con il titolo *La Divina commedia. Testo, concordanze, lessici, rimario, indici* (con prefazione di C. Tagliavini e nota tecnica di G. Caldara). Nell'analisi letteraria le concordanze sono utili per risalire alle interconnessioni presenti in un solo testo, o tra vari testi, in maniera intertestuale, ma oggi l'uso dei corpora e delle concordanze eseguite con i mezzi informatici è amplissimo e va dalla linguistica al marketing, dalla traduzione automatizzata alla gestione della documentazione giuridica, ecc. (Giuliano 2013, p. 5). Nel suo articolo intitolato *Per uno studio sulla "procedura delle concordanze"* Luca Andrea di Martino parla di

iterazioni di certe unità significanti, mediante le quali analogie, parallelismi, catene anaforiche percorrono le probabili connessioni tra

(6) La contesa è descritta in una delle prime storie della letteratura italiana come la *Storia della letteratura italiana del cav. Abate Girolamo Tiraboschi*, Tomo IV, 1795, p. 141.

(7) Nel 2021 è stato celebrato il settimo centenario della morte di Dante.

gli elementi. Le concordanze rendono evidenti, isolandole, le qualità delle relazioni prescelte, come se queste venissero contrassegnate “istologicamente”.

E continua:

la procedura delle concordanze funziona da mezzo isomorfo in grado di scernere le varie affinità riscontrabili tra le parti comunque collocate all'interno d'un insieme compatto (Di Martino 2012, p. 191).

2. 2. *Tra il Cantico delle creature di San Francesco e l'enciclica Laudato sì di papa Francesco*

All'interno della presente ricerca vale la pena di soffermarsi brevemente sulla tradizione ecologista francescana, tra le sue origini e il pensiero odierno, tra il *Cantico delle Creature* o *Cantico di Frate Sole*, la lauda di Francesco d'Assisi, scritta negli anni Venti del Duecento, e la lettera enciclica papale del 2015. Com'è noto la sensibilità francescana, almeno quella del primo periodo, si plasmò su canti e preghiere lasciati da Francesco. Il *Cantico*, in particolare, risale agli ultimi anni di vita del santo. È parere condiviso tra gli studiosi che il testo si divida in due parti tematiche: la prima, più lunga, contiene un elenco di elementi, astri, esseri viventi; la seconda, più breve, inizia con “sora nostra morte corporale”. Si crede che questa suddivisione rifletta due fasi nella composizione del testo, tra le quali intercorrono alcuni anni (Contini 1995)⁽⁸⁾.

Nell'ambito del pensiero cristiano, il testo di San Francesco rappresenta un'espressione precoce di sensibilità ambientale. Papa Francesco, nella lettera enciclica *Laudato sì*, si ricollega a questa sensibilità riprendendola per parlare della realtà contemporanea. Francesco crea una connessione testuale diretta citando un frammento del *Cantico delle creature* (vv. 5–19) e cerca strategie sia per avvicinare quel pensiero alla realtà contemporanea che per “tradurlo” nelle parole che suonano familiari all'uomo odierno. L'enciclica contiene una critica forte e una serie

(8) Per le notizie e le trascrizioni del testo v. Contini (1995, pp. 33–34); Pozzi (1971); Branca (1950).

di proposte per trasformare la nostra cultura, per vivere con maggiore rispetto della terra e degli esseri viventi. Una breve sintesi: nel primo capitolo, parlando della responsabilità comune di tutti gli abitanti del pianeta (“la casa comune”), il Papa auspica l’unità mondiale per affrontare la crisi ecologica in corso. Nel secondo capitolo il papa si confronta direttamente con la lauda di San Francesco. Al di là delle motivazioni teologiche, la preoccupazione principale del terzo capitolo è la tecnologia. Alcune parti del quarto capitolo propongono di vederci come un’unica collettività legata alla “casa comune”, in relazione allo sviluppo sostenibile. Il quinto capitolo contiene una critica del consumismo e i postulati per combatterlo, con una serie di proposte praticabili nella vita quotidiana di tutti. Nell’ultimo capitolo troviamo le proposte per un’educazione alla spiritualità ecologica.

