

---

# Modelli distribuzionali delle relazioni semantiche: il caso dell'iperonimia

Giulia Benotto

Università di Pisa

I modelli di semantica distribuzionale misurano la similarità semantica tra le parole, basandosi sul principio per il quale termini che appaiono in contesti simili sono semanticamente simili. Questa linea di ricerca è stata molto utilizzata negli ultimi decenni, in cui sono state proposte molte misure per il calcolo della similarità lessicale su base distribuzionale.

Parole semanticamente simili possono differire per il tipo di relazione semantica che intercorre tra loro. La similarità distribuzionale è modellata principalmente da misure simmetriche, come il coseno, che ben si prestano al riconoscimento di relazioni semantiche simmetriche, come la sinonimia. Questo rende difficoltoso applicare tali modelli al riconoscimento di relazioni semantiche non simmetriche, quali iperonimia/iponimia (*animale, cane*) o meronimia/olonimia (*finestra, palazzo*).

Questo lavoro si propone di valutare la possibilità di identificare e classificare la relazione di iperonimia utilizzando misure di similarità direzionali. A tale scopo è stata condotta una valutazione delle misure di similarità direzionali esistenti, nel riconoscere e classificare parole legate da iperonimia. Sono state inoltre individuate alcune caratteristiche peculiari di tale relazione, modellate in due nuove misure di similarità semantica direzionale, con risultati migliori rispetto allo stato dell'arte.

## 1. Semantica Distribuzionale

Il termine “Semantica Distribuzionale” definisce una famiglia di approcci all'analisi del significato nati in linguistica computazionale e nelle scienze cognitive, basati sull'ipotesi che la distribuzione statistica delle parole nei contesti giochi un ruolo determinante nel caratterizzare il loro comportamento semantico. A partire da Harris (1968) l'informazione distribuzionale ha avuto un ruolo fondamentale per la modellazione di vari

aspetti del linguaggio, ruolo riconosciuto anche per lo sviluppo di modelli computazionali del lessico. L'ipotesi distribuzionale è erede di una tradizione associazionista e combinatoria che assume come chiave fondamentale per l'esplorazione dello spazio semantico di una parola la ricostruzione delle modalità con le quali essa si combina con altre parole. Il lessico è concepito come uno spazio metrico nel quale le parole sono separate da distanze che dipendono dal grado di similarità semantica, misurata attraverso la distribuzione statistica di co-occorrenza nei testi. Il principio alla base è l'ipotesi distribuzionale, secondo cui due parole sono tanto più semanticamente simili quanto più tendono a ricorrere in contesti simili (Miller & Charles 1991). I modelli di similarità distribuzionale rappresentano il significato delle parole e misurano la similarità distribuzionale tra le parole basandosi sull'ipotesi che la prossimità nello spazio modelli la correlazione semantica.

Modelli di questo tipo trovano caratterizzazione fondamentale nelle parole del linguista inglese Firth (1957): "you shall know a word by the company it keeps". Sul piano cognitivo questo risponde a un modello del lessico mentale in cui i significati hanno rappresentazioni contestuali, come descritto da Charles (2000): "an abstraction of information in the set of natural linguistic context in which a word occurs".

Data l'ipotesi distribuzionale, è possibile applicare ai testi metodi computazionali per acquisire le proprietà semantiche attraverso elaborazioni matematiche delle distribuzioni delle parole nei testi. L'uso dei modelli informativi di tipo distribuzionale allo scopo di rappresentare il significato lessicale è un'area di ricerca che ha guadagnato importanza nell'ultima decade, grazie alla disponibilità di corpora testuali di grandi dimensioni e allo sviluppo di tecniche statistiche per l'estrazione degli schemi distribuzionali dei lessemi. Ciò ha permesso l'implementazione di modelli computazionali per la costruzione di spazi lessicali da applicare alla simulazione di diversi aspetti della competenza semantica. Le diverse realizzazioni dell'ipotesi distribuzionale sono accomunate dall'assunto che quantificare la similarità semantica tra due parole implichi valutare la misura in cui si sovrappongono i contesti linguistici nelle quali esse ricorrono. I modelli differiscono per vari parametri legati alle finalità teoriche o applicative di riferimento.

