

L'integrazione di informazioni contestuali e linguistiche nel riconoscimento automatico dell'ironia

Lorenzo Bernardini, Irina Prodanof

Dept. of Humanities, University of Pavia

lorenzo.bernardini01@ateneopv.it; irina.prodanof@unipv.it

Abstract

Italiano. L'ironia verbale è una figura retorica altamente complessa che appartiene al livello pragmatico del linguaggio. Finora, tuttavia, tutti i tentativi in campo computazionale volti al riconoscimento automatico dell'ironia si sono limitati a ricercare indizi linguistici che potessero segnalarne la presenza senza considerare fattori pragmatici e contestuali. In questo lavoro si è cercato di valutare la possibilità di integrare semplici fattori contestuali computabili con informazioni di tipo linguistico al fine di migliorare l'efficacia dei sistemi di riconoscimento automatico di ironia nei commenti dei lettori di quotidiani online.

English. *Verbal irony is a highly complex figure of speech which belongs to the pragmatic level of language. So far, however, all the computational approaches aimed at automatically recognizing irony have only searched for linguistic cues which could signal the presence of irony without taking into account pragmatic and contextual factors. In this work we have considered the possibility of integrating simple contextual and computable factors with linguistic information in order to improve the performance of irony detection systems in the comments of readers of web newspapers.*

1 Introduzione

L'ironia verbale è una figura retorica molto complessa che si colloca al livello pragmatico del linguaggio. Per quanto un ironista possa servirsi di elementi fonologici, prosodici, morfologici,

lessicali, sintattici e semantici per produrre ironia, quest'ultima non è una proprietà interna all'enunciato stesso e non è determinata dalle sue caratteristiche formali.

L'ironia è piuttosto un fenomeno interpretativo legato alle aspettative che un ascoltatore sviluppa riguardo alle intenzioni dell'autore di un enunciato prodotto in uno specifico contesto a partire da uno sterminato insieme di informazioni enciclopediche e contestuali.

2 Potenziali applicazioni e stato dell'arte

Il riconoscimento automatico di ironia porterebbe benefici nel campo della Sentiment Analysis. Non individuare l'intenzione ironica dell'autore di un enunciato può portare infatti a gravi fraintendimenti riguardo alle sue opinioni. Poiché l'uso dell'ironia è pervasivo e quantitativamente non trascurabile in molti contesti online, un solido sistema di Sentiment Analysis dovrebbe considerare i problemi legati all'uso dell'ironia e sviluppare metodi per il suo riconoscimento automatico.

Il riconoscimento automatico di ironia è tuttavia ancora a uno stadio pionieristico. Ad oggi tutti gli approcci computazionali al riconoscimento di ironia hanno cercato esclusivamente di individuare elementi linguistici interni al testo (grafici, semantici e lessicali) che potessero indicarne la presenza. Carvalho et al. (2009) hanno proposto una lista di indizi espliciti per individuare l'ironia nei commenti dei lettori di un quotidiano online in lingua portoghese (virgolette intorno ad aggettivi o nomi positivi, espressioni/simboli che indicano una risata nella scrittura (e.g. *lol*), interiezioni usate per esprimere sentimenti positivi (e.g. *viva*) e punteggiatura marcata). Reyes et al. (2013) hanno invece costruito un complesso modello considerando 11 parametri linguistici posti su 4 dimensioni (*signatures, unexpectedness, style, emotional scenarios*) per rilevare l'ironia nei *tweets*.

Per quanto questi e altri lavori abbiano ottenuto

risultati parzialmente soddisfacenti, tuttavia nessuno di essi ha mai preso in considerazione gli elementi contestuali che costituiscono il cuore del fenomeno dell'ironia, limitandosi a considerare esclusivamente il materiale linguistico alla ricerca di caratteristiche formali che possano segnalare la presenza. Tuttavia molto spesso l'informazione linguistica da sola non è sufficiente a stabilire se un enunciato è ironico o meno, tanto che lo stesso materiale linguistico, se proveniente da fonti o contesti differenti, può essere interpretato come ironico o meno.

