



---

# DISAMBIGUARE LE RETI SOCIALI

FEDERICO MARTELLI – ROBERTO NAVIGLI

*Prendendo le mosse dal concetto di rete sociale e di social network, solitamente impiegati in due accezioni distinte in lingua italiana, si presentano nuove prospettive di ricerca nell'ambito dell'elaborazione del linguaggio naturale e, in particolare, della disambiguazione automatica applicata alle reti sociali per eccellenza: i social network.*

**N**ell'era dei grandi avanzamenti tecnologici e del folgorante successo di internet, la nozione di rete sociale assume sempre maggiore importanza. La formalizzazione di tale concetto risale ai primi decenni del Novecento, quando la ricerca in ambito sociologico e antropologico iniziò a studiare le interazioni fra individui e le relative implicazioni sul loro comportamento. Nell'articolo *Class and Committees in a Norwegian Island Parish*, John Arundel Barnes scrive:

I find it convenient to talk of a social field [...] as a network. The image I have is of a set of points some of which are joined by lines. The points of the image are people, or sometimes groups, and the lines indicate which people interact with each other. We can of course think of the whole of social life as generating a network of this kind<sup>1</sup>.

Originariamente, il termine rete sociale veniva impiegato per indicare un insieme di individui connessi tra loro da relazioni. Circoscritto a un uso specialistico fino a qualche anno fa, esso, e soprattutto il suo equivalente inglese *social network*, si è ormai cristallizzato nell'uso entrando a far parte del vocabolario di quasi ogni parlante.

1. BARNES 1954, p. 43.

Oggi tale termine preserva il suo significato originario, mentre il suo anglicismo, ormai penetrato nella nostra lingua, fa principalmente riferimento al web e alle sue molteplici possibilità di aggregazione sociale, designando piattaforme quali Facebook, Twitter e Instagram, il cui scopo è quello di favorire la comunicazione tra i propri utenti, permettere loro di instaurare nuovi contatti e intessere relazioni sociali attraverso internet. Gradualmente, le reti sociali e i social network divengono oggetto di approfondimento di numerose discipline, dagli studi umanistici alle 'scienze dure'. Data la loro grande attualità, il presente scritto li pone al centro della trattazione per come visti dalle nuove prospettive di ricerca introdotte dall'intelligenza artificiale e, in particolare, da una sua branca, l'elaborazione del linguaggio naturale.

### I SOCIAL NETWORK

Reti sociali per antonomasia, i social network iniziano a diffondersi con l'arrivo del nuovo millennio e divengono rapidamente il luogo preferenziale dove poter comunicare, esprimere opinioni, scambiare informazioni e, più in generale, vivere una vera e propria vita sociale in formato digitale. Quale specchio della realtà in cui viviamo, i social network riflettono gli aspetti positivi e negativi della vita, amplificandone la risonanza. Naturalmente, tutto ciò apre a molteplici scenari. Innanzitutto, dato il loro impatto sulla società, i social sono visti con grande interesse da aziende e organizzazioni, non di rado disposte a investire ingenti somme in cambio di preziose indicazioni riguardo a preferenze e interessi degli utenti. La vendita di informazioni sensibili è divenuta oggi uno dei business più redditizi, tanto che i dati vengono spesso definiti «il nuovo petrolio».

Ormai non sorprendono più vicende come quella di Cambridge Analytica, che ha visto Facebook nuovamente coinvolta in uno scandalo di cessione di dati. Sfortunatamente, tali scandali non costituiscono casi isolati, al contrario colpiscono diversi colossi del cosiddetto BigTech quali Google e Twitter. La frequenza di tali episodi evidenzia l'importanza di tutelare la privacy attraverso la sua inclusione nel tessuto normativo. Purtroppo, nonostante l'attualità del dibattito e i tentativi di regolamentazione, pare che i grandi interessi industriali prevalgano sulla tutela della sfera privata di ognuno di noi. Oggi è sufficiente navigare in internet qualche minuto per ottenere una profilazione completa dell'utente, il quale potrà ricevere annunci