La nozione di spazio semantico si basa su un'analogia con lo spazio geometrico per cui il contenuto semantico di una parola è rappresentato dalla sua posizione in uno spazio definito in un sistema di coordinate, determinato dai contesti linguistici in cui tale parola può ricorrere. Uno spazio semantico di parole è definito dalla quadrupla di valori (T, B, M, S).

T è l'insieme delle parole che formano gli elementi che popolano lo spazio e di cui fornisce la rappresentazione semantica. B è la base che definisce le dimensioni dello spazio e contiene i contesti linguistici rispetto ai quali si valuta la similarità distribuzionale tra le parole. M è una matrice di co-occorrenza che fornisce una rappresentazione vettoriale di ogni parola in T. Il grado di distanza che intercorre tra vettori che rappresentano due parole in uno spazio è valutato misurando il coseno dell'angolo che essi formano. Due vettori hanno massima similarità quando sono allineati sulla stessa linea nella stessa direzione. In questo caso l'angolo ha ampiezza 0 e il coseno è 1, se i due vettori sono ortogonali (le parole che rappresentano non hanno caratteristiche comuni), l'angolo che si forma tra loro misura 90 e il coseno è 0. Alla base dei modelli distribuzionali risiede la possibilità di applicare ai testi metodi computazionali per acquisire le proprietà semantiche attraverso elaborazioni matematiche delle distribuzioni delle parole nei testi stessi. E' quindi possibile modellare relazioni di iperonimia utilizzando la semantica distribuzionale?

## 2. Modelli di similarità semantica utilizzati nella modellazione della relazione di iperonimia

I modelli distribuzionali caratterizzano fenomeni di similarità semantica tra parole vicine nello spazio distribuzionale, ma non rappresentano le relazioni semantiche che intercorrono tra parole (*cane* può essere simile ad *animale* e *gatto*, ma con relazioni diverse). Le parole sono collocate in spazi semantici comuni, in base alle loro rappresentazioni contestuali; la distanza tra parole è misurata tenendo conto della similarità semantica. Tale distanza è una relazione simmetrica: se una parola, *A*, è vicina a *B* nello spazio semantico, questo implica che *B* sia vicina ad *A*. Il modello non riesce a caratterizzare le diverse proprietà semantiche delle relazioni che legano le parole vicine. La distanza tra le parole è poco funzionale nel descrivere relazioni di tipo asimmetrico, come l'iperonimia. Considerando, ad esempio, la coppia *animale-cane*, legata dalla relazione di iperonimia, si può assumere che, se l'essere *cane* implica l'essere *animale*, l'essere *animale* non implica l'essere *cane*, essendo *animale* un termine più ampio di *cane*. Gli iperonimi sono termini semanticamente più ampi rispetto ai loro iponimi, sia a livello estensionale, (*animale* si riferisce a un insieme più ampio di entità rispetto a *cane*), sia a livello intensionale (*cane* ha proprietà più informative rispetto a quelle di *animale*, si può assumere infatti che il cane abbaia,

cosa non vera per tutti gli animali), quindi i sopraordinati risultano meno informativi rispetto ai concetti di livello base. Supponendo i concetti come organizzati in una gerarchia di categorie, che va dall'estremamente generico all'estremamente specifico, possiamo assumere che la classe più generica comprenda un alto numero di elementi, massimizzando l'accuratezza della classificazione. Le categorie più specifiche, invece, permettono maggiore accuratezza nella categorizzazione dei membri: sapere che qualcosa è un *canè* ci permette di inferire una serie di proprietà di quell'elemento. Di tutte le possibili categorie in una gerarchia, il livello base è il livello intermedio, compromesso tra l'accuratezza della classificazione data dal livello più generico e il potere predittivo del livello più specifico (Murphy 2003). Le categorie di base rappresentano il livello di caratterizzazione più naturale (Murphy & Lassaline 1997), dato che sono quelle prodotte più spesso, riconosciute più velocemente e apprese per prime.