Il contesto gioca dunque un ruolo fondamentale. Ma esistono fattori contestuali computabili che possano rivelarsi efficaci per riconoscere automaticamente l'ironia in testi online?

3 Analisi di una comunità online

Per individuare quali fattori contestuali possano rivelarsi computabili si è deciso di svolgere un'analisi empirica¹ riguardo all'uso dell'ironia all'interno di una comunità virtuale. L'analisi è stata realizzata sui commenti lasciati in calce agli articoli dai lettori abituali de *Il Fatto Quotidiano* online. L'analisi dei commenti si è svolta in tre fasi distinte.

Prima fase. Innanzitutto si è cercato di cogliere la significatività in termini quantitativi dei commenti ironici rispetto al totale dei commenti, verificando se i commenti ironici presentassero indizi linguistici espliciti della presenza di ironia.

Sono stati analizzati 20 commenti casuali a 6 articoli di giornale. Per ogni commento si è verificato:

- se l'autore fosse un commentatore abituale (oltre 500 commenti sul sito).
- se il commento fosse ironico o meno.
- se fossero presenti gli indizi espliciti proposti da Carvalho et al. (2009).

I numeri complessivi di quest'analisi empirica non possono certo costituire un campione statistico valido, ma sembrano tuttavia suggerire alcune tendenze degne di nota. Innanzitutto si è rilevato come l'ironia fosse presente in un numero considerevole di commenti. Almeno un quarto dei commenti a ciascun articolo presentava (almeno in parte) ironia, per un totale di 39 commenti su 120

(15 i casi di dubbia attribuzione). Questo dato suggerirebbe che l'ironia non debba essere trascurata come un fenomeno marginale in Sentiment Analysis.

In secondo luogo si è osservato come la maggior parte dei commentatori fossero abituali (85 su 120). Questo fatto è un presupposto fondamentale per avvalorare l'assunto che essi possano conoscersi tra loro e formare così una comunità.

Per quanto riguarda invece l'uso di indizi espliciti si è notato come essi comparissero soltanto in una minoranza, tuttavia significativa, dei commenti ironici (16/39). Ciò mostra che, servendosi solamente di questi indizi linguistici, molti dei commenti ironici non potrebbero assolutamente essere individuati. Per giunta la stessa serie di indizi appariva in quantità non trascurabile anche in commenti non ironici, seppure in proporzione minore (11/66). Pertanto l'uso di questi indizi linguistici, pur se inefficaci in molti casi di ironia e talvolta addirittura fuorvianti, non deve essere accantonato, vista comunque la maggior probabilità che essi siano presenti in un commento ironico rispetto ad uno non ironico.

Seconda fase. Nella seconda fase di analisi si è indagata l'attitudine ironica complessiva di una serie di commentatori abituali. Sono stati analizzati gli ultimi 25 commenti postati da 22 commentatori abituali scelti casualmente. Per ognuno di essi sono stati raccolti i seguenti dati:

- numero di commenti ironici tra gli ultimi 25 postati dall'utente.
- numero di commenti, tra quelli ironici sopra individuati, che presentassero gli indizi proposti da Carvalho et al. (2009).

In questa fase di analisi si è osservato come sembri possibile individuare un ristretto gruppo di commentatori che si contraddistinguono dagli altri poiché pare manifestare un'attitudine spiccatamente ironica (oltre 80% dei commenti sono interamente ironici). La maggioranza degli utenti sembra invece affidarsi più sporadicamente all'uso di ironia e spesso l'uso è limitato a parti ridotte del testo, perlopiù all'inizio o nel finale del commento. Si è notato inoltre come gli indizi linguistici siano pressoché assenti nell'ironia dei commentatori "spiccatamente" ironici, mentre sono decisamente più frequenti nei commenti ironici degli altri utenti. Volendo deliberatamente semplificare, sembrano delinearsi due tendenze di comportamento particolarmente significative:

¹Per una discussione più approfondita e maggiori dettagli sul lavoro di analisi si rimanda a Bernardini (2014). L'intero corpus è disponibile su richiesta.