A nostro parere, in *Laudato si* Papa Francesco cerca di intrecciare il linguaggio della lauda medievale con il discorso ai contemporanei. In che modo? Il papa tenta un aggiornamento del genere testuale e del linguaggio. Francesco d’Assisi aveva usato il genere delle benedizioni e lodi bibliche, un proposito che Bergoglio riprende nella preghiera che chiude la sua lettera enciclica. Nella parte principale invece egli passa dalla poesia alla prosa. Per aggiornare la lingua e il pensiero di San Francesco *Laudato si* si muove in due direzioni: da un lato la lingua del papa va in direzione della precisazione scientifica, dall’altro si affida a uno stile espressivo affettivo–familiare.

Per capire bene i modi è utile partire da dati statistici: nell’enciclica papale la parola “terra” presenta 77 occorrenze. Troviamo sia l’associazione “matre/madre terra” che la “sorella terra”. Se l’idea francescana di fratellanza (e sorellanza) non distingue tra gli esseri viventi con i loro processi fisiologici (come la morte), il mondo inanimato e i quattro elementi, nell’enciclica di papa Francesco le “creature” della lauda vengono anche designate, o “tradotte”, come “esseri viventi”, “organismi”, parti dell’ecosistema (vale la pena di notare che in *Laudato si* la voce lessicale “creature” è un lemma a frequenza alta, dato che presenta 77 occorrenze). Quello che per Francesco d’Assisi è l’aere, la fonte del “sustentamento”, diventa lo spazio di accumulo degli “inquinanti atmosferici” e la causa della morte prematura degli esseri viventi. Nell’enciclica il lemma “aria” è presente 5 volte, mentre l’inquinamento

atmosferico e il cambiamento climatico diventano argomenti discussi in porzioni più ampie del testo papale. Nel *Cantico delle creature* è menzionato il fuoco: oggi esso è visto nella funzione del “combustibile” (5 occorrenze), ed è ricordato perché riscalda le case dei ricchi e dei poveri, con un conseguente inquinamento dell’aria. La “sor’acqua”, con le sue 39 occorrenze nell’enciclica, è declinata come “l’acqua potabile”, “l’acqua pubblica” e si iscrive in una questione più generale di diritti, della gestione delle risorse e degli sprechi. L’espressione “specie” contra 28 occorrenze per riunire la ricchezza (detta anche “biodiversità”, 10 occorrenze) dei regni animale e vegetale. La strategia generale del papa è quindi quella di utilizzare la lauda per fondare e costruire le categorie concettuali del suo discorso.

Nel *Cantico delle creature* appare l’uso del pronome “noi”: si tratta di 1 occorrenza, cui si aggiunge il possessivo “nostra”, con 2 occorrenze. Il “noi” permette la creazione di un attore collettivo e di una dimensione comunitaria: è un aspetto che Papa Francesco amplia infinitamente nella sua proposta di rilancio del francescanesimo ecologista per unire la questione sociale a quella dell’ambiente. In *Laudato si* il pronome personale e il possessivo presentano un’alta frequenza: il primo conta 64 occorrenze; la declinazione del possessivo ne conta più di cento (tra cui vi è “nostra”, con 54 occorrenze, comprese “nostra madre (sic) terra”, “nostra oppressa e devastata terra”, “nostra casa comune”, ma anche “nostra dignità”, “nostra fede”, “nostra comprensione della pace”, “nostra identità”, e “nostra sensazione di radicamento”). sono 31 le occorrenze di “nostro”, tra cui il “nostro stesso DNA”, “il nostro senso di appartenenza”, “il nostro sentirci a casa”, “nostro Padre comune”; 12 le occorrenze del plurale “nostri”, tra cui i “nostri fratelli e sorelle”, i “nostri simili su cui si fonda ogni società civile”, “i nostri stili di vita, di produzione e di consumo”; 11 volte compare il plurale femminile “nostre”, più volte nelle “nostre città” e anche nelle “nostre lotte”). Se la famosissima sposa di Francesco d’Assisi era la povertà, le disuguaglianze sono tra le maggiori preoccupazioni del papa attuale: unendo due grandi temi il papa dice che la sostenibilità ambientale non si può ottenere senza intervenire sui problemi sociali. Lo stesso termine “ecologia” (presente 30 volte) si articola in una proposta “ambientale, economica e sociale”, e anche “culturale” “inseparabile dalla nozione del bene comune”.