pubblicitari targettizzati sulla base del traffico generato. Un altro aspetto allarmante riguarda l'utilizzo illecito delle piattaforme. Non immuni alle cattive intenzioni e al crimine, i social network, talvolta, vengono utilizzati da malviventi per il conseguimento dei propri fini. Sono ben note vicende di mobbing e stalking consumatesi per mezzo di questo canale e talvolta culminate in episodi tragici. D'altro canto, i social network possono essere anche destinati a finalità di ricerca. Ad esempio, l'enorme quantità di dati in essi contenuta può essere impiegata per i nobili scopi della scienza nelle sue più svariate discipline: dalla sociologia alla psicologia, dall'economia al diritto, dalla matematica all'informatica.

Una delle sfide più grandi, ai fini dell'indagine scientifica, è rappresentata proprio dalle modalità di analisi di una simile mole di dati e dalla loro difficile interpretazione. Alla luce dei suoi recenti sviluppi, l'elaborazione del linguaggio naturale, in particolare grazie alla disambiguazione automatica multilingue, potrebbe fornire risposte efficaci a tale problematica.

### L'ELABORAZIONE DEL LINGUAGGIO NATURALE E LA DISAMBIGUAZIONE AUTOMATICA MULTILINGUE

L'Nlp (*natural language processing*) è una branca dell'intelligenza artificiale avente come oggetto di studio il trattamento automatico del linguaggio naturale. In altre parole, si occupa d'individuare le tecniche attraverso le quali un computer può rappresentare, comprendere e riprodurre una lingua. Richiamando Daniel Jurafsky e James Martin, lo scopo dell'Nlp è:

To get computers to perform useful task involving human language, tasks like enabling human-machine communication, improving human-human communication, or simply doing useful processing of text and speech<sup>2</sup>.

Nell'Nlp, ogni attività eseguita dal computer viene indicata con il termine inglese *task*. Tra i più noti, notevole rilevanza assumono la disambiguazione e la traduzione automatica. La disambiguazione automatica, meglio nota come *word sense disambiguation* (Wsd), consiste nel determinare il significato corretto di una parola in uno specifico contesto. Oggetto di studio è l'ambiguità linguistica e, nello specifico, la natura polisemica delle parole di una lingua. Infatti, esse possono assumere diversi significati o, in termini linguistici, a un significante possono corrispondere più significati. Si consideri, ad esempio, la parola «piano». In taluni contesti assume

2. JURAFSKY – MARTIN 2006, p. 1.

il significato di lentamente, in altri designa uno strumento musicale, in altri ancora indica un concetto geometrico astratto. Quale scegliere? Appare evidente che, in base al contesto, il significato muti. Ma come può un computer distinguere correttamente tra tutti i significati possibili di una parola? Questa domanda pone diverse sfide: è evidente che disambiguare richieda una profonda conoscenza del mondo, da un lato, e una padronanza linguistica, dall'altro. Data la sua complessità, il Wsd viene considerato un task *AI-complete*, ovvero:

By analogy to NP-completeness in complexity theory, 'AI-complete' is a term, first coined by Fanya S. Montalvo, to indicate that the difficulty of a computational problem is equivalent to solving the central AI problem, i.e., making computers as intelligent as people<sup>3</sup>.

Ciononostante, diversi sistemi di disambiguazione automatica sono stati sviluppati con differenti approcci e strategie. A oggi, i più efficienti sono in grado di scalare su più lingue simultaneamente, ossia, sono capaci di disambiguare una parola indipendentemente dalla lingua di appartenenza. A tal fine, sono necessarie particolari risorse linguistiche come le cosiddette reti semantiche, ossia grafi in cui ciascun nodo rappresenta un concetto, mentre le connessioni tra nodi esplicitano le relazioni tra concetti. Babelfy<sup>4</sup>, noto sistema di disambiguazione basato su BabelNet<sup>5</sup> – tra le più grandi reti semantiche nonché dizionario enciclopedico multilingue – ha considerevolmente migliorato la disambiguazione automatica multilingue, ottenendo risultati significativamente migliori rispetto ad altri sistemi, allo stato dell'arte. La svolta introdotta da BabelNet è proprio da ricercarsi nel multilinguismo:

La vera sfida che ci poniamo con BabelNet è di fornire una base di conoscenza che faccia del multilinguismo la propria forza, assumendo l'esistenza di un unico concetto per tutte le lingue in cui esso può essere espresso, ma anche per quelle lingue in cui non esiste un termine per esprimerlo<sup>6</sup>.