- Commentatori quasi sempre ironici in quasi tutti i commenti e nell'inezienza del commento. La loro ironia è raramente "marcata" da indizi linguistici espliciti.
- Commentatori poco ironici. L'ironia è usata più di rado e spesso solo nella frase iniziale o conclusiva. La loro ironia presenta più frequentemente indizi linguistici espliciti.

Terza fase. La terza fase di analisi si è concentrata sul comportamento dei commentatori più ironici della comunità. Inizialmente sono state osservate le interazioni avvenute tra i quattro utenti più ironici individuati nella seconda fase di analisi e gli altri commentatori nell'arco di 3 mesi.

Attraverso quest'analisi sono emerse due considerazioni interessanti. In primo luogo si è visto come gli utenti abituali riconoscano e siano spesso consapevoli dello stile "spiccatamente ironico" di alcuni utenti. In secondo luogo è emerso come gli utenti "spiccatamente ironici" tendano a rispondere con più frequenza a commenti ironici che a commenti non ironici rispetto agli altri utenti.

Un altro aspetto peculiare dei commentatori spiccatamente ironici riguarda le loro scelte di *nickname* e *avatar*. Questo non è particolarmente sorprendente poiché il nome e l'immagine scelta rappresentano la *faccia* che un utente vuole mostrare di sé. Pertanto un utente che desideri apparire molto ironico tenderà plausibilmente a registrarsi con una *faccia* che risulti simpatica. Si è perciò cercato di verificare empiricamente se *nickname* e *avatar* divertenti potessero risultare indicativi per individuare commentatori spiccatamente ironici. Sono stati individuati manualmente 13 utenti con una *faccia* riconducibile a contesti divertenti. Di questi, 8 manifestavano un'attitudine spiccatamente ironica. In particolare tra i 7 utenti con una *faccia* riconducibile a contesti più specificamente satirici, 6 di essi apparivano spiccatamente ironici. Sembrerebbe pertanto che la scelta di usare un *nickname* e/o un *avatar* riconducibili a contesti divertenti, o tanto meglio satirici, possa davvero essere un ottimo indicatore per individuare quegli utenti che esibiscono una attitudine molto ironica all'interno della comunità.

4 Verso un modello integrato di informazioni linguistiche e contestuali

Le informazioni contestuali che un ascoltatore può utilizzare per attribuire ironia a un enunciato sono soggettive e potenzialmente infinite. Al momento

è però impensabile ipotizzare una macchina che possa contenere un insieme infinito e indefinito di credenze e conoscenze dal quale, in qualche modo, riconoscere l'ironia in un testo.

L'obiettivo deve essere dunque cercare di individuare ed estrarre alcune semplici informazioni contestuali computabili per poterle poi integrare con l'informazione linguistica. Il modello qui proposto integrerebbe informazioni di tre tipologie.

4.1 Informazioni relative all'attitudine ironica dell'enunciato

Dall'analisi dell'attitudine ironica della comunità online (Sez. 2, Seconda Fase) emergevano tendenze di comportamento differenti tra i commentatori quasi sempre ironici e gli altri utenti. In questo modello la prima operazione consisterebbe dunque nel riconoscere l'attitudine ironica generale di un commentatore per avviare una differente trattazione del materiale linguistico. Per fare ciò si dovrebbe innanzitutto considerare la *faccia* esibita dall'utente. Attraverso ricerche automatiche per parole o immagini sarebbe semplice collegare i *nickname* o gli *avatar* a contesti satirici e ipotizzare con buona probabilità che il commentatore sia spiccatamente ironico. Inoltre, poiché i commentatori molto ironici tendono con più probabilità a rispondere ad altri commenti ironici, si potrebbe estrarre manualmente una breve lista per considerare i commenti ai quali essi hanno replicato come più probabilmente ironici. Se un alto numero di commenti di uno stesso utente fosse poi commentato da più commentatori di questa lista, tale utente sarebbe inserito a sua volta nella lista (soprattutto in caso di *faccia* riconducibile a contesti divertenti o satirici).

