



UNIVERSITÀ DEGLI STUDI DI MILANO  
FACOLTÀ DI STUDI UMANISTICI

TECNICHE DI WORD EMBEDDING PER LO STUDIO  
DIACRONICO DEL SIGNIFICATO: GLI ARTICOLI DI REPUBBLICA  
DAL 1990 AL 2021

Relatore: Prof. Alfio Ferrara

Correlatore: Prof. Andrea Scala

Tesi di Laurea di:

Vittorio Mussin

Matr. Nr. 927230

ANNO ACCADEMICO 2020/2021

## **Ringraziamenti**

Ringrazio il Prof. Alfio Ferrara per avermi dato l'opportunità di approfondire gli argomenti di questa materia e per essere stato fonte di consigli e supporto durante i mesi che hanno portato allo sviluppo di questo lavoro.

Ringrazio il Prof. Andrea Scala per la disponibilità, gli utili accorgimenti e correzioni, e per avermi fornito le conoscenze necessarie per affrontare l'argomento dal punto di vista linguistico.

Ringrazio la mia famiglia e, in particolar modo, i miei genitori, che mi hanno supportato nella scelta degli studi, e che hanno la forza di supportarmi costantemente, oltre ad avere dato il loro contributo anche in questo lavoro, per l'aiuto negli argomenti statistici e per l'idea di visualizzare il cambiamento delle distanze di similarità in un piano cartesiano.

Ringrazio infine gli amici che mi sono stati vicini durante il percorso universitario, e non, per il loro sostegno e i momenti di sollievo.

## Indice

<b>Ringraziamenti</b> .....	i
<b>Indice</b> .....	ii
<b>Introduzione</b> .....	iii
<b>1: Individuare il problema: cos'è il mutamento semantico?</b> .....	1
1.1 Il mutamento semantico .....	2
1.2 Cause del mutamento .....	6
<b>2: Semantic shift nel Word embedding</b> .....	9
2.1 Semantica distribuzionale e contesto .....	9
2.2 Language models e principali modelli di Word Embedding .....	12
2.3 Diachronic embedding e Semantic shift .....	17
<b>3: Un caso di studio sugli articoli di Repubblica dal 1990 al 2021</b> .....	21
3.1 Confronto e descrizione dei corpora utilizzati .....	21
3.2 Training, allineamento e calcolo delle distanze .....	21
3.3 Valutazione dei modelli .....	22
3.4 Individuare le parole mutate semanticamente .....	24
3.5 Rappresentare il mutamento semantico .....	27
3.6 Una semantica diacronica distribuzionale .....	30
<b>Conclusioni</b> .....	35
<b>Bibliografia</b> .....	38

## Introduzione

Con l'avvento dell'informatica nell'analisi testuale è diventato possibile "informatizzare" persino il significato delle parole, pensando il linguaggio come un sistema combinatorio, in cui le frasi sono costituite da simboli che non hanno un significato di per sé, ma che assumono di senso solo quando rapportati alla loro probabilità di apparire in sequenza con altri simboli. In questo modo è possibile quantificare le relazioni che intercorrono tra le parole, considerate relazioni tra significati dalla semantica distribuzionale. Per limiti di questo lavoro non ci addentreremo nella questione se un approccio di questo tipo sia fondato dal punto di vista linguistico, ma ci limiteremo a verificare modelli informatici di questo tipo nel compito di individuare il mutamento di significato delle parole. Uno studio diacronico in una tale concezione del significato, infatti, fa corrispondere il mutamento semantico al cambiamento delle relazioni tra parole.

Perciò, utilizzando strumenti informatici di *word embedding*, si cercherà di capire quanto bene una semantica distribuzionale riesca a codificare il cambiamento di significato, avvalendoci delle metodologie correnti per la sua individuazione. In più, tenendo conto delle problematiche linguistiche e computazionali, si tenterà di comprendere che cosa questa semantica codifichi come mutamento e come questo si rapporti alle altre teorie sul cambiamento di significato.

Dapprima, infatti, si cercherà di comprendere come questo fenomeno sia stato concepito in linguistica in maniere diverse, per poi passare a una rassegna sommaria delle cause che possono portare al mutamento semantico, insieme ai fattori linguistici che lo garantiscono. In seguito, verranno introdotte le metodologie di costruzione e di confronto di modelli di *word embedding*, con attenzione ai principi e problemi teorici legati a una codifica distribuzionale del significato. Infine, questi modelli saranno verificati su una serie di articoli raccolti dall'archivio digitale di Repubblica, per un periodo che va indicativamente dal 1990 al 2021. Dalla loro valutazione si è portati a ritenere che essi siano in grado di individuare il mutamento dell'estensione dei significati, piuttosto che il mutamento dei significati. In una situazione non priva di problematiche, rappresentate sia da fattori strutturali, linguistici quanto computazionali, che contingenti, rispetto alla bontà dei modelli che abbiamo costruito. Tuttavia, le metodologie di confronto tra significati si

sono dimostrate efficaci quando questi venivano codificati correttamente, il che lascia sperare in risultati promettenti per lavori basati su strumenti migliori.

## 1: Individuare il problema: cos'è il mutamento semantico?

Uno dei casi più famosi e controversi sullo studio del mutamento semantico è rappresentato dall'indagine condotta nel 1931 da Jost Trier, il quale notò come nella lingua tedesca, tra Dodicesimo e Tredicesimo secolo, fossero avvenuti dei cambiamenti di significato nei termini legati alla sfera della conoscenza<sup>1</sup>. Termini che, a distanza di un secolo, mutarono semanticamente a seguito dei cambiamenti culturali e sociali verificatisi con il tramonto del feudalesimo. Per cui, la stessa parola che prima indicava «le conoscenze ed abilità dei cavalieri di corte»<sup>2</sup> passò ad indicare i saperi della scienza e dell'arte, per la necessità di riadattare il suo significato alla nuova realtà di riferimento. Questo studio fu più tardi contestato sia per la bontà delle fonti utilizzate, «in netta prevalenza testi del mistico tedesco Meister Eckhart»<sup>3</sup>, sia per la verificabilità dei risultati, dal momento che lavorare sul lessico astratto in fasi così antiche avrebbe consentito a Trier un certo grado di arbitrarietà<sup>4</sup>. Egli si rifaceva a una concezione secondo la quale «il significato delle parole non può essere analizzato considerandole singolarmente, ma solo guardando alla struttura in cui si inseriscono»<sup>5</sup>. Struttura rappresentata da tutte quelle relazioni che una parola intrattiene con le parole che esprimono idee vicine; per cui, ad esempio, il significato di *tecnica* sarebbe dato da tutte le relazioni che questa parola intrattiene con *sapere*, *conoscenza*, *arte*, *abilità*, e simili. Anche ipotizzando la correttezza dell'indagine di Trier, una riorganizzazione dei termini di una data area concettuale rimane tuttavia solo una delle molteplici facce in cui si può osservare il fenomeno del mutamento semantico. Infatti, per quanto rimanga costante considerarlo come una variazione nel tempo che concerne in qualche modo il significato<sup>6</sup>, in linguistica il mutamento semantico è inteso talora in maniera diversa.

---

<sup>1</sup> Come i termini per indicare «le conoscenze ed abilità dei cavalieri di corte», «le abilità artigiane di chi non apparteneva alla nobiltà» e «il sapere del nobile [...] vicino al sapientia latino», in alto-tedesco medio rispettivamente, *kunst*, *list* e *wisheit*. Cfr. Petricca P., *Semantica. Forme, Modelli e Problemi*, Milano, LED Edizioni Universitarie, 2019, pp. 29-33.

<sup>2</sup> Ibidem.

<sup>3</sup> Petricca P., *Semantica. Forme, Modelli e Problemi*, cit., pp. 29-33., nota 16.

<sup>4</sup> Casadei F., *Lessico e Semantica*, Roma, Carocci, 2003, p. 62.

<sup>5</sup> Ivi, p. 61.

<sup>6</sup> Mentre, se il significato cambia in sincronia (cioè a parità di coordinate temporali), per esempio nel modo in cui certe parole vengono usate da una particolare comunità di parlanti piuttosto che da un'altra, si parla allora di *variazione* di significato e non di mutamento, che implica invece una dimensione temporale. Cfr. Labov W., *Il continuo e il discreto nel linguaggio*, Bologna: Il Mulino, 1977.

### 1.1 Il mutamento semantico

Il significato delle parole cambia col passare del tempo e, se a una parola possono venire associati nuovi significati, può anche accadere che possa perdere il suo significato originario. Non c'è un solo modo in cui una parola può cambiare di significato e la casistica spesso non è nemmeno troppo trasparente, andandosi a concatenare vari tipi di fattori differenti. Ad esempio, una parola può mutare perché muta la cosa a cui fa riferimento, come nel caso di *galera*, che indica il luogo dove chi commette reato sconta la sua pena, un tempo consistente in lavori forzati a bordo di una particolare imbarcazione a remi, la “galera” appunto. Oppure, perché si usa una parola che si associa in qualche modo a ciò che si vuole esprimere, quando si parla cioè di “piedi” della montagna o di “costruire” il proprio futuro, o si usa il toponimo *Gorgonzola* per riferirsi al formaggio. Può accadere anche che una parola venga usata con un certo significato perché associata secondariamente a un'altra parola di forma simile, cui quel significato è tipico; com'è successo in francese, dove «*gazouiller* “cinguettare, mormorare” significa anche “emanare un cattivo odore” per associazione con *gaz*»<sup>7</sup>. Inoltre, una parola può vedere il suo significato espandersi, come nel caso odierno di *ripartenza*, termine coniato in ambito calcistico come sinonimo di contropiede e successivamente esteso a indicare più generalmente una “ripresa”; oppure restringersi, come nel caso di *movida*, che dal descrivere una situazione o ambiente piacevolmente movimentati, nel tempo si è specializzato a designare solamente un'intensa vita notturna tipica dei grandi centri urbani. O ancora, una parola col passare del tempo può finire per essere connotata positivamente o negativamente, come nel caso di *baiser* francese che «venne inizialmente usato come eufemismo con connotazioni oscene e oggi, a sua volta, è sostituito da *embrasser* “abbracciare”»<sup>8</sup>. Inoltre, i casi elencati finora possono presentarsi interferendo gli uni con gli altri, finendo per creare situazioni miste.

In tutte queste situazioni, tuttavia, va notato che a cambiare è il rapporto tra significanti e significati, ovvero la relazione tra *forma* (o *espressione*) e *contenuto* delle parole. Se il significato è infatti il contenuto di una forma, questo si può immaginare come una sostanza all'interno di un contenitore. Quando accade, per diversi motivi, che cambi

---

<sup>7</sup> Lazzeroni R., *Il mutamento linguistico*, in Lazzeroni R. (a cura di), *Linguistica storica*, Carocci editore, 2014, p. 27.

<sup>8</sup> Ivi, p. 25.

la qualità della sostanza, o che cambi la forma del contenitore, si avrà allora un mutamento nella relazione tra contenitore e contenuto. Questa è la concezione a cui fa riferimento la Semantica diacronica tradizionale, che concepisce il mutamento semantico come un fenomeno, relegato alla specifica parola, che avviene «ogni volta che un nuovo nome verrà collegato a un dato senso e/o un nuovo senso verrà collegato a un dato nome»<sup>9</sup>.

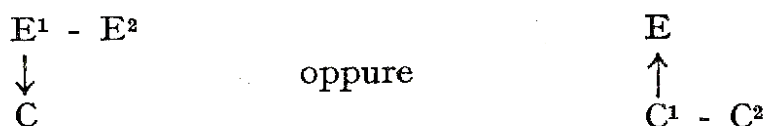


Figura 1, Schematizzazione delle associazioni tra parole: a sinistra si ha associazione di espressione, a destra di contenuto, dove E sta per "espressione" e C per "contenuto". Cfr. Coseriu E., *Teoria del linguaggio e linguistica generale*, Bari, Laterza, 1971, p. 254.