Recentemente sono stati proposti diversi modelli per la rappresentazione di relazioni asimmetriche (Weeds *et al.* 2004, Clarke 2009), basati sull'ipotesi di inclusione distribuzionale, utilizzata anche da Kotlerman *et al.* (2010) per l'identificazione del *Lexical Entailment* (implicazione lessicale), una relazione concettualmente più vasta rispetto a quella di iperonimia, che si applica a parole che possono essere inferite da altre parole in modo tale che sia possibile sostituire una parola con l'altra in determinati contesti. L'inferenza può essere determinata dalla sinonimia, dall'iperonimia e dalla meronimia. L'ipotesi di inclusione distribuzionale si basa sull'assunto per cui se  $u$  è un termine semanticamente più limitato rispetto a  $v$ , un numero significativo di features distribuzionali salienti di  $u$  saranno incluse nelle features di  $v$ .

È possibile utilizzare misure di inclusione distribuzionale per la classificazione dell'iperonimia? Le misure allo stato dell'arte sono funzionali per il riconoscimento di questa relazione? È possibile implementare delle variazioni di queste misure atte a migliorare il riconoscimento della relazione di iperonimia?

### 3. Misure di similarità distribuzionali

Per indagare la capacità delle misure di similarità direzionali nel classificare la relazione di iperonimia, abbiamo valutato il grado di correttezza sia nell'identificare l'iperonimo di un nome target, che nel discriminare termini legati dall'iperonimia da termini legati da relazioni simmetriche, come

i co-ponimi. I termini lessicali sono stati rappresentati come vettori di features distribuzionali estratte dal tensore TypeDm, derivato dal framework Distributional Memory (DM) (Baroni, Lenci 2010). DM rappresenta i dati distribuzionali come un insieme di tuple pesate  $((p1, l, p2), w)$  dove  $p1$  e  $p2$  sono parole,  $l$  è la dipendenza sintattica che le lega e  $w$  è il peso che stima la salienza di quella tupla utilizzando la Local Mutual Information (LMI) calcolata sulla frequenza dei legami. La Mutua Informazione (MI) è il rapporto tra la probabilità di occorrenza di due eventi congiunti (calcolata stimandone la frequenza) e la probabilità teorica del verificarsi degli eventi in maniera indipendente. Più alto è il valore della MI, più è plausibile che due eventi si verificino in maniera congiunta.

$$\text{Mutua Informazione} = \frac{f(e_1, e_2)}{f(e_1)f(e_2)}$$

La LMI è una variante di MI utilizzata per ridurre l'impatto dei dati poco frequenti. TypeDm contiene 30.693 lemmi e 25.336 link diretti e inversi formati da dipendenze sintattiche.

Le misure sono state valutate su un *corpus* derivato dal dataset BLESS (Baroni, Lenci 2011), che consiste di triplette che esprimono una relazione tra concetti target e concetti relatum. BLESS è composto da 200 concetti target, suddivisi in 17 classi (quali BIRD, FRUIT etc.). Per ogni concetto target BLESS include più parole, collegate al concetto target da una relazione semantica. Per questo esperimento è stato usato un sotto corpus derivato da BLESS formato da 14547 tuple, in modo che il termine relatum fosse attestato in TypeDM e le relazioni fossero le seguenti:

COORD: il relatum è un co-ponimo del concetto (*cocodrillo-lucertola*)

HYPHER: il relatum è un nome che è iperonimo del concetto (*cocodrillo-animale*)

MERO: il relatum è un nome che si riferisce a una parte del concetto (*cocodrillo-bocca*)

RANDOM-N: il relatum è un nome casuale che non ha relazioni con il concetto target (*cocodrillo-scarpa*).

Ogni parola del test set è rappresentata come un insieme di features distribuzionali. Per ogni feature esiste, nel tensore di TypeDM, una tupla corrispondente. Il peso della feature corrisponde al peso associato alla corrispondente tupla in TypeDM. Sono state applicate ai vettori derivati

dalle tuple tre misure direzionali allo stato dell'arte, per valutarne l'abilità nel discriminare l'iperonimia dalle altre relazioni semantiche, in particolare la co-iponimia (una misura simmetrica). In base all'ipotesi di fondo, le misure di similarità distribuzionale dovrebbero attribuire un punteggio più alto alle relazioni di iperonimia, asimmetriche e direzionali. Le misure allo stato dell'arte che sono state testate sono:

*WeedsPrec*( $u, v$ ): quantifica l'inclusione pesata delle features di un termine  $u$  all'interno delle features di un termine  $v$ .

$$\textit{WeedsPrec}(u, v) = \frac{\sum_{f \in F_u \cap F_v} w_u(f)}{\sum_{f \in F_u} w_u(f)}$$

*cosWeeds*( $u, v$ ): corrisponde alla media geometrica tra *WeedsPrec* e la similarità simmetrica tra  $u$  e  $v$ , misurata dal coseno dei loro vettori.

$$\textit{cosWeeds}(u, v) = \sqrt{\textit{WeedsPrec}(u, v) * \textit{cos}(u, v)}$$

*ClarkeDE*: misura simile a *WeedsPrec*, proposta da Clarke nel 2009.

$$\textit{ClarkeDE}(u, v) = \frac{\sum_{f \in F_u \cap F_v} \min(w_u(f), w_v(f))}{\sum_{f \in F_u} w_u(f)}$$

Sono state adottate due diverse metodologie di valutazione. La prima si basa sulla metodologia descritta da Baroni e Lenci (2010). Per ogni concetto, dati i punteggi di similarità del concetto con tutti i relata (per tutte le relazioni) nel test set, si seleziona il relatum con il punteggio più alto. Per ognuno dei 200 concetti di BLESS si ottengono 4 punteggi di similarità, uno per relazione. Si produce un boxplot, che riassume le distribuzioni dei punteggi per le relazioni nei 200 concetti. I boxplot (fig. 1), mostrano la mediana della distribuzione come una linea orizzontale in un box che si estende dal primo al terzo quartile. I valori al di fuori dell'intervallo sono rappresentati con dei cerchi.

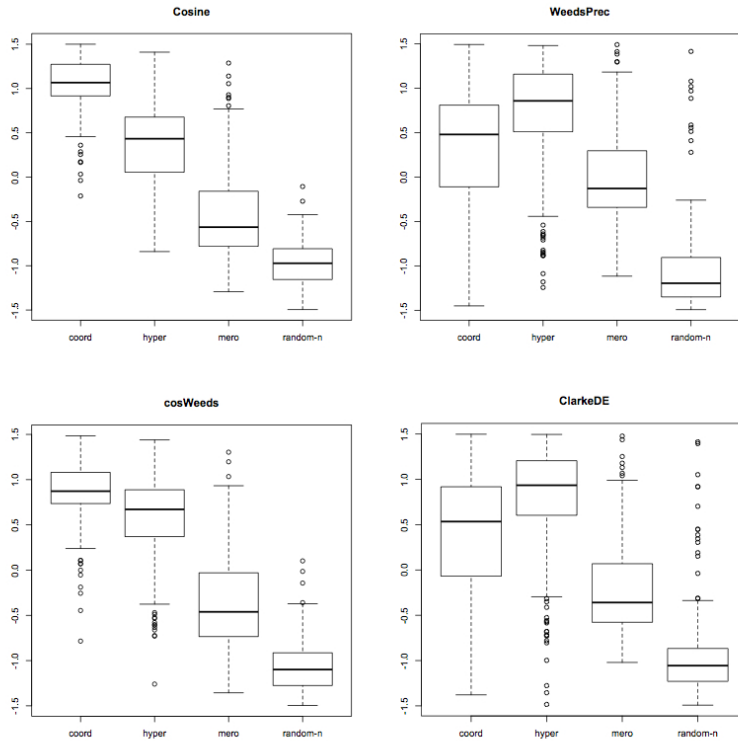


Figura 1. Boxplot per le misure di similarità distribuzionale allo stato dell'arte.

Dai risultati si può osservare che tutte le misure riescono a discriminare coppie che hanno un legame semantico da quelle che non lo hanno. Il coseno (usato come baseline) mostra una forte capacità nell'identificare i coordinati tra i vicini del target. Questo era prevedibile, essendo il coseno misura simmetrica, come la relazione di co-iponimia. Le misure direzionali favoriscono significativamente gli iperonimi, rispetto ai coordinati. L'unica eccezione è rappresentata da *cosWeeds*, che privilegia i coordinati se pur con poca differenza rispetto agli iperonimi, a causa della sua componente simmetrica. Dall'analisi dei boxplot si può dedurre che le misure direzionali basate sull'ipotesi distribuzionale di inclusione migliorino il riconoscimento e la classificazione degli iperonimi rispetto ai co-iponimi.