3. Design della ricerca

Uno degli obiettivi dell'applicazione della semiotica all'addestramento supervisionato delle reti neurali è migliorare la precisione nell'etichettatura del corpus, possibilmente riducendone l'estensione, risparmiando tempo e fondi di ricerca. Un problema relativo alla sentiment analysis è la mancanza di algoritmi per analizzare messaggi in lingua italiana. La maggior parte degli algoritmi viene generalmente addestrata su corpora in lingua inglese. Si sarebbe potuto ricorrere alla traduzione automatica, ma ciò avrebbe introdotto un'impresione non quantificabile nel risultato. Un'indagine preliminare sugli algoritmi esistenti sulla sentiment analysis ha portato all'identificazione di una libreria python, *feel_it* (Bianchi, Nozza et al. 2021), che, purtroppo, si concentra solo sulla polarità dei messaggi e non sulla soggettività. Tuttavia, per ottenere una stima indicativa sul grado di concordanza tra la classificazione effettuata dalla macchina e il giudizio umano, è stata effettuata una sentiment analysis sulla polarità di 200 messaggi sul tema "La giornata contro la violenza verso le donne" (25 novembre 2021). Gli studenti erano d'accordo con i giudizi della macchina nel 65% dei casi. È emerso un problema ulteriore sulla capacità individuale di interpretare una forma di comunicazione così breve e vivida. Ad esempio, uno degli studenti dichiara:

Questo lavoro, che ho trovato interessante e stimolante, è stato anche molto impegnativo per me, in quanto persona autistica. Spesso, a causa della mia naturale predisposizione a non cogliere spunti sociali, ho faticato a dedurre l'atteggiamento o l'emozione degli autori dei tweet dovendo basare la mia decisione esclusivamente sul contenuto testuale e non potendo approfondire il contesto per mantenere i risultati della ricerca il più puliti possibile. In questi casi ho portato il mio contributo scegliendo di non rispondere alla domanda «che cosa intendeva l'autore?» consapevole che i miei colleghi avrebbero dato sicuramente una risposta molto più precisa a questa domanda, ma chiedendomi «quale potrebbe essere una buona classificazione o etichetta per questo specifico tweet?» sulla base di schemi ricorrenti percepiti nell'uso di determinate parole, hashtag, punteggiatura, ecc.

Le discussioni valutative sono proseguite con il mio essere spesso più d'accordo con l'algoritmo che con i miei colleghi, portando a dibattiti costruttivi sulle ragioni che potrebbero aver portato l'algoritmo a compiere determinate scelte. Ciò è accaduto soprattutto quando i tweet sono stati etichettati in modo inaspettato (quando non palesemente impreciso) secondo il giudizio dei miei compagni di squadra ma perfettamente sensato per me, evidenziando le differenze tra la comprensione del sentiment attorno a un argomento mentre si è in una situazione di quotidianità, si ha una buona percezione dei segnali sociali, e la difficoltà di progettare una rete neurale che valuti le conversazioni in assenza di un contesto chiaro e univoco, senza alcun riferimento esterno che possa suggerire atteggiamenti impliciti, come immagini, suoni, linguaggio del corpo, ecc. (traduzione dall'inglese nostra).

La discussione ha portato gli studenti a formulare l'ipotesi che segue: se le nozioni, vaghe, empiriche e psicologiche di polarità e soggettività vengono sostituite con definizioni semiotiche, ancorate alla presenza, nel testo, di determinate caratteristiche linguistiche, il grado di incomprendimento nell'etichettatura umana dei corpora può essere ridotto, migliorando la capacità di classificazione della macchina.

Per verificare l'ipotesi, l'opposizione oggettività/soggettività è stata sostituita dalla categoria storia/discorso di Benveniste (1966, pp. 195–248), le cui caratteristiche linguistiche sono riassunte nella tab. 1.

<i>Caratteristiche</i>	Storia	Discorso
<i>Pronomi</i>	Lui, Lei, Lo,	Io, Tu, Noi
<i>Tempo verbale</i>	Preterito	Presente
<i>Avverbi di spazio</i>	Là	Qui
<i>Effetto di senso</i>	Oggettivo	Soggettivo
<i>Etichetta</i>	0	1

Altre caratteristiche linguistiche interessanti che sono state impiegate per etichettare i tweet sono la presenza della citazione e anche punti interrogativi. Alcuni esempi sono riportati in tab 2. I messaggi sono stati tradotti dall'italiano.