A un livello di analisi diverso si possono studiare non le relazioni tra forme e contenuti, ma tra contenuti e contenuti ( $C_1 - C_2$  in Figura 1), considerando come mutamento semantico solamente la riorganizzazione della struttura dei contenuti lessicali. In latino, per esempio, esistevano due termini opposti per indicare il colore nero: *ater* e *niger*, per riferirsi rispettivamente, al nero non brillante e al nero brillante, cioè lucido; nelle lingue romanze, invece, questa distinzione si è persa<sup>10</sup>. Indicare il colore nero con *ater* o *niger* piuttosto che solo con *nero* riguarda in generale un fatto d'organizzazione della realtà mediante unità lessicali, e più in particolare comporta che queste unità codifichino significati diversi. Se infatti si ipotizza che il significato sia composto da elementi più piccoli, come porzioni di significato, avere uno stesso contenuto suddiviso in due contenitori prima, e in uno solo poi, comporterà un cambiamento nella gestione di queste parti. Quando accade che in una lingua cambi la gestione delle porzioni di significato allora avverrà mutamento semantico e accadrà che i contenitori codificheranno

<sup>9</sup> Ullmann S., *Principi di Semantica*, traduzione di M. Mayer Modena e A. M. Finoli, Einaudi, 1977. Ullmann divide i mutamenti in due macrocategorie: i) quelli che avvengono per conservatorismo linguistico, dove il contenitore risulta più conservativo del contenuto, dettati da cambiamenti extralinguistici della cultura materiale e ideologica; ii) quelli che avvengono per innovazione linguistica, dovuti cioè a cause essenzialmente intralinguistiche. Il mutamento può avvenire inoltre in maniera differente a seconda dei rapporti tra nomi e significati: può avvenire cioè che a) il rapporto nome-significato cambi per associazioni tra sensi o che b) il rapporto nome-significato cambi per associazioni tra nomi. In entrambi i casi questo passaggio può avvenire per somiglianza o per contiguità. In a) se avviene somiglianza tra sensi si ha a metafora, se avviene contiguità tra sensi si ha la metonimia. In b) se avviene somiglianza tra nomi si ha il fenomeno dell'etimologia popolare, se avviene per contiguità di nomi si ha il fenomeno dell'ellissi.

<sup>10</sup> Coseriu E., *Teoria del linguaggio e linguistica generale*, Bari, Laterza, 1971, pp. 225-286.



significati diversi, essendo questi ultimi ristrutturati in porzioni differenti. Questo permette di discriminare due situazioni: quando avviene un mutamento di contenuto - una sostanza che si sposta intatta da un contenitore a un altro - da quando si verifica invece un mutamento nel contenuto - una sostanza che viene riorganizzata in porzioni differenti all'interno di diversi contenitori. Il primo tipo di fenomeno è indicato da Coseriu come "sostituzione", mentre il secondo è chiamato "modificazione", che costituisce la comparsa o scomparsa di una porzione di significato, o tratto distintivo<sup>11</sup>, e che implica sempre un'avvenuta sostituzione - perché un contenuto, per essere ristrutturato in contenitori differenti, deve necessariamente rompere il legame contenitore-contenuto. Queste due fondamentali tipologie di mutamento sono però relegate solo a una situazione in cui due o più parole entrano in opposizione, cioè in una relazione i cui elementi potrebbero essere sostituiti in una frase, mantenendo ugualmente una certa coerenza semantica, ma producendo di fatto significati diversi. Ad esempio, una pietra in latino può essere sia *atra* che *nigra*, e in ogni caso sarà di un colore nero, ma in un caso piuttosto che nell'altro sarà denotata da un diverso livello di brillantezza, e questa differenza di significato nella sostituzione degli elementi dell'opposizione fornisce proprio il tratto distintivo dell'opposizione stessa. Quindi, in verità, i tratti distintivi esistono solo a garanzia di un'opposizione, presente, in altri termini, tra quell'insieme di parole che «"coprono" una data area concettuale, delimitandosi a vicenda nel significato»<sup>12</sup>.

Riassumendo, finora si è considerato il mutamento semantico secondo due concezioni differenti, come mutamento del significato delle parole nel loro rapporto espressione-contenuto, e come mutamento dei tratti distintivi, nella ristrutturazione dei rapporti tra contenuti lessicali. Oltre a questo, se si presuppone che i fenomeni del linguaggio siano legati in modo imprescindibile alle facoltà mentali, il mutamento semantico può permetterci di osservare il ruolo dei fenomeni cognitivi nella lingua<sup>13</sup>, primo fra tutti, la metafora, consistente nell'associazione di un'esperienza con un'altra

---

<sup>11</sup> Ciò che è stato chiamato generalmente finora come "porzione di significato" si identifica in particolare con degli elementi detti "tratti distintivi", elementi differenziali in cui le unità di una certa struttura linguistica - in questo caso semantica - si possono analizzare. Dove questa struttura è fondata dall'esistenza di opposizioni distintive - come nell'esempio precedente di *ater/niger* "brillante/non-brillante". Se un'opposizione viene meno, vorrà dire che sono venuti meno i tratti distintivi che la garantivano.

<sup>12</sup> Casadei F., *Lessico e Semantica*, cit., p.59.

<sup>13</sup> Györi G., *Semantic Change as Linguistic Interpretation of The World*, John Ben-jamins Publishing Company, 2000.

cui si ha maggiore familiarità, per facilitarne la comprensione<sup>14</sup>. Ciò si può vedere considerando le diverse accezioni di una parola come *sostenere*, che può esser riferita tanto a oggetti concreti quale “muro, parete” quanto ad entità astratte come “argomentazione, tesi”. Secondo questa interpretazione «dal momento che il concetto ARGOMENTO è parzialmente compreso in termini della metafora UN ARGOMENTO È UNA COSTRUZIONE, il significato di “puntellare” - nel nostro esempio *sostenere* - nel concetto ARGOMENTO dipenderà dal significato che esso ha nel concetto COSTRUZIONE»<sup>15</sup>. In questo senso *sostenere* avrebbe subito un’espansione di significato<sup>16</sup>, o quanto meno un’espansione nella sua estensione, in virtù del meccanismo cognitivo generale della metafora. Metafora che, allora, non si configura tanto come uno strumento linguistico quanto un vero e proprio mezzo cognitivo, usato per pensare e comprendere il reale. È appunto tramite il mutamento semantico che si possono individuare le modifiche nella struttura concettuale, individuate nella ristrutturazione delle categorie concettuali:

The fact for example that the word *hawk* derives from PIE – per “Proto-Indo-European” – *\*kap-* “to grasp” suggests that referents of *hawk* were at once time not only referred to, but must also have been conceptualized initially as, instances of something grasping. Such a specific instance will be more or less central to a particular initial category under which it is first categorized. [...] Since referents of hawk are not simply grasping things but have a large degree of functional autonomy in our lives [...] in time they will be conceptually and semantically relegated to a different category and coded in the language accordingly.<sup>17</sup>

Se è infatti a partire dal modo in cui viene concettualizzata la realtà che viene usato il linguaggio, un mutamento di significato corrisponde sempre a un cambiamento nelle nostre categorie concettuali, e l’individuazione e lo studio del primo fenomeno può aiutare a comprendere meglio il secondo.

---

<sup>14</sup> Lakoff G., Johnson M., *Metaphors We Live By*, Bompiani, 1980.

<sup>15</sup> Ivi, pp.135-136.

<sup>16</sup> In realtà se si considerano i sensi di una parola come i significati che questa assume in particolari contesti, il significato, in quanto tale, di *sostenere* rimarrebbe immutato, vedendosi esteso solo rispetto ai suoi possibili referenti.

<sup>17</sup> Györi G., *Semantic Change as Linguistic Interpretation of The World*, cit.

## 1.2 Cause del mutamento

In base al punto di vista si può considerare il mutamento semantico come qualche cosa di diverso, a seconda che si cerchi il mutamento nel rapporto significante-significato, nella comparsa o scomparsa di tratti distintivi, o infine nella comparsa o scomparsa di categorie concettuali. Ognuna di queste prospettive è però suscettibile alle medesime cause, finché garantiscono le condizioni di avvenuto mutamento - come nel caso delle modificazioni nella semantica diacronica strutturale. Un primo ordine di cause può essere rappresentato dai motivi di carattere storico, in cui la lingua risulta essere più conservativa della civiltà materiale e ideologica, come nel caso di *galera* sopra citato. In un altro caso è possibile che il mutamento avvenga poiché dovuto alla diversa stratificazione sociale dei parlanti, quindi, alla trasposizione di significato da una varietà di lingua ad un'altra; per esempio, il latino «*ad-ripare* “giungere alla riva” ha allargato il suo significato nell'italiano *arrivare* [...] quando dal linguaggio della navigazione è passato all'uso comune»<sup>18</sup>. Ancora, un terzo gruppo è costituito dalle cause di natura psicologica, come fattori emotivi e tabù, di cui un esempio è il termine usato per riferirsi ad “orso” nelle lingue indoeuropee che non hanno continuato il termine originario indoeuropeo, e l'hanno sostituito «da varie circonlocuzioni di tipo inconfondibilmente eufemistico e propiziatorio: lituano *lakys* “leccatore”, russo *medved* “mangiatore di miele”, mediogallese *melfochyn* “maiale da miele”»<sup>19</sup>. Infine, può accadere che si abbia mutamento per influenza straniera, cui un caso emblematico è quello di *realizzare*, che in italiano ha accolto il senso di “accorgersi, capire” dall'inglese *to realize*, senso di cui originariamente il termine italiano era sprovvisto.

Il fatto che il significato possa cambiare è reso possibile da alcune proprietà del linguaggio verbale che riguardano il segno linguistico, nell'arbitrarietà dell'associazione tra espressione e contenuto, e il significato, nella sua indeterminatezza. L'arbitrarietà si manifesta in generale nel fatto che «se i segni linguistici non fossero fondamentalmente arbitrari, le parole delle diverse lingue dovrebbero essere tutte molto simili»<sup>20</sup> e «allo stesso modo [...] parole simili nelle diverse lingue dovrebbero designare cose o concetti simili»<sup>21</sup>. In particolare, inoltre, l'arbitrarietà non si pone solo nel rapporto significante-

---

<sup>18</sup> Lazzeroni R., *Il mutamento linguistico*, cit., p. 24.

<sup>19</sup> Ullmann S., *Principi di Semantica*, cit., p. 219.

<sup>20</sup> Berruto G., Cerruti M. S., *La linguistica. Un corso introduttivo*, UTET De Agostini, 2017, p.8.

<sup>21</sup> Ibidem.

significato, ma anche nei confronti dell'organizzazione dei contenuti lessicali, poiché, come si è visto per la semantica diacronica strutturale, una lingua può organizzare i significati in maniera diversa anche a distanza di tempo. Diversamente, il significato è intrinsecamente indeterminato, nel senso che risulta molto difficile dare una definizione esatta di un qualsiasi significato, o assegnare ai contenuti una categoria netta e precisamente delimitata. Se, per esempio, si considera la parola *tazza*, è difficile dire se un oggetto senza manico, e un po' più alto e stretto di una tazza, sia ancora una tazza<sup>22</sup> e, allo stesso modo, è difficile dire quale sia il confine tra una tazza di questo genere e un oggetto come un bicchiere. Anche se «si tratta di un problema di descrizione formale più che un problema della lingua»<sup>23</sup>, la vaghezza semantica si configura come un fenomeno fondamentale, perché garantisce una grande flessibilità ed estendibilità ai significati delle parole nella misura in cui vengono usati dai parlanti. A conti fatti, quindi, se l'associazione tra contenitore e contenuto è un fatto arbitrario, come anche l'organizzazione dei contenuti, un contenitore che prima ospitava un certo contenuto può accoglierne facilmente un altro, e organizzarlo liberamente. Analogamente, se il significato non è determinato né determinabile precisamente, è possibile che possa ridimensionarsi, modificarsi o, non essendoci confini netti tra significati, intersecarsi con altri sensi vicini senza difficoltà. In questo modo il rapporto tra contenitore e contenuto diventa molto labile ed è facile che si possa incrinare.

Il mutamento semantico si configura quindi come un fenomeno molto comune, tanto da non dover essere necessariamente cercato in periodi molto lunghi, dal momento che, come nel caso di *movida*, può capitare che una parola cambi di significato anche nell'arco di soli vent'anni. Inoltre, si configura come un fenomeno difficilmente prevedibile nelle sue singole manifestazioni, e fortemente connesso a fattori di ordine storico-sociale e psicologico-cognitivo. In questo lavoro, ad ogni modo, non si tratterà del mutamento semantico né in chiave cognitiva né in chiave strutturalista, ma, limitatamente ai criteri della semantica diacronica tradizionale, si tenterà di capire quanto bene una semantica distribuzionale riesca a codificare il cambiamento di significato, inteso come qualunque mutamento nel rapporto espressione-contenuto. Negli ultimi capitoli, infatti, sulla scorta dei risultati ottenuti da un modello informatico di *word*

---

<sup>22</sup> Casadei F., *Lessico e Semantica*, cit., pp.37-38.

<sup>23</sup> Labov W., *Il continuo e il discreto nel linguaggio*, cit., p.160.

*embedding*, si cercherà di dare una spiegazione del perché alcune parole siano effettivamente mutate e insieme del perché il modello le consideri tali.

## 2: Semantic shift nel Word embedding

### 2.1 Semantica distribuzionale e contesto

Uno “spazio semantico” o *word embedding*, è uno spazio multidimensionale in cui si può associare a ogni parola una coordinata. Le tecniche che hanno lo scopo di rappresentare le parole in uno spazio di questo genere si rifanno alla teoria della semantica distribuzionale, che codifica il significato delle parole in base alla loro posizione reciproca nello spazio, in termini di una relazione di vicinanza, detta similarità semantica<sup>24</sup>. Ciò implica che a una vicinanza tra parole corrisponda una vicinanza di significato, che non rispecchia tuttavia una relazione semantica “classica”<sup>25</sup>, ma unicamente una distanza ravvicinata nello spazio.