Il secondo tipo di valutazione effettuata si basa sull'Average Precision (AP), un metodo utilizzato in Information Retrieval che combina

precisione, rango di occorrenza e recall. Per ogni misura di similarità è stata calcolata AP rispetto alle 4 relazioni di BLESS. Il massimo punteggio possibile di AP, 1, si ottiene quando tutti i relata legati al target dalla relazione in esame riportano punteggi di similarità semantica più alti rispetto ai relata non legati al target da quella particolare relazione (se in un ideale ordinamento per punteggio di similarità semantica, i relata legati al target dalla relazione in esame occorressero tutti ai primi posti). Anche i risultati ottenuti con questa metodologia di valutazione confermano il fatto che le misure di similarità direzionali hanno performance migliori rispetto al coseno nel riconoscere e classificare gli iperonimi.

<i>Misura</i>	COORD	HYPHER	MERO	RANDOM-N
<i>Cosine</i>	0.79	0.23	0.21	0.30
<i>WeedsPrec</i>	0.45	0.40	0.31	0.32
<i>cosWeeds</i>	0.69	0.29	0.23	0.30
<i>ClarkeDE</i>	0.45	0.39	0.28	0.33

Tabella 1. Valori di Average Precision per le misure direzionali allo stato dell'arte.

Confrontando i risultati ottenuti nel valutare le relazioni di co-ipponimia e quelle di iponimia si nota che tutte le misure analizzate finora hanno riportato valori di AP più alti per la relazione di co-ipponimia rispetto a quella di iponimia, probabilmente perché queste misure di similarità sono state concepite per il riconoscimento del lexical entailment. Questo fa sì che le features di inclusione che usano siano soddisfatte anche dai coordinati.

Utilizzare misure di similarità direzionali è una strada promettente per la classificazione dell'ipponimia, ma l'analisi delle misure allo stato dell'arte rivela che è possibile migliorarle perché siano performanti per il riconoscimento di tale relazione. A tale scopo sono state implementate due variazioni delle misure sopra descritte che caratterizzano le proprietà della relazione di iponimia.

## 4. Misure di similarità direzionale

La prima misura implementata è basata sull'idea che un termine più ampio dovrebbe poter essere trovato anche in contesti nei quali il termine meno ampio non è utilizzato. Se  $v$  è un termine semanticamente più ampio di  $u$ , le features distribuzionali di  $u$  sono anche features distribuzionali



di  $v$ , al contrario ci sono features di  $v$  che non sono features di  $u$ . Considerando la coppia di termini *animale-cane*, si può assumere che le features distribuzionali di *animale* comprendano le features di *cane* (tra le features di *animale* si potranno trovare proprietà quali “abbaiare”, caratterizzanti del *cane*), ma che le features di *cane* non comprendano molte delle features di *animale* (fra le features di *animale* vi sarà “miagolare”, che non è una feature di *cane*).

$$InvCL(u, v) = \sqrt{ClarkeDE(u, v) * (1 - ClarkeDE(v, u))}$$

Questa misura valuta non solo l’inclusione delle features del termine iponimo in quelle dell’iperonimo, ma anche la non-inclusione delle features dell’iperonimo in quelle dell’iponimo.

La seconda misura è basata sull’intuizione che i termini subordinati si applicano a un insieme di termini che appartengono alla stessa categoria. Si assume, in altri termini, che gli iponimi di un iperonimo comune formino un insieme di termini contrastivi in relazione all’iperonimo. Si può dire che, considerando gli iponimi del termine *animale*, termini come *cane*, *gatto*, *cavallo* formano l’insieme dei termini di livello più basso rispetto ad *animale* e sono in contrasto tra loro avendo proprietà diverse e quindi features distribuzionali diverse. Se un termine  $v$  è semanticamente più ampio di un termine  $u$  le features di  $u$  sono incluse nelle features di  $v$  (le features di *cane* sono presenti anche tra le features di *animale*), è pur vero che anche le features degli altri iponimi complementari a  $u$  rispetto a  $v$  saranno incluse nelle features di  $v$  (considerando *gatto* come complementare di *cane* rispetto all’iperonimo *animale*, si può assumere che ci siano anche molte features di *gatto* che compaiono tra quelle di *animale*).