3. MALLERY 1994, p. 47.

4. <<http://babelfy.org/>> [11-10-2018].

5. <<https://babelnet.org/>> [11-10-2018].

6. GROSSI – NAVIGLI 2017, p. 67.

Il successo di simili sistemi è testimoniato dalla vertiginosa crescita di un settore strategico che avrà un impatto sempre più forte sulle nostre vite. Già oggi sono molteplici i campi di applicazione della disambiguazione e nuove possibilità di utilizzo si esplorano ormai quotidianamente. La disambiguazione multilingue dei social network potrebbe aprire interessanti frontiere di analisi delle relazioni sociali e fornire utili informazioni di tipo semantico per successive indagini scientifiche. Uno degli ostacoli maggiori che dovrà essere superato nel contesto social riguarda la corretta interpretazione della varietà linguistica utilizzata che, come si vedrà in seguito, presenta caratteristiche che la rendono, talvolta, particolarmente difficile da decifrare.

### INSEGNARE AI COMPUTER UNA NUOVA LINGUA

Grazie alle strategie e ai modelli individuati dall'analisi delle reti sociali – moderna metodologia di indagine delle relazioni tra individui – è possibile analizzare automaticamente le connessioni tra utenti. Tuttavia, un computer sarebbe in grado di comprendere il significato di ciò che gli utenti dicono? Sarebbe possibile, ad esempio, individuare i concetti più ricorrenti, indipendentemente dalla lingua dell'utente? Uno dei maggiori ostacoli da superare è rappresentato dalla cosiddetta «lingua dei social», definizione con la quale s'intende la varietà linguistica utilizzata all'interno dei social network. Questa varietà non è presente solo nella lingua italiana bensì in numerosi altri idiomi e si colloca in seno alla nota *comunicazione mediata dal computer* (Cmc) che studia, in tutte le sue sfaccettature, un'originale forma di comunicazione: quella che avviene tramite gli strumenti digitali. Nella tradizione linguistica italiana, molteplici sono le denominazioni utilizzate per indicare la lingua dei social network e, più in generale, l'italiano del mondo digitale, tra cui citiamo: «italiano digitato»<sup>7</sup>, «scrittura liquida»<sup>8</sup> e «e-taliano»<sup>9</sup>. A livello globale, invece, un termine impiegato per indicare il gergo utilizzato nei social network è *netspeak*, di cui illustriamo i principali tratti caratteristici. La scrittura tipica dei social è interattiva e dialogica, ovvero avviene in un contesto in cui sia il mittente sia il destinatario sono parte attiva. Ciò comporta che il significato possa essere ricostruito solo in un contesto comunicativo spazio-temporale ben definito. La sintassi risulta frammentata e destrutturata. Sono numerose le allusioni e gli anglicismi, mentre è quasi del tutto assente la punteggiatura<sup>10</sup>. Inoltre, si fa largo uso di simboli quali le cosiddette *emoticon*, che intendono trasmettere efficacemente uno stato d'animo,

7. ANTONELLI 2009.

8. FIORENTINO 2011.

9. ANTONELLI 2014.

10. D'ACHILLE 2017.

o il cancelletto (#) utilizzato per creare *hashtag*, ossia espressioni sintetiche impiegate nei social network per etichettare i post. Gli hashtag riassumono, a volte scherzosamente, un fenomeno, un concetto o un avvenimento. Spesso sono di difficile interpretazione, in quanto possono contenere acronimi e abbreviazioni note solo a un gruppo ristretto di utenti, oppure perché comprimono in una locuzione unica più parole non separate da spazi. Inoltre, hashtag particolarmente ambigui quali #adr, #wow, #self ecc. per la loro interpretazione richiedono un'attenta analisi del contesto in cui appaiono. Grazie ai moderni algoritmi e a risorse linguistiche sempre più complete, la disambiguazione dispone oggi degli strumenti necessari per affrontare un simile task.