Lo snodo fondamentale della semantica distribuzionale è l’ipotesi distribuzionale, che presuppone che due parole siano semanticamente simili quando possano comparire in contesti simili<sup>26</sup>. Per questo motivo, una volta che vengono registrati i dati statistici riguardo il contesto d’occorrenza di una parola, si può misurare la sua similarità con le altre parole in base alla condivisione o meno dei medesimi dati. La procedura di calcolo delle posizioni delle parole, tale che ogni parola risulti essere vicina alle parole che hanno un contesto simile, verrà trattata in 2.2.

Considerare lo spazio semantico come formato da parole, ovvero elementi semantico-lessicali, porta a prediligere un contesto linguistico della stessa tipologia, intendendo il contesto come l’insieme degli elementi intratestuali - semantico-lessicali - in cui si trova inserita una parola. Ciò succede anche se, in linea di principio, si potrebbe selezionare un tipo di contesto differente, non solo semantico-lessicale. A riguardo, è stato dimostrato che il tipo di contesto utilizzato condiziona fortemente il tipo di significato che la similarità semantica codifica nello spazio semantico<sup>27</sup>, e che, in particolare, l’implementazione di un contesto a dipendenza sintattica porta a codificare delle

---

<sup>24</sup> Lenci A., *Modelli distribuzionali del lessico: metodi computazionali per l’analisi semantica* in «Informatica Umanistica», 3 (2010), pp. 57–69.

<sup>25</sup> Come iperonima e iponimia.

<sup>26</sup> Se vengono quindi usate in maniera simile e in simili contesti: si pensi per esempio, all’uso italiano delle parole “dottore” e “medico”, che possono ricorrere in contesti linguistici molto simili.

<sup>27</sup> Levy O., Goldberg Y., *Dependency-based word embeddings in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Baltimora, Association for Computational Linguistics, 2014, pp. 302-308.

similarità semantiche assimilabili a relazioni di coiponimia<sup>28</sup>. Nel nostro lavoro, tuttavia, considereremo sempre il contesto come il co-testo semantico-lessicale presente a una certa distanza arbitraria, detta finestra di contesto<sup>29</sup>, da una parola data.

Nell'ipotesi distribuzionale si può distinguere una versione “forte” e una “debole”, differenti per il diverso peso che è attribuito al contesto nella costituzione del significato<sup>30</sup>. Nel primo approccio la distribuzione linguistica di una parola è considerata la causa prima del suo significato, mentre nel secondo il comportamento distribuzionale di una parola è visto esclusivamente come correlativo del suo comportamento semantico, partendo da un presupposto contrario, cioè che sia il significato di una parola a determinarne il comportamento nel contesto<sup>31</sup>. Una distinzione, quindi, basata sulla differenza tra contenuto e comportamento semantico e sulla questione se sia l'uno, o al contrario l'altro, a venire determinato dal contesto, secondo due sfondi teorici diversi: il primo caso, cui si fa riferimento a Zellig Harris, condivide un approccio strutturalista, mentre il secondo, cui si fa riferimento a George Miller e Walter Charles, eredita un'impostazione cognitivista della semantica.

Va segnalato però che, a prescindere dall'atteggiamento teorico, ci sono in verità alcuni aspetti propri dell'impostazione distribuzionale che colgono bene certe caratteristiche del significato. In una semantica distribuzionale, infatti, la rappresentazione del significato è intrinsecamente quantitativa e graduale - di contro, per

---

<sup>28</sup> Per coiponimo si intende «ciascuno dei termini aventi in comune una relazione di iponimia con uno stesso iperonimo», dove l'iponimia è la «subordinazione semantica di un vocabolo rispetto a un altro di significato più generico ed esteso», che viene detto appunto iperonimo. Cfr. *Il Nuovo De Mauro*, [20/10/2021], <https://dizionario.internazionale.it/parola/coiponimo>, Internazionale e *Il Nuovo De Mauro*, [20/10/2021], <https://dizionario.internazionale.it/parola/iponimia>, Internazionale.

<sup>29</sup> Per esempio, data la frase: “L'acqua in natura è alla base di tutte le forme di vita conosciute”, se consideriamo la parola “tutte” e stabiliamo che debbano rientrare nel suo contesto solo le parole in una finestra di contesto di ampiezza uguale a tre, il contesto di “tutte” sarà allora [“alla”, “base”, “di”, “le”, “forme”, “di”]. Si è notato anche che l'impiego di una finestra corta porta a codificare informazioni più di tipo sintattico, mentre al contrario con finestre larghe la similarità semantica codifica più una prossimità d'argomento; per questo motivo spesso si usa una finestra di contesto di media ampiezza, in genere di grandezza cinque.

<sup>30</sup> Lenci A., *Distributional semantics in linguistic and cognitive research* in «Italian journal of linguistics», 20(1) (2008), pp. 1-31.

<sup>31</sup> A. Lenci (Ibidem) a proposito cita anche Levin, il quale afferma, in riferimento alle strutture argomentali dei verbi, che “argument structures might in turn be derivable to a large extent from the meaning of words.” Cfr. Levin B., *English verb classes and alternations: A preliminary investigation*, Chicago, University of Chicago press, 1993.

esempio, a una semantica componenziale<sup>32</sup>. Questa è una caratteristica che sembra essere in linea soprattutto con la proprietà dei segni linguistici di avere confini indefiniti e sfumati, detta “vaghezza semantica”, che costituisce un grosso limite per semantiche d’impostazione qualitativa. Proprio a causa della vaghezza, infatti, risulta problematico assegnare una categoria a un elemento che presenta caratteri di più categorie o a elementi a confine tra categorie<sup>33</sup>. Un problema di questo genere non si pone invece per una semantica distribuzionale - in cui la codifica del significato è misurabile - per la quale non esistono elementi a confine di categorie assolute, ma solo elementi più o meno vicini tra di loro.

Bisogna precisare, tuttavia, che la stessa nozione di similarità semantica, che nello spazio semantico codifica il significato, è vaga e problematica essa stessa, in quanto non distingue tra loro le diverse relazioni semantiche, quali l’iponimia e l’iperonimia, che, essendo relazioni asimmetriche, non possono essere codificate nello spazio semantico, se non a particolari condizioni<sup>34</sup>. Infatti, non sembra che la similarità semantica codifichi una effettiva similarità tra parole vicine - quale può essere una similarità funzionale tra coiponimi - quanto piuttosto una vaga somiglianza<sup>35</sup>. Un problema diverso, connesso sempre alla struttura dello spazio semantico, consiste nella difficoltà di rendere conto del principio di composizionalità semantica, secondo cui il significato di un’espressione risulta ricavabile dal significato dei suoi membri e dal modo in cui questi sono combinati. Infatti, sebbene lo spazio semantico sembri codificare una qualche forma di composizionalità, questa è relegata solo alla misura dei suoi elementi<sup>36</sup>, e non si è ancora riuscita a codificare in modo soddisfacente per combinazioni di più parole<sup>37</sup>.

---

<sup>32</sup> dove il significato è rappresentato da attributi discreti a opposizioni binarie, e di natura qualitativa: ad esempio, se il significato di “uomo” è rappresentato con i seguenti attributi: [+animato, +umano, +adulto, +maschio], il significato di “bambina” sarà: [+animato, +umano, -adulto, -maschio].

<sup>33</sup> Casadei F., *Lessico e Semantica*, cit.

<sup>34</sup> Benotto G., *Modelli distribuzionali delle relazioni semantiche: il caso dell’iperonimia* in *Animali, Umani, Macchine. Atti del convegno 2012 del CODISCO*, Roma-Messina, Corisco, 2013, pp. 85-97. Si noti che la difficoltà nel codificare relazioni asimmetriche è connessa anche alla problematica dell’inferenza, interpretabile come un insieme di relazioni di iponimia, iperonimia e coiponimia. Cfr. Lenci A., *Distributional semantics in linguistic and cognitive research*, cit.

<sup>35</sup> Lenci A., *Distributional Models of Word Meaning*, in «Annual Review of Linguistics», 4 (2018), pp. 151-171.

<sup>36</sup> I modelli di semantica distribuzionale possono eseguire somme tra i vettori delle parole, e si è notato che sono molto performanti nello svolgere tali operazioni. Per esempio, l’espressione [re]-[uomo]+[donna] (dove il vettore è indicato con [ ]) dà come risultato un vettore molto simile a quello di [regina].

<sup>37</sup> Lenci A., *Distributional semantics in linguistic and cognitive research*, cit.



Il ruolo fondante delle distribuzioni linguistiche, quindi dei documenti testuali da cui estrarle, rende molto comodo l'utilizzo dei modelli di semantica distribuzionale per via della crescente disponibilità di grandi collezioni di testi in formato digitale, ed è uno dei motivi del loro successo in linguistica computazionale e in psicologia. Bisogna aggiungere però che questi modelli sono molto suscettibili ai corpora da cui attingono i dati, poiché tutta la procedura d'elaborazione dello spazio semantico è ad essi molto vincolata, sia nella quantità di parole rappresentabili, sia nella qualità dell'operazione di calcolo delle coordinate, che è tanto più accurata tanto più estese sono le informazioni riguardo le distribuzioni. Inoltre, un problema a ciò derivato è la difficoltà di discernere le diverse accezioni e i diversi sensi con cui una parola può essere usata, per via, rispettivamente, dell'assenza nel corpus di una distinzione tra un'accezione e un'altra rispetto a una stessa stringa di testo, e per il tipo di contesto usato come base di distribuzione, che se non contiene informazioni su un certo tipo di senso - ad esempio pragmatico - non potrà codificarne l'informazione. È chiaro quindi che lo spazio semantico è fortemente limitato dai documenti con cui è costruito, e non potrà che rappresentare lo spazio semantico di un uso della lingua, che sarà più o meno rappresentativo della lingua intera quanto lo sono i documenti che si esaminano rispetto all'insieme dei testi prodotti in tale lingua.

Il *word embedding* si inquadra allora come un modello semantico empirico distribuzionale; e se i suoi usi spaziano dal campo dell'Information Retrieval<sup>38</sup> ai sistemi di traduzione automatica<sup>39</sup>, alla classificazione dei testi<sup>40</sup>, è perché permette di eseguire in modo molto facile operazioni su dati semantico-lessicali, cosa altrimenti piuttosto complessa senza l'impiego di vettori su cui è possibile svolgere analoghe operazioni di calcolo.

## 2.2 *Language models e principali modelli di Word Embedding*

Ci sono diverse tipologie di dati che si possono ricavare dalle distribuzioni linguistiche. L'informazione più basilica è la frequenza di occorrenza assoluta di una

---

<sup>38</sup> Galke L., Saleh A., Scherp A., *Word embeddings for practical information retrieval* in «INFORMATIK», P275 (2017).

<sup>39</sup> Bojanowski P., Grave E., Joulin A., Mikolov T., *Enriching word vectors with subword information* in «Transactions of the Association for Computational Linguistics», 5 (2017), pp. 135-146.

<sup>40</sup> Wang S. I., Manning C. D., *Baselines and bigrams: Simple, good sentiment and topic classification* in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Jeju Island, Association for Computational Linguistics, 2012, pp. 90-94.

parola, che può essere relativizzata al numero totale delle parole del corpus, e in questo modo usata per calcolare la *Maximum likelihood estimation*, ovvero la probabilità di estrarre casualmente dal corpus una parola data, utile anche per calcolare il valore di *Pointwise mutual information*<sup>41</sup> nelle collocazioni<sup>42</sup>. Invece, per quanto riguarda la procedura d'elaborazione del sistema di coordinate, si possono individuare due gruppi, distinti in base alle diverse modalità di costruzione dello spazio semantico.

Da una parte sono raggruppati tutti quei modelli esclusivamente statistici, in cui sono gli stessi dati di distribuzione a formare i valori delle coordinate dei vettori, senza il passaggio verso un'elaborazione ulteriore. In questo caso il vettore di una parola può essere ricavato dal conteggio di collocazioni tra la parola e ogni altra parola del corpus, oppure dal conteggio di quante volte ogni parola del corpus compare nella finestra di contesto della parola data, ovvero la co-occorrenza<sup>43</sup>. In quest'ultimo caso si viene a formare una matrice di co-occorrenza, dove il vettore di una parola è formato dalla frequenza di co-occorrenza rispetto a ogni altra parola del corpus, e il suo numero di dimensioni è determinato dalla grandezza del vocabolario ( $V$ ) del corpus. Prendendo in considerazione come corpus le seguenti frasi: i) *ogni giorno vado a scuola con l'autobus*, ii) *ogni sera vado a teatro con l'autobus*, iii) *ogni giorno vado a scuola a piedi*; considerando una finestra di contesto uguale a tre, la matrice di co-occorrenza risulterà essere la seguente (Tabella 1):

	a	con	giorno	l'autobus	ogni	piedi	scuola	sera	teatro	vado
a	1	2	2	2	3	2	3	1	1	4
con	2	0	0	2	0	0	1	0	1	2
giorno	2	0	0	0	2	0	2	0	0	2
l'autobus	2	2	0	0	0	0	1	0	1	0
ogni	3	0	2	0	0	0	0	1	0	3
piedi	2	0	0	0	0	0	1	0	0	0
scuola	3	1	2	1	0	1	0	0	0	2
sera	1	0	0	0	1	0	0	0	1	1
teatro	1	1	0	1	0	0	0	1	0	1
vado	4	2	2	0	3	0	2	1	1	0

Tabella 1, Matrice di co-occorrenza.