4.2 Informazioni relative alla rete di rapporti nella comunità

In secondo luogo bisognerebbe analizzare come i commenti siano stati recepiti dalla comunità. Nelle risposte a un commento possono infatti trovarsi chiari indizi che tale commento sia stato recepito come ironico. Gli indizi da ricercare nelle risposte a un commento sarebbero:

- Espressioni che indicano una risata, soprattutto nell'incipit o isolati senza ulteriore testo.
- Parole appartenenti al campo semantico di "ironia".

Inoltre, poiché le possibilità di interpretare correttamente un enunciato ironico sono incrementate

dalla familiarità esistente tra coloro che interagiscono, il rilievo degli indizi di ironia presenti nelle risposte a un commento potrebbe essere ponderato in proporzione al numero totale di interazioni sul sito tra il commentatore originale e colui che risponde.

4.3 Informazioni di natura linguistica

Il ricorso ad alcune informazioni contestuali non escluderebbe comunque l'uso di elementi linguistici nel processo di individuazione automatica di ironia. I sistemi ad oggi implementati hanno ottenuto risultati parzialmente soddisfacenti e sarebbe errato ignorare l'aiuto che potrebbe derivarne. Gli indizi proposti da Carvalho et al. (2009) sembrano soprattutto convincenti poiché facilmente rilevabili e pensati appositamente per un identico contesto. Oltre a questi indizi, anche i puntini di sospensione e i giochi di parole si sono rivelati dei buoni segnalatori della presenza di ironia durante l'analisi empirica.

4.4 Modello integrato

Le serie di informazioni contestuali e linguistiche sarebbero integrate in un unico sistema di identificazione automatica di ironia.

Innanzitutto si isolerebbero gli utenti che esibiscono un'attitudine ironica molto spiccata all'interno della comunità. L'appartenenza a questo gruppo condizionerebbe la successiva trattazione del materiale linguistico sia interno al commento sia nelle sue risposte (v. Appendice A). Se l'utente dovesse appartenere al gruppo dei commentatori spiccatamente ironici:

- il suo commento sarà considerato molto probabilmente ironico a prescindere da indizi linguistici presenti nel commento stesso e nelle risposte.
- la presenza di indizi linguistici nel commento ne determinerebbe la classificazione come ironico. L'assenza sarebbe ininfluenza.
- la presenza di indizi linguistici nel commento ne determinerebbe la classificazione come ironico. L'assenza influirebbe solo quando, in numerose risposte, nessuna li presentasse.
- se classificati come ironici, i commenti sarebbero considerati ironici nella loro interezza.

Se invece l'utente non dovesse appartenere al gruppo dei commentatori spiccatamente ironici:

- il suo commento sarà considerato ironico solo in presenza di indizi linguistici nel commento stesso o nelle sue risposte.

- la presenza di indizi linguistici nel commento inciderebbe positivamente sulla sua classificazione come ironico. L'assenza inciderebbe negativamente.
- la presenza di indizi linguistici nelle risposte inciderebbe positivamente sulla sua classificazione come ironico. L'assenza inciderebbe negativamente.
- se i commenti fossero classificati come ironici a partire da indizi linguistici interni al commento stesso, saranno ritenute ironiche solo le frasi contenenti tali indizi, soprattutto al principio o alla fine del commento.

5 Conclusioni

In questo lavoro è stata presentata la possibilità di usare informazioni contestuali per identificare automaticamente l'ironia nei commenti dei lettori abituali di giornali online. A tal fine è stato proposto un possibile approccio computazionale che identifichi i commentatori maggiormente ironici di una comunità, suggerendo un trattamento differente del materiale linguistico tra essi e gli altri commentatori. L'integrazione di informazioni contestuali e informazioni linguistiche potrebbe influire positivamente sull'efficacia dei sistemi di riconoscimento automatico di ironia, che avrebbero una funzione importante nel campo della Sentiment Analysis.