La misura è così strutturata:

$$COL(u, v) = \sqrt{ClarkeDE(u, v) * ClarkeDE(comp(u, v))}$$

In questa formula,  $y$  rappresenta il termine complementare di  $u$  rispetto a  $v$ . Si è assunto che fosse possibile approssimare le proprietà complementari di un iponimo individuandone il co-iponimo più vicino in termini distribuzionali e selezionando tutte le features di tale termine che non sono anche features dell’iponimo in esame. Il termine più vicino all’iponimo è stato calcolato utilizzando il coseno (essendo il coseno la misura che

meglio individua similarità simmetriche, come quella di co-iponimia) che ha permesso di individuare, per ogni iponimo il co-iponimo più simile. Si suppone quindi che l'iponimo in esame e il co-iponimo più simile a esso rispetto a  $\nu$  condividano molte features. Selezionare solo le features del co-iponimo che non sono caratterizzanti dell'iponimo in esame dovrebbe dare una buona approssimazione di come è fatto il complementare di tale termine.

Per la valutazione di tali misure abbiamo usato la metodologia descritta in precedenza. I risultati hanno confermato le intuizioni e le aspettative alla base di queste misure.

La valutazione con i boxplot (fig. 2) evidenzia l'andamento già visto nell'analisi delle misure di similarità direzionali. *invCL* e *COL* favoriscono in maniera significativa gli iperonimi rispetto ai coordinati. In particolare, analizzando *invCL* si può notare che la capacità di classificare l'iperonimia migliora, ma in maniera poco significativa rispetto a *ClarkeDE*, mentre aumenta significativamente la capacità di discriminare questa relazione dalla relazione di coordinazione, per la quale le performance diminuiscono. Lo stesso avviene in relazione alla misura *COL*, per la quale aumenta in maniera significativa la capacità di classificare la relazione di iperonimia, contestualmente a una diminuzione in relazione ai coordinati, che fa sì che questa misura si comporti meglio nel classificare la relazione di iperonimia e nel discriminarla dalla co-iponimia.

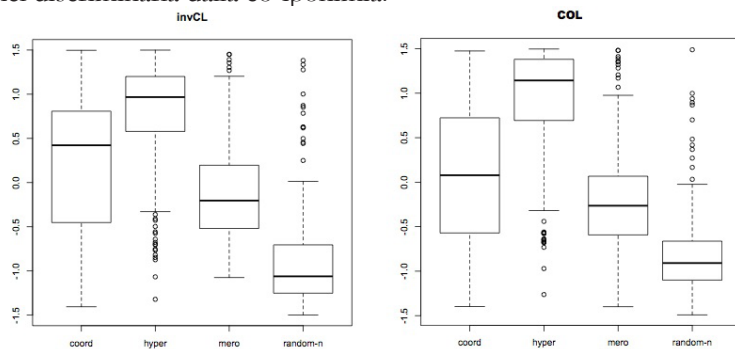


Figura 2. Boxplot per le misure di similarità sviluppate per la classificazione della relazione di iperonimia.

Per quanto riguarda la valutazione con Average Precision (tabella 2), le due misure qui implementate riportano risultati significativamente mi-

gliori rispetto alle misure allo stato dell'arte.

<i>Misura</i>	COORD	HYPER	MERO	RANDOM-N
invCL	0.38	0.40	0.31	0.34
COL	0.36	0.42	0.32	0.3

Tabella 2. Valori di average Precisione per le misure direzionali sviluppate per il riconoscimento dell'iperonimia.

Queste presentano valori alti per la relazione di iperonimia, ma valori ancora più alti per la relazione di co-iponimia. Le nuove misure non solo migliorano il punteggio di AP in relazione alla iperonimia, ma riescono anche a discriminare la relazione di iperonimia da quella di co-iponimia. Sia *invCL* che *COL*, infatti, riportano punteggi di AP più alti rispetto alle misure allo stato dell'arte in relazione all'iperonimia, ma soprattutto riportano punteggi significativamente più bassi per la relazione di co-iponimia, anche rispetto a quelli ottenuti per la relazione di iponimia, cosa che non si verificava utilizzando le misure direzionali allo stato dell'arte. Tra le due misure qui implementate, *COL* ha una performance migliore dato che aumenta maggiormente il divario tra il punteggio riportato per la relazione di co-iponimia e quella di iponimia.