<sup>41</sup> Che è così calcolato:  $pmi(a, b) = \log \frac{p(a, b)}{p(a)p(b)}$ ; cioè come la probabilità della collocazione  $(a, b)$ , come osservata rispetto alla *MLE*, rispetto alla probabilità attesa della combinazione  $(a, b)$ , ipotizzando l'indipendenza di  $a$  e di  $b$ . Le collocazioni sono combinazioni particolarmente frequenti di parole, giustificate non tanto da un legame semantico quanto invece dall'esser consolidate dall'uso. Cfr. Tiberii P., *Dizionario delle collocazioni*, Bologna, Zanichelli, 2012, p. 3.

<sup>42</sup> Lenci A., Montemagni S., Pirelli V., *Testo e computer. Elementi di linguistica computazionale*, Roma, Carocci, 2016.

<sup>43</sup> La co-occorrenza si verifica quando delle parole occorrono assieme nella stessa finestra di contesto.

Dove il vettore, per esempio, di *vado* (in Tabella 1) sarà [4,2,2,0,3,0,2,1,1,0], dove “4” rappresenta il valore di *vado* nella prima dimensione *a*, “2” nella seconda dimensione *con*, e così via. In virtù dell’ipotesi distribuzionale si può pensare che parole che abbiano frequenze di co-occorrenza simili siano simili anche per significato, e in questo modo si giustifica la creazione di uno spazio semantico, che avrà dimensione  $V \times V$ . Ciononostante, il problema principale di questo approccio è che i vettori così formati risultano composti da un numero notevole di dimensioni, pari alla grandezza di  $V$ , presentando tuttavia un alto numero di valori vuoti, soprattutto per  $V$  molto grandi. Modelli di questo tipo, implementati in realtà da strumenti informatici più avanzati, sono usati nel campo della classificazione dei testi, come il filtro spam delle e-mail<sup>44</sup>.

A un secondo gruppo invece appartengono i modelli probabilistici, in cui la posizione delle parole si ricava a partire dalle loro probabilità di occorrenza. Queste sono calcolate stimando la probabilità di una certa sequenza di  $n$  parole  $w$ :  $P(w_1, w_2, w_3, \dots, w_n)$ <sup>45</sup> o la probabilità d’occorrenza di una parola  $w$  di posizione  $n$  rispetto a tutte le parole precedenti:  $P(w_n | w_1, w_2, \dots, w_{n-1})$ , che si assume, per la proprietà di Markov, essere approssimabile alla probabilità d’occorrenza di una parola  $w$  di posizione  $n$  rispetto solo a  $N$  parole precedenti:  $P(w_n | w_{n-N+1}, \dots, w_{n-1})$ <sup>46</sup>. Ciò rende possibile stimare l’occorrenza di una parola a partire dalla presenza di poche parole precedenti, ma non solo. Infatti, questi modelli sono attivamente impiegati nel *word embedding*, in due principali architetture, di cui l’una ha la funzione di predire le parole nella finestra di contesto di una parola  $w_t$  data, l’altra viceversa di predire una parola  $w_t$  a partire da un contesto noto; si parla rispettivamente di *Skip-gram* e *Continuous Bag-of-Words*. Per generare la posizione delle parole dalle stime di probabilità si opera nel seguente modo: si crea una rete neurale avente come input ogni parola del vocabolario ( $V$ ) del corpus e come output una stima di probabilità che ogni parola occorra nel contesto delle parole di input, in modo tale che una parola  $w$  restituirà come output un numero  $V$  di stime. Successivamente si considerano come vettore quei “parametri” (di dimensione  $D$  arbitraria) che nella rete neurale mediano il passaggio dall’input all’output. In particolare,

<sup>44</sup> Wang Z. J., Liu Y., Wang Z. J., *E-mail filtration and classification based on variable weights of the Bayesian algorithm*. In «Applied Mechanics and Materials», 513 (2014), pp. 2111-2114.

<sup>45</sup> Che viene scomposta in  $P(w_1) P(w_2|w_1) P(w_3|w_1, w_2) \dots P(w_n | w_1, w_2, \dots, w_{n-1})$  per il teorema della probabilità composta.

<sup>46</sup> Jurafsky D., & Martin J. H., *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Draft of December 30, 2020, cap. 3.

a ogni parola di input è assegnato un vettore *one-hot*<sup>47</sup>, che selezionerà tra i parametri quelli che regolano la stima di probabilità che tutte le altre parole di  $V$  occorranza nel contesto di  $w$ . Ai parametri vengono inizialmente assegnati valori casuali e ad ogni  $w$  di input vengono tutti aggiornati (ottimizzati) in modo che l'output sia il più simile possibile all'effettiva occorrenza delle parole. I parametri selezionati da  $w$  costituiranno così il suo vettore nello spazio semantico, spazio che avrà quindi dimensione  $D \times V$ .

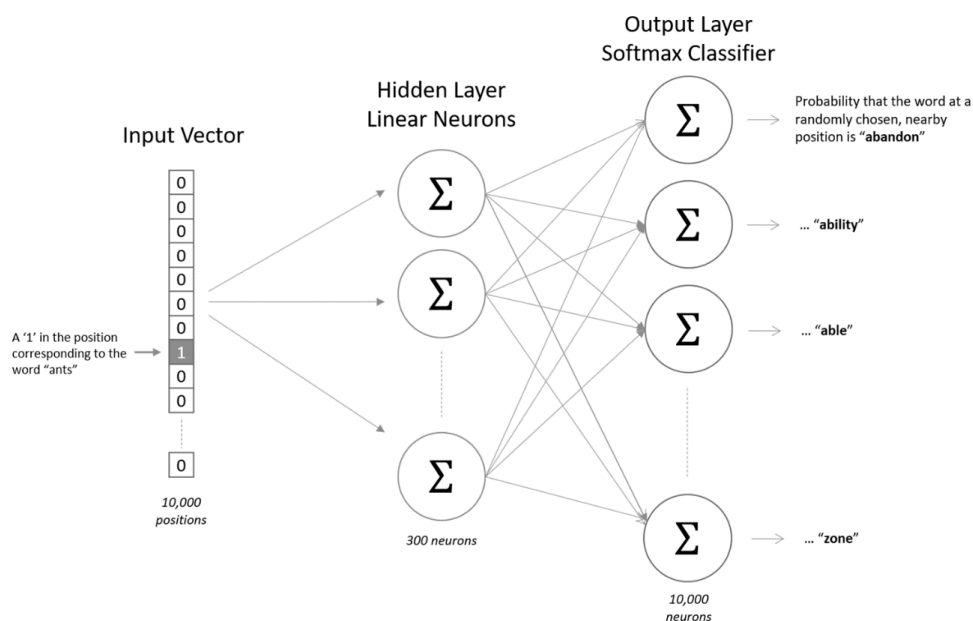


Figura 2, Architettura della rete neurale in Word2Vec. Cfr. McCormick C., *Word2Vec Tutorial - The Skip-Gram Model*, 2016, [20/10/2021], <http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/>.

I vettori così creati non hanno una dimensione fisica, e nemmeno statistica, come poteva essere per i vettori costruiti da una matrice di co-occorrenza, ma rappresentano unicamente dei parametri che riescono ad ottimizzare le probabilità d'occorrenza delle parole di  $V$  nel contesto di  $w$  con la loro l'occorrenza effettiva - e ciò fa sì che questi vettori non siano per noi interpretabili. Non sono nemmeno dei parametri assoluti, ma anzi estremamente relativi sia ad ogni singolo modello, che ad ogni singola elaborazione dello stesso modello con  $D$  e  $V$  uguali, e ciò comporta che ogni spazio semantico, nelle sue posizioni assolute, sia intrinsecamente indipendente.

Nonostante questo, i modelli di *Word Embedding*, si sono dimostrati in grado di calcolare valori di similarità che risultano essere molto simili ai punteggi assegnati

<sup>47</sup> Ovvero un vettore costituito da una serie di zeri, e un solo valore di "1", indicante la posizione di  $w$ .

sperimentalmente dai parlanti, e di sapere svolgere alcuni compiti di categorizzazione, come determinare la categoria lessicale delle parole (*Part-Of-Speech tagging*)<sup>48</sup>. Infine, danno buoni risultati in problemi di analogia e nel cogliere il significato di relazioni<sup>49</sup>, quindi nella somma di vettori di parole in relazione semantica tra loro. L'esempio più famoso è l'espressione [re] – [uomo] + [donna], che risulta essere un vettore molto vicino a quello di [regina]. Queste capacità, inoltre, potendole confrontare con i risultati ottenuti dai parlanti, vengono usate anche per valutare il rendimento dei modelli di *word embedding* rispetto appunto a un metro di riferimento umano.

Il modello più rappresentativo è Word2Vec, introdotto nel 2013 da Mikolov et al.<sup>50</sup>, fondato sulle architetture *Skip-gram* e *Continuous Bag-of-Words*, che elaborano sempre i dati dell'“intorno”<sup>51</sup> delle parole, e rendono questo modello orientato verso le distribuzioni relative. Al contrario GloVe<sup>52</sup> (*Global Vectors*) lavora a partire dalla matrice di co-occorrenza delle parole del corpus, e vettorializza le parole secondo i parametri che ottimizzano la stima delle co-occorrenze.

Approcci di questo genere considerano il linguaggio come un sistema combinatorio, in cui le frasi sono costituite da simboli che non hanno un significato di per sé ma che assumono di senso solo quando rapportati alla probabilità di apparire insieme ad altri simboli. In questo senso, è come se ci confrontassimo con una lingua scritta in un alfabeto a noi ignoto, che codifica una parola con uno e un solo simbolo, e l'unico nostro modo per cogliere il significato dei suoi elementi fosse capire come siano relazionati, in base al modo in cui si combinano tra di loro. Così, un modello di *word embedding* cerca di capire cosa significhi un simbolo in base a come esso compare combinato in sequenza con altri simboli, in una situazione in cui ci possono essere svariate centinaia di migliaia

---

<sup>48</sup> Arseniev-Koehler A., Foster J. G., *Sociolinguistic Properties of Word Embeddings*, 2020, SocArXiv doi:10.31235/osf.io/b8kud.

<sup>49</sup> Jurafsky D., Martin J. H., *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Draft of December 30, 2020, cap. 6.

<sup>50</sup> Mikolov T., Chen K., Corrado G., Dean J., *Efficient estimation of word representations in vector space* in *Proceedings of Workshop at International Conference on Learning Representations*, Scottsdale, US, ICLR, 2013.

<sup>51</sup> Cioè la finestra di contesto.

<sup>52</sup> Pennington J., Socher R., Manning C.D., *Glove: Global vectors for word representation* in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, QA, EMNLP, 2014, pp. 1532–1543.

di elementi che si combinano tra di loro, come succede nel caso delle parole per corpora testuali di grandi dimensioni.

Per limiti di questo lavoro, tuttavia, non ci addentreremo nella questione se un approccio di questo tipo sia fondato dal punto di vista linguistico, né discuteremo le cause semantiche, sintattiche, pragmatiche e testuali che generano particolari sequenze di parole, ma ci limiteremo a verificare questi modelli nel compito di individuare il mutamento di significato delle parole.

### 2.3 *Diachronic embedding e Semantic shift*

Potendo realizzare uno spazio semantico che rappresenti il significato di un grande numero di parole, si è pensato di impiegare questi modelli per costruire, e poi confrontare, spazi semantici di differenti periodi storici, con l'obiettivo di individuare e quantificare dei cambiamenti nel significato delle parole. Come già accennato in precedenza, tuttavia, gli spazi vettoriali generati da modelli di *word embedding* sono intrinsecamente indipendenti l'uno dall'altro, in virtù dell'iniziale assegnazione casuale dei parametri nel livello "intermedio" - l'*"hidden layer"* - della rete neurale, rendendo quindi due spazi semantici non direttamente confrontabili. L'insieme delle tecniche per operare un confronto di questo tipo, insieme alle modalità di misura del mutamento semantico delle parole, viene designato con il nome di *diachronic embedding*.