Al momento stiamo ampliando la ricerca valutando l'influenza di informazioni come il tipo di quotidiano, l'argomento della notizia e la lunghezza del commento su un corpus di commenti più ampio e costruito su più quotidiani.

Ovviamente un'integrazione di informazioni contestuali così basilari non risolverebbe completamente il problema di come identificare automaticamente l'ironia in testi online.

Tuttavia questo lavoro riflette la ferma convinzione che questi progressivi tentativi di integrare informazioni contestuali semplici e computabili con l'informazione linguistica siano oggi la migliore strada da percorrere per tentare di affrontare automaticamente fenomeni di natura pragmatica così complessa e sfaccettata come l'ironia.

Ringraziamenti

Vorremmo ringraziare Caterina Mauri, Silvia Pareti, Diego Frassinelli e Bonnie Webber. Un ringraziamento va inoltre ad Alessandra Cervone.

Bibliografia

- Katharina Barbe. 1995. *Irony in context*. John Benjamins Publishing Co. Amsterdam.
- Lorenzo Bernardini. 2014. Un nuovo parametro per l'individuazione automatica dell'ironia: la fonte dell'enunciato. Tesi della Magistrale in Linguistica Teorica e Applicata. Università di Pavia, Pavia.
- Penelope Brown and Steven Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge University Press. Cambridge.
- Paula Carvalho, Luís Sarmiento, Mário J. Silva and Eugénio De Oliveira. 2009. Clues for detecting irony in user-generated contents: Oh...!! it's "so easy" ;-). *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*. ACM. 53–56.
- Herbert H. Clark and Richard J. Gerrig. 1984. On the pretense theory of irony. *Journal of Experimental Psychology General*. 121–126.
- Sigmund Freud. 1948. *Der Witz und seine Beziehung zum Unbewussten: Gesammelte Werke*, 6. Imago Publishing Co. London.
- Rachel Giora. 1995. On irony and negation. *Discourse processes*. 19(2): 239–265.
- Erving Goffman. 1967. *Interaction ritual: essays on face-to-face behavior*. Aldine Publishing Co. Chicago.
- Herbert P. Grice. 1978. Logic and conversation. *Syntax and semantics: Vol. 9. Pragmatics*. Academic. New York.
- Norman Knox. 1972. On the classification of ironies. *Modern Philology*. 70: 53–62.
- Rada Mihalcea and Carlo Strapparava. 2006. Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence*. 22(2): 126–142.
- Walter Nash. 1985. *The language of humour: Style and technique in comic discourse*. Longman. Londra.
- Antonio Reyes, Paolo Rosso and Davide Buscaldi. 2009. Humor in the blogosphere: First clues for a verbal humor taxonomy. *Journal of Intelligent Systems*. 18(4): 311–331.
- Antonio Reyes and Paolo Rosso. 2011. Mining subjective knowledge from customer reviews: A specific case of irony detection. *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis*. Association for Computational Linguistics. 118–124.
- Antonio Reyes, Paolo Rosso and Tony Veale. 2013. A multidimensional approach for detecting irony in Twitter. *Language Resources and Evaluation*. 47: 239–268.
- Antonio Sarmiento, Paolo Rosso and Tony Veale. 2013. Automatic creation of a reference corpus for political opinion mining in user-generated content. *Proceedings of the 1st international CIKM workshop on topic-sentiment analysis for mass opinion*. ACM. 29–36.
- Dan Sperber and Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Blackwell. Oxford.
- Wolf-Dieter Stempel. 1976. Ironie als Sprechhandlung. In Wolfgang Preisendanz and Rainer Warning. *Das Komische*. Fink. Munich. 205–237.
- Akira Utsumi. 1996. A unified theory of irony and its computational formalization. *Proceedings of the 16th conference on computational linguistics*. Association for Computational Linguistics. Morristown, NJ. 962–967.
- Tony Veale and Yanfen Hao. 2010. Detecting irony in creative comparisons. *Proceedings of 19th European conference on artificial intelligence-ECAI 2010*. IOS Press. Amsterdam. 765–770.
- Cynthia M. Whissell. 1989. The dictionary of affect in language. In Robert Plutchik and Henry Kellerman (Eds.). *Emotion: Theory, Research and Experience*. Academic Press. New York. 113–131.