<i>Messaggio</i>	<i>Etichetta</i>
Ho appena comprato #LaudatoSi l'enciclica del @Pontifex_it Francesco su uomo e ambiente... ve la consiglio.	Discorso
Ah è questo nell'Enciclica di @Pontifex_it?! Torneremo allora tutti a scavare la terra a mani nude? #LaudatoSi.	Discorso
Caro @Pontifex_it per questo ti seguirò sempre con speranza combattiva e senza arrendermi mai! #cop21 #laudatosi http://roma.corriere.it/giubileo-2015/ .	Discorso
@matteorenzi oggi sei partito con la tua #enciclica vuoi per caso copiare il santo padre @Pontifex_it #LaudatoSi.	Discorso
@Pontifex_it: La proprietà della casa ha molta importanza per la dignità delle persone e per lo sviluppo delle famiglie. #LaudatoSi.	Storia
Marina Tomarro: "Partire dalla #LaudatoSi per una nuova visione #Politica" @vaticannews_it.	Storia
#LaudatoSi, un monito quello del @Pontifex_it che va colta, a difesa del creato @fai_cisl.	Storia
"Non siamo fatti per vivere nel cemento". Non lo dice il CNR ma @Pontifex_it nel punto 44 dell'enciclica #LaudatoSi.	Storia

La scelta del metodo e della libreria di Python è stata determinata dalla letteratura tecnica sull'argomento (Mirjalili e Rashka 2019). Le caratteristiche delle tecniche di addestramento della rete neurale sono riassunte in tab. 3.

<i>Metodo</i>	Discesa del gradiente stocastico.
<i>Algoritmo di addestramento:</i>	SGDClassifier(loss='log', random_state=1, n_int=5) dalla libreria Python Scikit Learn.
<i>Training set</i>	234 tweet, metà etichettati come "storia", metà etichettati come "discorso" dagli studenti.
<i>Test set</i>	234 tweet, metà etichettati come "storia", metà etichettati come "discorso" dagli studenti.
<i>Punteggio di accuratezza</i>	97%

4. Risultati

Il punteggio di accuratezza del 97% è stato ottenuto utilizzando un corpus molto piccolo di messaggi italiani etichettati da un ristretto numero di studenti utilizzando categorie semiotiche collegate alle caratteristiche linguistiche e non ai vaghi e discutibili concetti di oggettività/soggettività utilizzati nella letteratura sull'analisi dei sentimenti. Tuttavia, la sicurezza della macchina durante l'etichettatura dei messaggi sembra bassa. Tab 4 riporta due esempi:

<i>Messaggio</i>	<i>Predizione</i>	<i>Probabilità</i>
Tanti professionisti, opinionisti, mezzi di comunicazione e centri di potere [...] vivono e riflettono a partire dalla comodità di uno sviluppo e di una qualità della vita che non sono alla portata della maggior parte della popolazione mondiale #laudatosi @Pontifex_it.	Storia	56,21%

#laudatosi 137 Dal momento che tutto è intimamente relazionato [...], propongo di soffermarci adesso a riflettere sui diversi elementi di una ecologia integrale, che comprenda chiaramente le dimensioni umane e sociali. #PapaFrancesco #ClimateAction @forumalcentro @OndaCivica_EU.	Discorso	52.08%
---	----------	--------

5. Limiti e implicazioni

Il modello classificatore utilizzato è tra i più semplici. L'opposizione 0/1 è ovviamente una semplificazione eccessiva, poiché in letteratura troviamo esempi di generi misti, come l'autobiografia (Starobinski 1961). Tuttavia, la ricerca ha il solo scopo di dimostrare l'utilità della semiotica. L'opposizione binaria si adatta bene allo scopo di addestrare un semplice classificatore, poiché le reti neurali diventano progressivamente imprecise quando la classificazione prevede l'uso di più etichette, implicando l'uso di sofisticate tecniche di deep learning (Ziqi e Lei 2018) e ricerca transdisciplinare.

Le categorie semiotiche garantiscono a chi deve etichettare il corpus definizioni univoche basate su forme di espressione e contenuti testuali e non su buon senso, stereotipi, folk-semiotica, nozioni psicologiche o epistemologicamente discutibili, migliorando la ricerca sulla sentiment analysis e sull'elaborazione del linguaggio naturale. Il miglioramento della precisione dell'etichettatura potrebbe implicare un risparmio di tempo e di fondi e la possibilità di addestrare algoritmi specifici su piccoli corpora che riflettano una specifica domanda di ricerca come, nel nostro esempio, «tweet che reagiscono alle credenze religiose».