Per questo genere di confronto si individua per prima cosa una tipologia di documenti testuali rilevante per uno studio diacronico, dopodiché, una volta che i documenti del corpus entro cui si vuole misurare mutamento sono raccolti, si procede a dividere il corpus complessivo in diverse "tappe" in corrispondenza di sub-corpora diacronicamente omogenei. Nel fare questo è necessario che i sub-corpora abbiano un contenuto analogo, presentando cioè abbastanza parole in comune da rendere in seguito utile un confronto, e che abbiano anche una dimensione non troppo diversa: che siano simili quindi per qualità e per quantità. Allora si procede a rendere gli spazi semantici confrontabili, allineando i vettori delle parole tra uno spazio e l'altro, facendo sì che una parola abbia la stessa posizione in entrambi gli spazi. In primo luogo, si possono allineare i vettori delle parole operando delle trasformazioni geometriche<sup>53</sup> tra i differenti spazi

---

<sup>53</sup> Come traslazione, rotazione e ridimensionamento uniforme.

vettoriali, presupponendo che essi siano diversi solo per il loro orientamento reciproco<sup>54</sup>. Altrimenti, utilizzando una procedura diversa, è possibile costruire gli spazi semantici in modo che siano allineati già in partenza, secondo la metodologia proposta da Kim et al.: «We chronologically train the model by initializing word vectors for subsequent years with the word vectors obtained from previous years»<sup>55</sup>. Con l’idea di ottenere «incrementally updated diachronic embedding models»<sup>56</sup>, ovvero modelli in cui la nuova posizione delle parole venga aggiornata a ogni tappa diacronica, essendo calcolata in base alla posizione precedente. Diversamente, è possibile calcolare la posizione delle parole dapprima in base alla distribuzione delle parole in tutte le tappe, per poi usare questa posizione “complessiva” per aggiornare singolarmente la posizione delle parole per ogni tappa<sup>57</sup>. Utilizzando una di queste modalità di allineamento si arriverà in una situazione in cui si hanno più spazi semantici allineati e, quindi, confrontabili, corrispondenti alle diverse tappe tramite cui è stato suddiviso il corpus complessivo.

Essendo il significato di una parola determinato dalla sua posizione nello spazio semantico, si può pensare di quantificare il suo grado di mutamento in base allo spostamento della sua posizione da uno spazio semantico anteriore a uno spazio semantico recente. Questo spostamento si può misurare o in termini di distanza della medesima parola  $w$  dalla posizione  $p_1$  alla posizione  $p_2$ , o in termini di discrepanza tra la posizione  $p_1$  e la posizione  $p_2$  tramite il calcolo dello scarto quadratico medio (in inglese *Mean squared error*, *MSE*) tra tutte le dimensioni  $i$  di  $w$  nel primo spazio  $S_1$  e nel secondo spazio  $S_2$  (1)<sup>58</sup>.

$$MSE = \frac{\sum_{i=1}^n (S_{1i} - S_{2i})^2}{n} \quad (1)$$

---

<sup>54</sup> Hamilton W. L., Leskovec J., Jurafsky D., *Diachronic word embeddings reveal statistical laws of semantic change*, 2016, arXiv preprint arXiv:1605.09096.

<sup>55</sup> Kim Y. et al., *Temporal Analysis of Language through Neural Language Models*, 2014, arXiv preprint arXiv:1405.3515, p.1.

<sup>56</sup> Kutuzov A. et al., *Diachronic Word Embeddings and Semantic Shifts: a Survey*, 2018, arXiv preprint arXiv:1806.03537, p. 1390.

<sup>57</sup> Yao Z. et al., *Dynamic word embeddings for evolving semantic discovery*, in: *Proceedings of the eleventh acm international conference on web search and data mining*, 2018, pp. 673–681.

<sup>58</sup> Peng H. et al., *Incrementally Learning the Hierarchical Softmax Function for Neural Language Models*, in: «Proceedings of the AAAI Conference on Artificial Intelligence», 31 (2017), pp. 3267–3273.

Dove  $n$  è il numero delle dimensioni di  $w$  e  $S_{1_i} - S_{2_i}$  è lo scarto tra la dimensione  $i$  nel primo spazio e nel secondo. In questi due casi si avrebbe allora un maggiore mutamento semantico in corrispondenza di, rispettivamente, una maggior distanza e una maggiore discrepanza, tra posizione iniziale e finale. Tuttavia, come è stato accennato all'inizio di questo capitolo, in una semantica distribuzionale il significato di una parola è codificato non dalla sua posizione assoluta, ma dalla sua posizione rispetto alle altre parole dello spazio. Motivo per cui, misurare il mutamento semantico di una parola in termini di spostamento della sua posizione assoluta risulta essere quantomeno contraddittorio con le premesse della stessa semantica distribuzionale. Inoltre, risulterebbe scorretto quando accada che, a subire mutamento di significato non sia solo una parola, ma tutto il campo semantico<sup>59</sup> a cui quella parola appartiene, in una situazione in cui la parola cambierebbe di posizione assoluta, mantenendo tuttavia inalterata la sua posizione relativa, e quindi il suo significato.

Per tener fede all'assunto della semantica distribuzionale, si può misurare il mutamento semantico di una parola in termini di variazione delle distanze tra essa e le altre parole dello spazio; di modo da considerare una parola mutata solo quando cambino i suoi rapporti di similarità con le altre parole. A riguardo, in letteratura si chiama *second-order embedding* il vettore di una parola formato dalle distanze tra quella parola e tutte le altre parole del corpus<sup>60</sup>. Inoltre, similmente alle modalità indicate per confrontare il vettore di una parola nel tempo, si può misurare il mutamento semantico di una parola misurando lo spostamento, o la discrepanza, del suo "vettore delle distanze" tra una posizione iniziale e una finale. Analogamente, si potrebbe valutare, considerando una parola  $w$ , non il vettore delle distanze tra  $w$  e tutte le parole del corpus, ma il vettore delle distanze tra  $w$  e un insieme arbitrario  $P$  di parole; tale in particolare è la procedura che si è adottata in questo lavoro, come si vedrà più in dettaglio in 3.4.

In una semantica distribuzionale va precisato che, siccome il significato è legato alla distribuzione delle parole, un mutamento di significato corrisponde a un cambiamento in questa distribuzione. In una situazione, come si è visto in 2.1, in cui il

---

<sup>59</sup> Per campo semantico si intende quell'insieme di parole che «"coprono" una data area concettuale, delimitandosi a vicenda nel significato». Cfr. Casadei F., *Lessico e Semantica*, cit., p.59.

<sup>60</sup> Eger S., Mehler A., *On the linearity of semantic change: Investigating meaning variation via dynamic graph models*, 2017, arXiv preprint arXiv:1704.02497.



significato può essere inteso tanto come una causa quanto come un effetto della distribuzione. Individuare il mutamento semantico nel cambiamento della distribuzione di una parola vuol dire individuare se c'è cambiamento nei suoi contesti d'occorrenza, cosa che può esser legata, o esser del tutto indipendente, al suo significato effettivo. Può essere indipendente quando una certa parola, che possa presentarsi in una vasta gamma di contesti differenti, venga registrata in un sub-corpus solo in certi tipi di contesti e non in altri. Quando succede, il modello può interpretare questa parola come semanticamente vicina a parole diverse tra un sub-corpus e l'altro. Per cui, al momento di misurare il cambiamento del suo significato, la parola risulterebbe mutata poiché vicina a parole differenti in periodi diversi, benché il suo significato rimanga invariato. È il caso della parola *cancelliera*, che può apparire mutata per via delle associazioni in anni recenti con la parola *Merkel*, che non si verificherebbero per periodi anteriori al 2005, data dell'elezione di Angela Merkel a Cancelliera federale della Germania. Casi come questo verranno trattati più nel dettaglio nel capitolo successivo, dove si analizzeranno in riferimento a uno studio del mutamento semantico in una collezione di articoli dall'archivio digitale di Repubblica.

Riassumendo, quindi, lo studio del mutamento semantico in un modello distribuzionale comporta ostacoli sia dal punto di vista computazionale, nell'allineamento di spazi altrimenti indipendenti, che dal punto di vista linguistico, nella necessità di registrare le occorrenze in una maniera non "scomposta" tra diversi sub-corpora. Sono proprio problemi di questo genere che vanno tenuti in considerazione nella valutazione del mutamento, onde evitare errori, che sono tanto più probabili quanti meno dati si hanno riguardo il contesto d'occorrenza delle parole, come si vedrà nel prossimo capitolo.

### 3: Un caso di studio sugli articoli di Repubblica dal 1990 al 2021

#### 3.1 Confronto e descrizione dei corpora utilizzati

In questo capitolo verrà affrontato il mutamento di significato in una serie di testi raccolti dall'archivio digitale di Repubblica, in un arco di tempo che va indicativamente dal 1990 al 2021. I testi sono stati raccolti in due periodi di riferimento: dal 1994 al 1997 e dal 2019 al 2021. Questa divisione individua due corpora, di dimensione simile, entro cui è stato misurato il mutamento semantico delle parole, prendendo come punto di riferimento il caso di *ripartenza*, la cui prima attestazione nell'archivio di Repubblica è datata all'anno 1993, e il cui uso è odiernamente accostato alla situazione di ripresa successiva alla pandemia di COVID-19.

Il primo corpus, che fa riferimento al periodo 1994-1997, è formato da 160.147 articoli, 3.636.960 frasi, 80.892.581 parole e 545.884 parole uniche, mentre il secondo, che fa riferimento al periodo 2019-2021, è formato da 187.061 articoli, 3.104.483 frasi, 76.405.707 parole e 469.569 parole uniche. Prima di essere utilizzati come base di *training* per il modello distribuzionale, questi corpora sono stati inizialmente processati tramite una “tokenizzazione” in *Python* secondo la funzione *nlk.tokenize.word\_tokenize* dalla libreria *nlk.tokenize*, dopodiché sono stati rimossi i simboli di punteggiatura. Questi corpora, contenendo testi giornalistici, sono rappresentativi solo dell'italiano scritto e, oltretutto, della variante linguistica dell'italiano giornalistico. Si è voluto procedere al confronto di testi simili sociolinguisticamente per isolare al meglio la variabile diacronica e per avere un maggior grado di confrontabilità anche tra la quantità di parole condivise tra i due corpora, che è pari a 250.314 parole uniche.

#### 3.2 Training, allineamento e calcolo delle distanze

Per il confronto dei modelli semantici derivati dal primo e dal secondo corpus, rispettivamente *A(1994-1997)* e *B(2019-2021)*, si sono sperimentate due diverse modalità di allineamento, generando perciò complessivamente quattro spazi semantici. Due spazi, *S<sub>A</sub>incr.* e *S<sub>B</sub>incr.*, fanno riferimento alla modalità di allineamento, citata nel capitolo precedente, volta ad ottenere «incrementally updated [...] models»<sup>61</sup>. Mentre gli altri due spazi, *S<sub>A</sub>comp.* e *S<sub>B</sub>comp.*, sono stati allineati secondo la modalità, citata sempre nel capitolo

---

<sup>61</sup> Kutuzov A. et al., *Diachronic Word Embeddings and Semantic Shifts: a Survey*, cit., p. 1390.

precedente, che utilizza le “posizione complessive” delle parole come base di partenza per ricavare le loro posizioni nei singoli spazi semantici.

Al fine di generare questi spazi semantici, per entrambe le modalità di allineamento, sono state allenate<sup>62</sup> in *Python* due reti neurali, con i corpora e le modalità sopracitate, basate sul funzionamento di Word2Vec tramite la funzione *models.word2vec* dalla libreria *gensim.models*, lasciando invariati i parametri di base. Inoltre, durante la fase di *training* si sono individuate le collocazioni di bigrammi<sup>63</sup> utilizzando in *Python* la funzione *models.phrases* dalla libreria *gensim.models.phrases*, lasciando invariati i parametri di base. Infine, per quanto riguarda la misura della similarità semantica, le distanze tra vettori sono state sempre calcolate come *cosine similarity* (2).

$$\text{cosine similarity} = \cos(\vartheta) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|} \quad (2)$$

Dove  $v_1 \cdot v_2$  è il prodotto scalare tra i due vettori  $v_1$  e  $v_2$  e  $\|v_1\| \|v_2\|$  è il prodotto delle loro norme. Questa distanza restituisce un valore compreso tra -1 e 1, dove, un valore di similarità pari 1 indica che due vettori sono simili, un valore di -1 indica al contrario che due vettori sono opposti, e infine un valore di 0 che non c'è correlazione tra i vettori.

### 3.3 Valutazione dei modelli

Come è stato accennato in 2.2, i modelli distribuzionali vengono valutati confrontando le loro performance rispetto a performance umane in medesimi compiti. In questo lavoro, in particolare, i modelli che abbiamo ottenuto sono stati valutati nel test dell'analogia, dove, data una relazione tra quattro elementi  $a, b, c, d$  per cui  $a:b = c:d$ , il modello cerca di predire un elemento a partire da tre elementi noti. Questo viene valutato per diversi compiti specifici, come l'abilità del modello nel saper codificare la relazione singolare-plurale, per esempio nel caso dell'analogia *ratto:ratti = uccello:x*, in cui viene valutato se  $x$  corrisponde o meno a *uccelli*. Per la valutazione dei quattro modelli creati ( $S_A^{incr.}$ ,  $S_B^{incr.}$ ,  $S_A^{comp.}$  e  $S_B^{comp.}$ ), sono state utilizzate le analogie prodotte nel lavoro di Berardi G. et al.<sup>64</sup> dalla trasposizione in italiano del test di analogia di Google, sviluppato da Mikolov T. et al.<sup>65</sup>. Le prestazioni, riportate in Tabella 2, sono

<sup>62</sup> Nella fase di *training* della rete neurale.