Per 'insegnare' a una macchina a comprendere il significato di ciò che gli utenti dicono nei social, non è sufficiente una mera analisi statistica, ma è necessario ricorrere ad approcci e strategie ben più avanzate.

Il nostro team di ricerca presso il dipartimento di informatica dell'Università Sapienza di Roma sta sviluppando un efficace sistema di disambiguazione multilingue dei social network, basato sulla tecnologia introdotta da BabelNet. L'obiettivo è, da un lato, l'arricchimento della rete semantica con nuove lessicalizzazioni provenienti dal mondo dei social quali, appunto, hashtag e gergalismi della Rete; dall'altro, lo sviluppo di un innovativo algoritmo di disambiguazione che sia in grado di elaborare e collegare tali informazioni. Un simile sistema potrebbe trovare applicazione in numerosi ambiti. Grazie a questa nuova tecnologia, capace di analizzare e comprendere i contenuti presenti nei social network, sarebbe possibile individuare sia le tendenze principali sia quelle emergenti, ma anche prevedere l'insorgere di orientamenti potenzialmente pericolosi



#### BIBLIOGRAFIA

- G. ANTONELLI, *Scrivere e digitare* in T. GREGORY (diretta da), *Treccani XXI secolo*, II, Roma 2009, pp. 243-253.
- G. ANTONELLI, *L'e-taliano: una nuova realtà tra le varietà linguistiche italiane?*, atti del XII Congresso Società Internazionale di Linguistica e Filologia Italiana (Helsinki, 18-20 giugno 2012), in E. GARAVELLI ET AL. (a cura di), *Dal manoscritto al web: canali e modalità di trasmissione dell'italiano*, II, Franco Cesati, Firenze 2014, pp. 537-556.
- J.A. BARNES, *Class and Committees in a Norwegian Island Parish*, «Human relations» VII (1954) 1, pp. 39-58.
- P. D'ACHILLE, *L'italiano contemporaneo*, il Mulino, Bologna 2003.
- P. D'ACHILLE, *I social network e la lingua italiana, tra neologismi e anglicismi*, Accademia della Crusca: <<http://www.accademiadellacrusca.it/it/scaffali-digitali/articolo/social-network-lingua-italiana-neologismi-anglicismi>> [11-10-2018].
- G. FIORENTINO, *Scrittura liquida e grammatica essenziale*, in U. CARDINALE (a cura di), *A scuola d'italiano a 150 anni dall'Unità*, il Mulino, Bologna 2011, pp. 219-241.
- T. GROSSI – R. NAVIGLI, *Quando la macchina comprende ciò che diciamo*, «Gnosis» XXIII (2017) 2, pp. 62-69: <[http://gnosis.aisi.gov.it/gnosis/Rivista51.nsf/ServNavig/51-23.pdf/\\$File/51-23.pdf?OpenElement](http://gnosis.aisi.gov.it/gnosis/Rivista51.nsf/ServNavig/51-23.pdf/$File/51-23.pdf?OpenElement)> [11-10-2018].
- D. JURAFSKY – J. MARTIN, *Speech and language processing*, Prentice Hall, Upper Saddle River 2006.
- J.C. MALLERY, *Thinking about Foreign Policy: finding an appropriate role for artificially intelligent computer*, Department of Political Science and Artificial Intelligence Laboratory, MIT, Cambridge (1994): <[cite-seerx.ist.psu.edu/viewdoc/download?doi=10.1.1.50.3333&rep=rep1&type=pdf](http://seerx.ist.psu.edu/viewdoc/download?doi=10.1.1.50.3333&rep=rep1&type=pdf)> [26-10-2018].