Ad esempio, l'opposizione tra polarità negativa e positiva potrebbe essere sostituita con la categoria timica, che «si articola a sua volta in euforia/disforia (con aforia come termine neutro) e gioca un ruolo fondamentale nella trasformazione dei microuniversi semantici in assiologie» (Greimas e Courtés 1979, p. 360). Tuttavia, per la ricchezza e l'articolazione delle categorie semantiche sono necessarie tecniche di deep learning più sofisticate.

Lo stesso insieme di categorie semiotiche potrebbe essere ulteriormente esteso all'analisi di dati non verbali, fornendo un vocabolario standard alle scienze sociali e alla teoria dell'informazione (Akrich e Latour 1992).

La domanda di ricerca più interessante che emerge dai nostri risultati riguarda lo stato della rete neurale, quando considerata come attore sociale (Akrich e Latour 1992). Nell'Unione Europea, con la risoluzione del Parlamento Europeo del 16 febbraio 2017 (2015/2103(INL)) è iniziato un dibattito politico sulla possibilità di attribuire uno status giuridico alle “intelligenze artificiali”. Il Parlamento ha chiesto alla Commissione europea di creare «uno status giuridico specifico per i robot nel lungo periodo, in modo che almeno i robot autonomi più sofisticati possano essere affermati come aventi lo status di persone elettroniche incaricate di riparare i danni che possono causare». Partendo dall'etichetta “Intelligenza artificiale”, le strategie retoriche possono essere utilizzate nei testi per attribuire agli algoritmi valori semantici sia non umani (es. “artificiale”) che umani (“intelligenza”) (Galofaro 2021, pp. 10–11). L'attribuzione di uno status giuridico a tali algoritmi non dipende solo dal conflitto tra le narrazioni — definendole “intelligenti” o riducendole a “metodi statistici”, ma anche dal loro coinvolgimento dialettico nelle relazioni sociali strumentali o empatiche.

6. The network within

Le prospettive ambivalenti sull'“Intelligenza artificiale” possono forse cambiare se adottiamo la nozione di *network within*:

A stringent actor–network approach cannot determine a priori which are the actors–knots of the network in question; above all, it cannot postulate their dimension, scale, consistency, and cohesion. In effect, if each knot is itself a network, one may often have to describe and define the “network within” that constitutes each actor. Only by taking into consideration this network within, is it possible to account for the various “translations” an actor faces (or, in other words, for the ways in which the network constituting a given actor is re–configured upon its entering into a relationship with other actors–networks) (Parolin and Mattozzi 2020, p. 48).

Invece di considerare i nodi inclusi nell'actor network come preesistenti ad essa, essi devono essere visti come prodotti dalle relazioni che la legano all'oggetto. Ogni attore, come l'utente di un'auto guidata da un'IA, il prigioniero in attesa che un'IA decida sulla libertà vigilata, il bambino che interagisce con un'infermiera IA, il Parlamento europeo che discute del fantasma nella macchina dell'IA, non è un hub connesso e ricollegato tramite fili ad altri attori, ma il risultato di un processo di individuazione locale risultante dalle relazioni globali. A loro volta, gli attori sono da considerarsi, ricorsivamente, reti.

7. Conclusioni: l'IA come nuovo mito

La così detta "Intelligenza Artificiale" può essere considerata un nuovo mito, in grado di esercitare una prensione sui così detti nuovi media, rappresentati da Twitter nel nostro caso studio. Se l'elaborazione di un mito risponde al tentativo di risolvere una contraddizione interna alla cultura di un'epoca e di una società, l'algoritmo incarna il tentativo di portare la competenza umana oltre i limiti dell'individuo, i quali confliggono con le dimensioni globali del fenomeno comunicativo. Come si è scritto in 2.1, lo sviluppo stesso dell'informatica avviene con l'obiettivo di manipolare in pochi minuti corpora larghissimi, non di rado riferibili agli studi umanistici, rivolgendosi a quella stessa scienza statistica che è la base dell'apprendimento delle reti neurali (par. 4).