<sup>63</sup> Ovvero collocazioni formate da due parole.

<sup>64</sup> Berardi G., Esuli A., Marcheggiani D., *Word Embeddings Go to Italy: A Comparison of Models and Training Datasets*, in *6th Italian Information Retrieval Workshop*, Cagliari, 2015.

<sup>65</sup> Mikolov T. et al., *Efficient estimation of word representations in vector space*, cit.

relative al rapporto, per ogni compito, tra test corretti e test totali. Inoltre, affianco a ogni valore è presente, in parentesi tonde, anche la differenza tra il nostro risultato e quello ottenuto, nello stesso test, dal modello Word2Vec<sup>66</sup> di Berardi G. et al. basato su BOOKS, una collezione di 31.432 e-books gratuiti scritti in italiano<sup>67</sup>.

	<i>Incremental update</i>		<i>Posizione complessiva</i>	
	$S_A^{incr.}$	$S_B^{incr.}$	$S_A^{comp.}$	$S_B^{comp.}$
<i>Compito valutato</i>	<i>A (1994-1997)</i>	<i>B (2019-2021)</i>	<i>A (1994-1997)</i>	<i>B (2019-2021)</i>
Capital-common-countries	9.74% (-75%)	14.50% (-70%)	12.06% (-73%)	15.42% (-69%)
Capital-world	7.12% (-40%)	8.50% (-39%)	8.52% (-39%)	7.32% (-40%)
Currency	0.48% (-3%)	0.24% (-3%)	0.96% (-3%)	0.00% (-4%)
City-in-state	2.82% (-21%)	3.23% (-20%)	3.26% (-20%)	3.13% (-20%)
Regione capoluogo	5.56% (-18%)	7.31% (-16%)	8.19% (-15%)	9.36% (-14%)
Family	31.58% (-36%)	28.95% (-39%)	34.50% (-33%)	28.36% (-40%)
Semantic accuracy	7.13% (-31%)	8.20% (-30%)	8.33% (-30%)	7.67% (-31%)
Adjective-to-adverb	4.52% (-13%)	5.81% (-12%)	6.67% (-11%)	7.20% (-11%)
Opposite	4.94% (-23%)	3.75% (-24%)	5.34% (-22%)	3.75% (-24%)
Comparative	8.33% (0%)	8.33% (0%)	8.33% (0%)	8.33% (0%)
Superlative (assoluto)	16.48% (-32%)	19.46% (-29%)	24.54% (-23%)	22.09% (-26%)
Present-participle (gerundio)	37.40% (-40%)	47.68% (-30%)	59.78% (-18%)	53.83% (-24%)
Nationality-adjective	19.70% (-58%)	15.20% (-62%)	26.64% (-51%)	16.26% (-61%)
Past-tense	29.94% (-31%)	34.07% (-27%)	39.58% (-22%)	34.85% (-27%)
Plural	8.17% (-46%)	9.05% (-45%)	10.24% (-44%)	8.41% (-46%)
Plural-verbs (3rd person)	68.82% (-17%)	70.01% (-16%)	73.80% (-12%)	66.67% (-19%)
Plural-verbs (1st person)	12.55% (-33%)	17.10% (-28%)	21.56% (-23%)	16.85% (-28%)
Remote-past-verbs (1st person)	7.05% (-27%)	13.46% (-21%)	13.46% (-21%)	11.54% (-23%)
Noun-masculine-feminine-singular	12.37% (-45%)	27.89% (-30%)	31.05% (-27%)	31.58% (-26%)
Noun-masculine-feminine-plural	3.31% (-26%)	8.46% (-21%)	14.58% (-14%)	7.50% (-22%)
Syntactic accuracy	17.85% (-37%)	19.98% (-35%)	24.25% (-30%)	20.66% (-34%)
<b>Total accuracy</b>	<b>15.78%</b> (-32%)	<b>17.70%</b> (-30%)	<b>21.05%</b> (-26%)	<b>18.05%</b> (-29%)

Tabella 2, Valutazione dei 4 modelli creati, distinti secondo le due modalità di allineamento utilizzate e i diversi periodi temporali di riferimento.

Rispetto al modello sopracitato di Berardi G. et al., i nostri modelli ottengono allo stesso modo bassi risultati, presumibilmente per via della scarsità di dati testuali con cui sono stati costruiti. In Tabella 3 si può infatti notare come le dimensioni dei corpora utilizzati in questo lavoro siano molto ridotte se paragonate al corpus BOOKS, un fatto che dipende

<sup>66</sup> Con l'architettura *Skip-gram*.

<sup>67</sup> Berardi G. et al., *Word Embeddings Go to Italy: A Comparison of Models and Training Datasets*, cit.

per gran parte anche dall'effettiva disponibilità, in formato elettronico, di dati testuali diacronici per la lingua italiana.

	Frasi	Parole	Parole uniche
BOOKS	232.777.927	2.222.726.367	9.443.100
$A(1994-1997) + B(2019-2021)$	6.741.443	157.298.288	765.139

Tabella 3, Dimensioni di BOOKS e dei corpora raccolti dell'archivio digitale di Repubblica per i periodi 1994-1997 e 2019-2021.

### 3.4 Individuare le parole mutate semanticamente

Una volta creati gli spazi semantici diacronici da confrontare, sono state individuate quelle parole che hanno subito un maggiore mutamento semantico, secondo le indicazioni in 2.3. In particolare, considerando una parola  $w$ , si è dapprima ottenuto il “vettore delle distanze”, calcolato tra  $w$  e un insieme arbitrario  $P$  di parole, per confrontare in seguito la discrepanza tra il “vettore delle distanze” del primo periodo, e il “vettore delle distanze” del secondo periodo. Più nel dettaglio, prendendo in considerazione un insieme  $P_n$  di parole, di  $n$  elementi, si è calcolato il valore di  $MSE$  (3) rispetto a: i) la distanza (2) calcolata nello spazio  $S_{A(1994-1997)}$  tra  $w$  e gli elementi di  $P$  ( $d_A$ ) e ii) la distanza (2) calcolata nello spazio  $S_{B(2019-2021)}$  tra  $w$  e gli elementi di  $P$  ( $d_B$ ). Applicando quindi la seguente formula:

$$MSE = \frac{\sum_{i=1}^n (d_{A_i} - d_{B_i})^2}{n} \quad (3)$$

Dove  $n$  è il numero di elementi di  $P$  e  $d_{A_i} - d_{B_i}$  è lo scarto delle distanze, calcolate in un caso tra gli spazi  $S_{Aincr.}$  e  $S_{Bincr.}$ , e in un altro tra gli spazi  $S_{Acomp.}$  e  $S_{Bcomp.}$ .

Nell'individuazione delle parole mutate sono stati utilizzati due insiemi  $P$  differenti: in un caso come l'insieme delle dieci parole più vicine a  $w$  nello spazio  $A(1994-1997)$  e in un altro come un insieme di 1843 parole, appartenenti al lessico fondamentale dell'italiano<sup>68</sup> secondo *Il Nuovo vocabolario di base della lingua italiana*<sup>69</sup>. L'insieme  $P$  risulta così formato dalle parole vicine in un caso - l'intorno di  $w$  nello spazio semantico

<sup>68</sup> Per Lessico fondamentale si intende quel vocabolario «formato da circa 2.000 vocaboli di massima frequenza, che “coprono” da soli in media il 90% di qualunque testo scritto o parlato» cfr. Casadei F., *Lessico e Semantica*, cit.

<sup>69</sup> De Mauro T., Chiari I., *Il Nuovo vocabolario di base della lingua italiana*, 2016, Internazionale, [20/10/2021], <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana>.

- e dalle parole del lessico fondamentale nell'altro. In questo modo si avrà mutamento semantico quando cambiano i rapporti di similarità tra una parola  $w$  e il suo intorno o tra  $w$  e il lessico fondamentale.

L'intorno e il lessico fondamentale vanno a configurarsi così come i punti di riferimento entro cui si misura il mutamento, assumendo che l'intorno sia formato da parole che hanno una effettiva relazione semantica con la parola  $w$ , di modo che possano fornire informazioni sul suo significato. Tuttavia, in questo bisogna essere cauti poiché capita che le parole vicine riflettano invece contesti molto specifici, che non indicano necessariamente una vicinanza di significato, e che vengano registrate come simili solo per l'alto grado di associazione tra i loro contesti, come nel caso sopracitato dell'associazione *cancelliera* – *Merkel*. Per quanto riguarda la seconda metodologia, d'altra parte, si assume che l'insieme delle parole del lessico fondamentale sia formato da parole uniformemente distribuite nello spazio, in modo da poter rappresentare, seppur approssimativamente, lo spazio semantico stesso. Di seguito (Tabella 4) per ogni metodologia di allineamento sono riportate due liste delle parole<sup>70</sup> cambiate maggiormente secondo l'indice di *MSE*, tra le parole con frequenza maggiore di 30 in entrambi i corpora.

---

<sup>70</sup> Selezionate da un dizionario di 660.000 parole cfr. Napoletano F., *paroleitaliane*, [20/10/2021], <https://github.com/napolux/paroleitaliane>.

Allineamento tramite <i>incremental update</i>							
<i>P</i> formato dalle 10 parole più vicine a <i>w</i>				<i>P</i> formato dal Lessico Fondamentale			
<i>MSE</i>	<i>w</i>	$f_A$ ('94-'97)	$f_B$ ('19-'21)	<i>MSE</i>	<i>w</i>	$f_A$ ('94-'97)	$f_B$ ('19-'21)
0.58	disuso	138	197	0.12	ansa	34	1547
0.58	euro	1001	61018	0.05	euro	1001	61018
0.54	ansa	34	1547	0.03	dismisura	265	207
0.53	ritroso	178	148	0.03	ingrandire	33	308
0.49	dismisura	265	207	0.03	portale	58	2234
0.47	rileto	141	248	0.03	isolamento	1369	19567
0.44	galla	552	358	0.03	malate	145	1850
0.40	oltremanica	32	120	0.03	inciampo	51	283
0.39	community	56	591	0.03	soquadro	140	89
0.38	ridosso	878	1208	0.03	step	37	709
0.37	priori	371	152	0.03	fotogramma	152	699
0.37	pin	53	87	0.03	intensive	39	3865
0.32	palio	538	955	0.03	vicinato	99	444
0.31	intensive	39	3865	0.03	autonomia	193	123
0.30	osta	265	211	0.03	core	415	233
0.29	vicinato	99	444	0.03	assembramento	37	1719
0.28	guariti	46	16636	0.03	comprensivo	199	1182
0.28	procinto	294	254	0.03	arancione	206	4714
0.28	ripartenza	32	4475	0.03	indeterminato	392	804
Allineamento iniziando la posizione complessiva							
<i>P</i> formato dalle 10 parole più vicine a <i>w</i>				<i>P</i> formato dal Lessico Fondamentale			
<i>MSE</i>	<i>w</i>	$f_A$ ('94-'97)	$f_B$ ('19-'21)	<i>MSE</i>	<i>w</i>	$f_A$ ('94-'97)	$f_B$ ('19-'21)
0.37	arrendo	43	46	0.04	euro	1001	61018
0.28	euro	1001	61018	0.04	ansa	34	1547
0.25	contempo	602	1079	0.02	isolamento	1369	19567
0.18	post	2716	14434	0.02	paciere	42	35
0.14	protrae	71	85	0.02	capri	63	39
0.12	dimette	281	243	0.01	on	2178	5957
0.12	popolari	4419	1569	0.01	fotogramma	152	699
0.11	isolamento	1369	19567	0.01	vergogno	188	168
0.10	positivi	2273	44904	0.01	contempo	602	1079
0.10	impegneranno	49	36	0.01	logicamente	68	52
0.10	recheranno	59	62	0.01	positivo	5548	15235
0.10	attivo	3010	3547	0.01	perché	2146	525
0.10	cinofile	39	132	0.01	provare	2177	3478
0.10	marchi	3284	1222	0.01	applicativo	52	65
0.09	facenti	114	132	0.01	capezzale	161	46
0.09	osta	265	211	0.01	governance	102	1032
0.09	capri	63	39	0.01	contrasto	2423	3421
0.08	novantuno	3845	2766	0.01	hip	200	238
0.08	contrasto	2423	3421	0.01	inciampo	51	283

Tabella 4, Liste delle parole mutate maggiormente: a sinistra l'insieme *P* è formato dalle 10 parole più vicine a *w*, a destra dalle parole del Lessico fondamentale italiano.

Si può notare come l'impiego del lessico fondamentale riduca notevolmente la presenza di forme flesse, suscettibili a problemi di contestualità e bassa frequenza. Se una parola compare poche volte nel corpus, infatti, le sue parole vicine spesso saranno parole connesse alla situazione a cui la parola di partenza fa riferimento, più che parole con un significato effettivamente simile. Questi risultati saranno esaminati più nel dettaglio in 3.6, si rinvia quindi a questa sezione la discussione sulla loro attendibilità.