## 5. Conclusioni

L'intento di questo lavoro era l'analisi della capacità delle misure di similarità direzionali nella classificazione della relazione di iperonimia. La componente direzionale di tali misure appariva una caratteristica ottimale per il riconoscimento della relazione di iperonimia, essendo questa una relazione semantica asimmetrica. La valutazione delle misure allo stato dell'arte ha parzialmente confermato questa intuizione, mettendo in risalto la capacità delle misure direzionali di classificare le relazioni di iperonimia, ma evidenziando anche una limitata capacità di discriminare la relazione di iperonimia rispetto a quella di co-iponimia.

Le misure di similarità semantica direzionali fin qui analizzate non erano state sviluppate con lo scopo di classificare le relazioni di iperonimia, ma per il *lexical entailment* (o implicazione lessicale), un tipo di relazione più ampio rispetto all'iperonimia. Sono state quindi sviluppate delle misure di similarità semantica direzionali costruite in modo da formalizzare aspetti prototipici della relazione di iperonimia, per migliorare il riconoscimento di quest'ultima.

Le misure implementate, *invCL* e *COL*, sono state costruite intorno alle proprietà peculiari dell'iperonimia. Gli esperimenti confermano la capacità di queste due misure di riconoscere le relazioni di iperonimia, riportano infatti migliori capacità di classificare questa relazione e anche di discriminarla rispetto alla relazione di co-iperonimia, rispetto alle misure allo stato dell'arte.

Queste due misure confermano l'ipotesi iniziale e mostrano che le misure di similarità semantica direzionali rappresentano un buon punto di partenza per il riconoscimento e la classificazione della relazione di iperonimia in maniera non-supervisionata, fornendo un promettente punto di partenza per sviluppi futuri.

## Bibliografia

- Baroni M., Lenci A. (2010) Distributional Memory: A general framework for corpus-based semantics, *Computational Linguistics*, 36(4), pp. 673–721.
- Baroni M., Lenci A. (2011) How we BLESSed distributional semantic evaluation, in *Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics*, pp. 1-10.
- Charles W. G. (2000) Contextual correlates of meaning, *Applied Psycholinguistics*, 21.4 (2000), pp. 505–524.
- Clarke D. (2009) Context-theoretic semantics for natural language: an overview, in *Proceedings of the EACL 2009 Workshop on GEMS: GEometrical Models of Natural Language Semantics*, pp. 112–119.
- Firth J. R. (1957) *Papers in Linguistics*, London, Oxford University Press.
- Geffet M., Dagan I. (2005) The distributional inclusion hypotheses and lexical entailment, in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*.
- Harris Z. S. (1968) *Mathematical Structures of Language*, New York, Wiley.
- Kotlerman L., Dagan I., Szpektor I., Zhitomirsky-Geffet M. (2010) Directional distributional similarity for lexical inference, *Natural Language Engineering*, 16(04), pp. 359–389.
- Miller G. A., Charles W. G. (1991) Contextual correlates of semantic simi-

- larity, *Language and Cognitive Processes*, 6.1(1991), pp. 1–28.
- Murphy, G. L. (2004) *The big book of concepts*, Cambridge MA, The MIT Press.
- Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical structure in concepts and the basic level of categorization, in Lamberts, Koen, Shanks, David R. (eds.) (1997) *Knowledge, concepts and categories. Studies in cognition*, Cambridge MA, The MIT Press, pp. 93-131.
- Murphy, M. L. (2003) *Semantic Relations and the lexicon: Antonymy, Synonymy and Other Paradigms*, Cambridge, Cambridge University Press.
- Rogers T.T., McClelland J.L. (2003) *Semantic cognition: A parallel distributed processing approach*, Cambridge MA, The MIT Press.
- Weeds J., Weir D., McCarthy D. (2004) Characterising measures of lexical distributional similarity, in *Proceedings of COLING 2004*, pp. 1015–1021.