Osservando il nostro caso studio, si deve ammettere che anche la comunicazione di Papa Francesco ha un carattere mitico. Essa cerca di porre rimedio alla contraddizione tra il sapere economico e quello ecologico e tra due valori profondi reperibili in essi: il monoteismo, che vede un intervento trasformativo del divino nella storia, e la sacralità della natura, della temporalità ciclica della riproduzione. Per farlo, come si è visto al par. 2.2, il Papa reinventa una tradizione: amplifica alcuni temi presenti nel *Cantico delle creature* e ne narcotizza altri. Figure come madre terra e sorella acqua sottolineano la creaturalità dell'ambiente cui siamo legati, mentre la povertà, sorella del santo, è il principale problema del Papa. A tal proposito, Le Goff (1999, p. 34) paragona San Martino e San Francesco, stabilendo che all'epoca del

primo la contraddizione principale percepita era evidentemente la distribuzione dei beni (è la figura del mantello diviso col povero), mentre all'epoca del secondo la contraddizione è l'accettazione o il rifiuto dell'economia monetaria.

L'operazione mitopoietica di Papa Francesco si incontra e si scontra con i nuovi miti: la sua comunicazione sociale è ripresa, rilanciata e commentata da altri attori individuali o collettivi (uomini politici, leader d'opinione, singoli utenti, gruppi ecologisti) che in tal modo la sanzionano positivamente o negativamente. Tali *destinanti* sanzionatori non sono solo *antropomorfi* (umani), ma anche *non pienamente umani*: questo è lo statuto problematico delle reti neurali che analizzano la comunicazione e la classificano in termini di efficacia, grado di soggettività, polarità.

L'inconsistenza del mito della rete neurale è dimostrata dal fatto che tali reti sono addestrate sulla base di nozioni filosoficamente mal formulate, tanto vaghe quanto sono precisi gli algoritmi di apprendimento. Il nostro esperimento dimostra che la semiotica può rendersi utile se non altro per chiarire meglio a cosa corrispondano, nel testo, nozioni come polarità e soggettività per migliorare l'addestramento, dunque per intervenire sul lato "umano" della rete.

In ogni caso, l'addestramento di una rete neurale è una forma di scrittura il cui risultato è l'archivio di una competenza umana collettiva (es. la competenza media di un gruppo di studenti) al fine di replicarne le prestazioni, quali emettere una sanzione finale positiva o negativa su un determinato testo. Da un punto di vista semiotico, una rete neurale è un termine complesso, sia umano sia non umano.

Come abbiamo visto nel nostro caso studio, tale attore ibrido incarna un destinante cui la cultura attribuisce la proprietà di considerare il flusso della comunicazione ecologista nel proprio complesso classificandola come oggettiva o soggettiva e contribuendo a misurarne l'efficacia e l'impatto sull'insieme della società: una meta-rappresentazione dell'autocomunicazione sociale sul tema del proprio rapporto con l'ambiente. I flussi di comunicazione mitica che attraversano l'ibrido tra natura e cultura chiamato "Gaia" da Bruno Latour (2017) costituiscono un meccanismo cibernetico di regolazione in grado di restituire un significato al nostro rapporto con l'ambiente? Questo resta da vedere.

Riferimenti bibliografici

- Akrich M. e B. Latour (1992) “A Summary of Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies”, in W.E. Bijker e J. Law (a cura di), *Shaping Technology/Building Society*, MIT press, Cambridge MA, 259–264.
- Baker P. (2006) *Using corpora in Discourse Analysis*, Continuum, London–New York.
- Benveniste, E. (1966–1974) *Problèmes de linguistique générale*, Gallimard, Paris, vol. 1.
- Beretta E., A. Santangelo, B. Lepri, A. Vetrò e J.C. De Martin (2019) *The invisible power of fairness. How machine learning shapes democracy*, in *Proceedings of the 32nd Canadian Conference on Artificial Intelligence*, Springer, Berlin, 238–250.
- Bianchi F., D. Nozza e D. Hovy (2021) “FEEL–IT: Emotion and Sentiment Classification for the Italian Language”, in *Proceedings of the 11th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Association for Computational Linguistics*, “ACL–Anthology”: 76–83, <https://aclanthology.org/2021.wassa-1.8> (consultato il 28 febbraio 2022).
- Branca V. (1950) *Il cantico di frate sole. Studio delle fonti e testo critico*, Olschki, Firenze.
- Burrell J. (2015) *How the Machine Thinks: Understanding Opacity in Machine Learning Algorithms*, <http://ssrn.com/abstract=2660674> (consultato il 1 marzo 2022).
- Ceriani G. (2016) “Reputazione, popolarità, sentiment. Discorsi sul sesso in rete e problematiche di ricerca integrata” in G. Ferraro e A.M. Lorusso (a cura di), *Nuove forme d’interazione: dal web al mobile*, Libellule, Lecce, 43–52.
- Contini G. (1995) *Poeti del Duecento. Testi arcaici, scuola siciliana, poesia cortese*, Ricciardi–Mondadori, Milano.
- Di Martino L.A. (2012) *Per uno studio sulla procedura delle concordanze*, “Enthymema”, VII (2012), 183–200.
- Galofaro F. (2021) *On the Juridical Relevance of the Phenomenological Notion of Person in Max Scheler and Edith Stein*, “International Journal for the Semiotics of Law”, 35: 1317–1331, <https://doi.org/10.1007/s11196-021-09823-z> (consultato il 28 febbraio 2022).