### 3.5 *Rappresentare il mutamento semantico*

Per comprendere meglio il cambiamento dei rapporti di similarità è possibile visualizzare il mutamento delle distanze in un piano cartesiano. Se, in periodi diversi, la distanza tra due parole risulta mutata vorrà dire che sarà avvenuto un cambiamento nel significato dell'una o dell'altra parola. In questo modo si può rappresentare il mutamento semantico di una parola rispetto alle altre parole in termini del mutamento delle loro relazioni di similarità semantica.

Per visualizzare il mutamento semantico di una parola  $w$ , dato un insieme  $P$  di parole, si possono prendere in esame le distanze<sup>71</sup>  $d(w, w_p)_A$  e  $d(w, w_p)_B$ <sup>72</sup>, inserendole come punti in un piano cartesiano con dimensioni  $(d_B; d_A)$ . Dove ogni punto rappresenta la parola  $w_p$  cui è misurata la distanza da  $w$  e, più precisamente, la posizione della distanza  $d(w, w_p)$  tra gli spazi che rappresentano  $A(1994-1997)$  e  $B(2019-2021)$ . L'origine si identifica così con la posizione di  $w$ , interpretabile come quella parola che ha una distanza  $d_A$  e  $d_B$  pari a 0 da sé stessa.

In base alla posizione delle parole  $w_p$  si possono avere principalmente due situazioni. Se delle parole si dispongono lungo la bisettrice (dove  $d_A = d_B$ ; con un coefficiente angolare  $m$  - della retta che passa dall'origine alla parola data - prossimo a 1), allora il significato di  $w$  che quelle parole indicano - in virtù della relazione di similarità - non cambia tra lo spazio  $S_{A(1994-1997)}$  e lo spazio  $S_{B(2019-2021)}$ . Se invece delle parole si dispongono in prossimità di un asse, allora il significato di  $w$  che quelle parole indicano sarà peculiare dello spazio  $S_{A(1994-1997)}$  o dello spazio  $S_{B(2019-2021)}$ , a seconda dell'asse cui sono vicine (un  $m$  molto basso denota un significato peculiare a  $S_{B(2019-2021)}$ , mentre al contrario, peculiare a  $S_{A(1994-1997)}$  per valori molto alti). Potendo realizzare

---

<sup>71</sup> Calcolate in questo caso come  $1 - \text{CosineSimilarity}(w, w_p)$ .

<sup>72</sup> Dove  $w_p \in P$ .



grafici di questo tipo, è possibile visualizzare le parole che tra i due corpora sono caratterizzate da un marcato mutamento semantico e interpretare quali siano quei significati che cambiano tra il 1994 e il 2021.

Considerando la parola *portale*, in Figura 3, si può notare come le sue parole vicine in  $S_{A(1994-1997)}$  e in  $S_{B(2019-2021)}$  denotino il significato che essa assume nel 1994-1997 piuttosto che nel 2019-2021. Ad esempio, nello spazio  $S_{B(2019-2021)}$ , *portale* risulta molto vicina a *sito*, mentre nello spazio  $S_{A(1994-1997)}$ , risulta vicina a parole come *portico*, *chiostro* e *cortile*, indicando un significato di portale prossimo all’ambito dell’architettura, così come quello delle sue parole vicine.

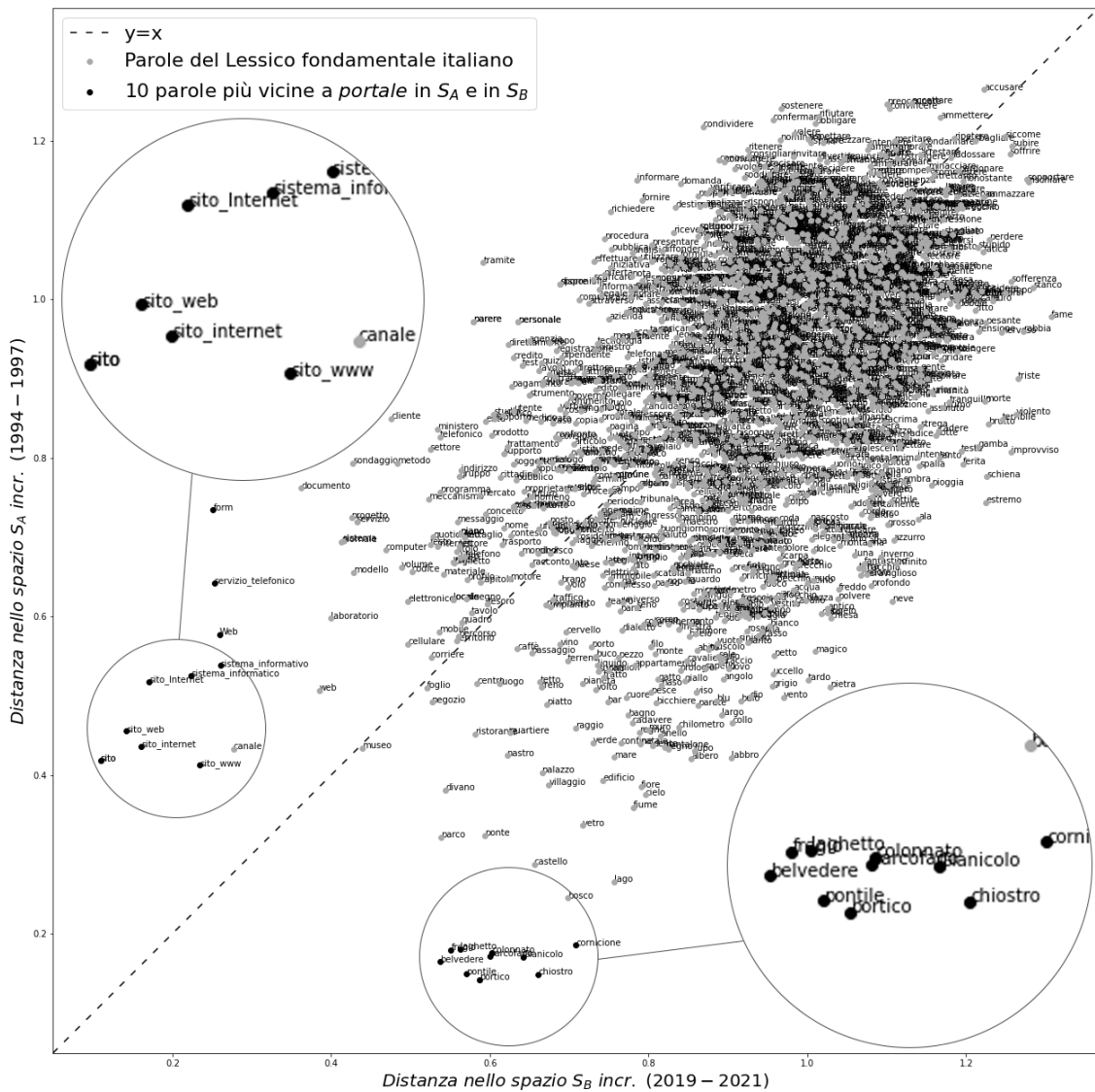


Figura 3, Rappresentazione del mutamento dei rapporti di similarità della parola *portale*, tra gli spazi  $S_{Aincr.}$  e  $S_{Bincr.}$ .

Nel caso del rapporto tra *portale* e *sito*<sup>73</sup> si può dire che essi siano - nel periodo recente - sinonimi, mentre lo stesso non si può dire - nel periodo anteriore - per *portico*, *chiosstro* e *cortile*<sup>74</sup>, che, più che esser sinonimi, sono parole che si riferiscono allo stesso ambito concettuale, che nei capitoli precedenti abbiamo definito essere corrispondente al campo semantico. Come è infatti già stato accennato in 2.1, la similarità semantica può codificare una vasta e indistinta gamma di relazioni semantiche, configurandosi in questo senso come uno dei limiti principali di una semantica distribuzionale, anche per quanto riguarda l'individuazione del mutamento semantico, come si vedrà in 3.6.

Infine, per quanto riguarda la rappresentazione del mutamento, si noti come, unitamente a una diversa conformazione delle parole vicine, cambi anche la conformazione delle parole del lessico fondamentale. Questo vuol dire che in una semantica distribuzionale il mutamento non è un fenomeno che riguarda solo l'intorno delle parole, ma anche molte delle parole dello spazio semantico - rappresentate dal lessico fondamentale. Si osservi quindi la conformazione delle parole del lessico fondamentale per parole che si suppone abbiano mantenuto inalterato il loro significato, come nel caso riportato di seguito (Figura 4) della parola *uomo*.

---

<sup>73</sup> La similarità semantica tra *portale* e *sito*, nello spazio del periodo 2019-2021, è pari a 0,89.

<sup>74</sup> La similarità semantica tra *portale* e *portico*, *chiosstro* e *cortile*, nello spazio del periodo 1994-1997 è pari rispettivamente a 0,86, 0,85 e 0,77.

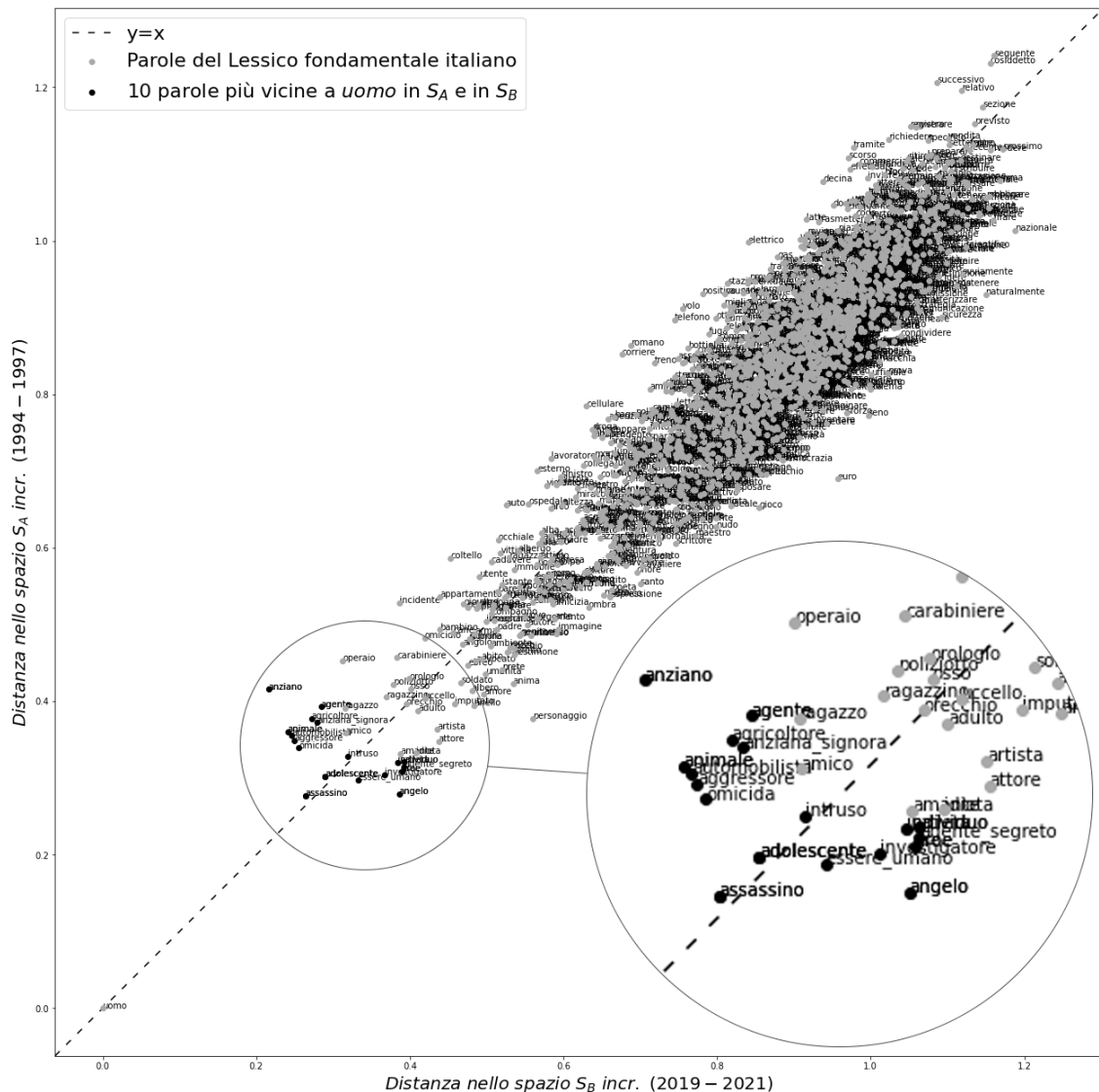


Figura 4, Rappresentazione del mutamento dei rapporti di similarità della parola uomo, tra gli spazi  $S_{A\text{incr.}}$  e  $S_{B\text{incr.}}$ .