Appendici

A

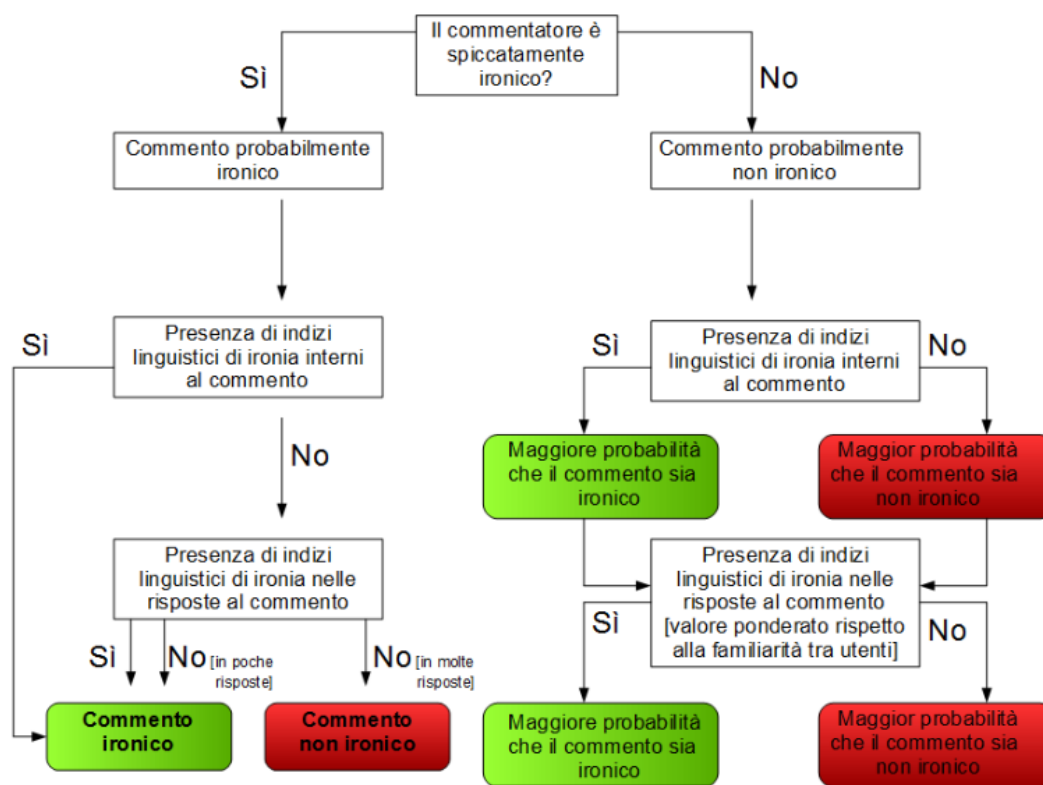


Figura 1: La figura 1 rappresenta il modello integrato di informazioni linguistiche e contestuali per il riconoscimento automatico di ironia. La classificazione dell'utente come spiccatamente ironico implicherebbe una differente trattazione del materiale linguistico contenuto nel commento stesso e nelle sue risposte. Per i commentatori ironici sarebbe sufficiente la presenza di un qualsiasi indizio nel commento o nelle sue risposte affinché esso sia classificato come interamente ironico, mentre l'unica condizione perché sia classificato come non ironico sarebbe la presenza di più risposte senza alcun indizio di ironia. Per gli altri commentatori invece la classificazione del commento come ironico dipenderebbe maggiormente dalla presenza di indizi sia nel commento che nelle sue risposte (ponderandone la rilevanza in base alla familiarità tra gli utenti). Inoltre i commenti classificati come ironici a partire da indizi presenti nel commento stesso saranno considerati ironici limitatamente alle porzioni di testo nelle quali tali indizi sono presenti.