- Gigliozzi G. (2003) *Introduzione all'uso del computer negli studi letterari*, Bruno Mondadori, Milano.
- Giuliano L. (2013) *Il valore delle parole. L'analisi automatica dei testi in Web 2.0*, Data Science (Dipartimento di Scienze statistiche), Roma, file:///E:/Convegni%20e%20saggi/1%20Convegni%20e%20conferenze/2021.12.21%20Circe%20workshop/ValoreDelleParoleV2a.pdf (consultato il 28 febbraio 2022).
- Greimas A.J. e J. Courtés (1979) *Sémiotique: dictionnaire raisonné de la théorie du langage*, Hachette, Paris (trad. it. *Semiotica: dizionario ragionato della teoria del linguaggio*, Bruno Mondadori, Milano, 2007).
- Grubbs D. (2020) *Understanding Political Twitter: Using tweet sentiment analysis to understand the global political atmosphere*, <https://towardsdatascience.com/understanding-political-twitter-ce3476a38377> (consultato il 28 febbraio 2022).
- Kumar A. e D. Gupta (2021) "Sentiment Analysis as a Restricted NLP Problem", in F. Pinarbasi e M. Nurdan Taskiran (a cura di), *Natural Language Processing for Global and Local Business*, IGI Global, Hershey PA, 65–97.
- Latour B. (2017) *Facing Gaia: Eight Lectures on the New Climatic Regime*, Polity Press, Medford MA.
- Le Goff, J. (1999) *Saint François d'Assise*, Gallimard, Paris (trad. it. *San Francesco d'Assisi*, Laterza, Bari, 2002).
- Liu B. (2021) *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, San Rafael, California.
- Lorusso A.M. e P. Peverini (a cura di) (2017) *Il racconto di Francesco. La comunicazione del Papa nell'era della connessione globale*, Luiss University Press, Roma.
- Lucatti E. (2022) *Di foglie, aria ed espressione: le linee sciamaniche della semiotica e la questione del suo statuto scientifico*, Esculapio, Bologna.
- Mirjalili V. e S. Raschka (2019) *Python Machine Learning*, Packt Publishing, Birmingham.
- Parolin L. e A. Mattozzi (2020) *How to account for tacit knowledge in innovation processes: the concept of "network within"*, "RASK", 51: 39–54.
- Pozzi G. (1971) *Rileggendo il Cantico di frate sole*, "Messaggero serafico", 60, 8: 2–32.
- Putnam H. (2002) *The collapse of the fact/value dichotomy and other essays*, Harvard University Press, Cambridge MA.

- Santangelo A. (2016) *Twitter e l'analisi semiotica di un programma televisivo*, in G. Ferraro e A.M. Lorusso (a cura di), *Nuove forme d'interazione: dal web al mobile*, Libellule, Lecce, 67–86.
- Starobinski J. (1961) *L'Œil vivant: Corneille, Racine, La Bruyère, Rousseau, Stendhal*, Gallimard, Paris.
- Viveiros de Castro E. (2017) *A inconstância da alma selvagem*, Ubu editora, São Paulo.
- Ziqi Z. e L. Lei (2018) *Hate Speech Detection: A Solved Problem? The Challenging Case of Long Tail on Twitter*, "Semantic Web", 1 (0): 1–5.