### 3.6 Una semantica diacronica distribuzionale

Per quale motivo la parola *arancione* viene percepita come mutata dal modello distribuzionale insieme a *portale*, *ripartenza*, *ansa*, *vicinato* e *soqqadro*? Non sembra infatti che tutte le parole considerate mutate dal modello abbiano oggi un significato realmente diverso da quello che avevano negli anni Novanta. Inoltre, le parole che verosimilmente sono mutate, come *movida*, *ripartenza* e *portale*, oltre a non aver subito mutamento semantico allo stesso modo, agli occhi del modello non appaiono mutate di molto rispetto ad altre parole il cui significato non è cambiato affatto. Tutto ciò è dovuto ad una scorretta codifica nell'estensione dei significati tra un periodo e l'altro, intendendo

l'estensione di un significato come la gamma degli oggetti a cui quel significato si può applicare.

Come è stato già ripetuto, la qualità della codifica del significato di una parola è legata alla bontà dei dati sulla sua distribuzione linguistica, nella generazione di rapporti di similarità semantica da contesti simili. Se si considera che acquisire dati sulla distribuzione linguistica di una parola è approssimabile a registrare l'insieme degli oggetti a cui si può applicare il suo significato, un errore nella codifica della sua estensione comporta la codifica di un significato errato. Errori di questo tipo sono tanto più probabili quanto meno è specifico il significato di una parola, poiché parole con una grande estensione occorrono in contesti molto variegati, che possono essere registrati in un corpus in maniera parziale. Mentre, viceversa, è più facile cogliere l'estensione di un significato specifico, poiché la gamma dei contesti in cui può comparire è relativamente limitata. Per questo, quando i dati distribuzionali di una parola sono insufficienti a codificare l'estensione del suo significato, quest'ultimo avrà una codifica sbagliata.

Era stato osservato, a fine 2.3, che il mutamento dei contesti d'occorrenza di una parola può non corrispondere al mutamento del suo significato. Qui si aggiunga che il giudizio del modello riguardo il mutamento di significato di una parola, oltre a poter essere causato da una scorretta codifica del suo significato, può essere dovuto anche a una differente codifica dell'estensione del suo significato tra spazi semantici diversi. Caso in cui il significato può apparire mutato per due ragioni: o perché codificato da due estensioni parziali diverse o perché effettivamente mutato, dunque avente estensioni realmente differenti.

Per questo motivo, nel valutare quanto bene una semantica distribuzionale riesca a codificare il mutamento semantico, si discrimineranno due situazioni fondamentali: i) quando sia avvenuta una codifica incompleta dell'estensione di un significato tra i due periodi; ii) quando l'estensione sia stata codificata correttamente e una parola appaia di conseguenza mutata per motivi più o meno semanticamente giustificabili. Procedendo in questo modo si può modificare la tabella precedente discriminando queste due situazioni, come illustrato in Tabella 5, dove, rispetto agli spazi  $S_A^{incr.}$  e  $S_B^{incr.}$ , sono riportate in grassetto le parole che si possono ritenere mutate e, in corsivo, le parole che sono considerate mutate per una codifica parziale dell'estensione del loro significato tra uno

spazio e l'altro. Invece, il significato delle parole che non sono riportate né in corsivo né in grassetto non è stato codificato correttamente in entrambi gli spazi, fatta eccezione per *ansa*, che subisce una sovrapposizione di entità diverse, dove *ansa*, “curvatura, insenatura”, viene confusa invece con l'agenzia *ANSA*. Inoltre, sono riportate anche le due parole  $w_t$  più vicine  $w$  nei due spazi semantici.

<i>MSE</i>	$w$	$f_A$	$f_B$	Prime due $w_t$ in $S_A^{incr}$ .	Prime due $w_t$ in $S_B^{incr}$ .
0.12	<i>ansa</i>	34	1547	isolotto, isoletta	BARCELLONA, BERLINO
0.05	<b>euro</b>	1001	61018	Euro, unione monetaria	lire, sterline
0.03	dismisura	265	207	disumanità, solarità	ritmo sostenuto, vertiginosamente
0.03	ingrandire	33	308	editare, abbellire	riquadro, farfallomania
0.03	<b>portale</b>	58	2234	portico, chiostro	sito, sito web
0.03	isolamento	1369	19567	abbandono, indigenza	quarantena, autoisolamento
0.03	malate	145	1850	portatrici, primitive	nelle Marche, cinquecentottantatre
0.03	inciampo	51	283	errore strategico, povero cristo	foibe, Vele
0.03	soquadro	140	89	colpii, agevolissimo	disposizione_gratuitamente, a soquadro
0.03	step	37	709	funky, swing	standard, aspetti_fondamentali
0.03	fotogramma	152	699	brano, canovaccio	ARTICOLO, pezzo intitolato
0.03	intensive	39	3865	ginecologiche, I20	nei reparti, reparti ospedalieri
0.03	vicinato	99	444	corpo carbonizzato, brusio	alberghi ristoranti, bar ristoranti
0.03	autonomia	193	123	OSSESSIVO, ipotizzarlo	intenderci, sonia
0.03	core	415	233	Caminito, Casanova	Set, Genius
0.03	<i>assembramento</i>	37	1719	accampamento, andirivieni	affollamento, assembramenti
0.03	comprensivo	199	1182	oneroso, equo	Comprensivo, Giannina Gaslini
0.03	<b>arancione</b>	206	4714	color, marrone	gialla, arancioni
0.03	indeterminato	392	804	indefinito, bovide	full time, lavorativo

Tabella 5, Parole con frequenza maggiore di 30 cambiate di più secondo l'indice MSE, rispetto alla modalità di allineamento tramite incremental update e all'insieme di  $P$  formato dalle parole del lessico fondamentale italiano.

In penultima posizione è stata segnalata in grassetto la parola *arancione*: un caso di mutamento semantico per ellissi<sup>75</sup> in cui il contenuto dell'espressione *regione arancione* si è spostato nel contenitore *arancione*<sup>76</sup>. Questo è uno di quei casi per cui secondo la semantica diacronica tradizionale sarebbe avvenuto un mutamento di significato, essendo cambiata la relazione tra forma e contenuto, mentre lo stesso non varrebbe per la semantica diacronica strutturale o cognitiva. Per quanto riguarda invece la parola *euro*, si

<sup>75</sup> Cfr. Nota 9.

<sup>76</sup> Si noti che, con i nostri dati, *regione arancione*, in base alla funzione *models.phrases* dalla libreria *gensim.models.phrases*, non forma una collocazione.

può pensare che il suo significato sia mutato in virtù del cambiamento della sua estensione, che ha sostituito dal 2001 l'estensione del significato di *lira/lire*.

Si può assumere che, in una semantica distribuzionale, il significato di una parola, essendo determinato dalle relazioni di similarità con le altre parole, possa essere colto a partire dalle sue parole vicine, in quanto più rappresentative del suo significato. Si può osservare come in Tabella 5 le parole vicine a una certa parola possano essere inerenti o non inerenti al suo significato. Se non sono inerenti vorrà dire che il significato della parola non è stato codificato correttamente, e perciò non avrebbe senso operare un confronto del suo significato tra periodi diversi. Se, viceversa, sono inerenti al suo significato, esse possono esserlo per motivi più o meno semanticamente giustificabili. In particolare, non si può dire che ci sia una relazione semantica<sup>77</sup> tra parole che sono vicine a una certa parola in quanto referenti fisici della stessa, come nel caso della parola *cancelliera*, che nello spazio del periodo 2019-2021 ha un valore di similarità con *Merkel* pari a 0,76, risultando appunto *Merkel* la parola più vicina a *cancelliera*. Anche se esiste comunque un legame referenziale tra le due parole, sembra inefficace pensare che un tale cambiamento di referente possa codificare un mutamento di significato per la parola *cancelliera*.

Diversamente, si può dire che una parola sia cambiata quando siano avvenuti dei mutamenti semanticamente giustificabili nelle relazioni tra questa e le parole vicine che indicano il suo significato. Un caso di questo genere è rappresentato dalla parola *portale*, che nel periodo 1994-1997 è associata a parole dell'ambito dell'architettura come *portico*, *chiostro* e *cortile*, mentre nel periodo 2019-2021 a parole che fanno riferimento alla realtà di internet, venendo *portale* inteso come sinonimo di *sito* per anni successivi all'avvento della rete. Inoltre, si può osservare come, in un periodo, le parole vicine a *portale* codifichino l'appartenenza a uno stesso campo semantico mentre, in un altro, una relazione di sinonimia. La disomogeneità delle relazioni semantiche codificate da parole vicine è connessa alla natura della similarità semantica che, come detto appunto in 2.1, non corrisponde a una relazione semantica precisa, ma solo a un certo grado di somiglianza tra due significati.

---

<sup>77</sup> Se non di tipo referenziale.

Sulla base del caso di *portale* si può allora generalizzare affermando che, in un modello distribuzionale, una parola muta di significato quando sussista un cambiamento nelle relazioni semantiche che essa intrattiene con le sue parole vicine. Se perciò si verifica che una parola venga usata come sinonimo per un'altra, è verosimile pensare che il suo significato sia in qualche modo cambiato, avvicinandosi a quello del suo nuovo sinonimo. Proprio questo caso è stato colto dal modello per la parola *ripartenza*, che nello spazio del periodo più recente è fortemente associata a *ripresa*, con un valore di similarità di 0,89, mentre nel periodo 1994-1997 è associata a espressioni d'ambito calcistico.

Infine, da questi casi emerge che la similarità semantica comprende di fatto diversi fenomeni, che possono anche avere poco a che fare con il significato, il che si configura come un problema strutturale della semantica distribuzionale. Invece, per quanto riguarda il cambiamento di significato, al problema della similarità si aggiunge la necessità di codificare correttamente, e allo stesso modo, l'estensione di un significato in spazi diversi, cosa che altrimenti porta a una valutazione scorretta del suo mutamento. Al netto di queste problematiche, in conclusione, è possibile codificare felicemente il cambiamento di significato delle parole, anche se tuttora le metodologie di codifica dei significati risultano a volte inefficienti, inficiando l'individuazione del mutamento semantico, come si può osservare dalla grande quantità di falsi positivi negli elenchi di parole mutate riportati in questo testo.

## Conclusioni

Come le diverse semantiche diacroniche, illustrate in 1.1, individuano tante facce di uno stesso fenomeno, e sussistono senza escludersi a vicenda, così non si potrebbe dire che una semantica distribuzionale non possa cogliere in qualche maniera il mutamento di significato delle parole. Al netto dei problemi sopra illustrati, una semantica distribuzionale, infatti, considera mutamento semantico qualsiasi cambiamento nell'estensione di un significato. Quanto un mutamento dell'estensione sia un mutamento anche del significato è possibile valutarlo esaminando la giustificabilità semantica delle nuove relazioni di similarità che una parola intrattiene con le parole vicine e le altre parole dello spazio. Quello su cui la semantica distribuzionale deve interrogarsi, è il legame che intercorre tra i due fenomeni, poiché, se è vero che l'aggiunta di un oggetto come *Merkel* all'estensione di *cancelliera* non è rilevante ai fini di stabilire il mutamento di significato di *cancelliera*, verosimilmente l'aggiunta di un oggetto appartenente alla categoria degli animali lo sarebbe. Viceversa, un altro interrogativo da porsi è se esistano dei mutamenti semantici che non causino un cambiamento nell'estensione, tali cioè da non potere essere individuati dal modello.

Un fatto strutturale da tenere a mente è che, con le metodologie attuali, una semantica diacronica distribuzionale non sembra in grado di distinguere la qualità del mutamento, quanto unicamente il maggiore o minore grado di diversità di significato tra periodi diversi. Questo è chiaro nel caso citato in precedenza di *portale*, il cui significato per “sito” in periodi recenti si configura come un'aggiunta per uso metaforico<sup>78</sup> - e non una vera e propria modifica - rispetto al suo significato originario. In questo senso, inoltre, non si può dire che *portale* nel 2021 abbia perso il suo significato per “elemento architettonico”, cosa che tuttavia il modello non è in grado di stabilire. I mutamenti che una semantica distribuzionale può individuare sono tutti, e solo, quelli che comportano un mutamento nell'estensione di un significato. Al che possono sorgere dei dubbi riguardo le finalità di una semantica diacronica distribuzionale dal punto di vista del contributo linguistico apportato. Al momento, infatti, essa si limita a una semplice constatazione di quali parole abbiano subito un qualche cambiamento semantico, senza tuttavia essere in grado di spiegare o giustificare le cause del loro mutamento. Ad

---

<sup>78</sup> Cfr. nota 9.



esempio, sempre nel caso di *portale*, un modello distribuzionale non ha elementi per capire che il suo nuovo significato per *sito* è legato a un mutamento avvenuto dapprima in lingua inglese e, in un secondo momento, riprodotto in italiano sul calco di *portal* inglese<sup>79</sup>. Tuttavia, è possibile che una delle applicazioni dei modelli diacronici distribuzionali - oltre a quantificare il mutamento - consista nell'individuare con più precisione i limiti temporali di un mutamento semantico. Infatti, siccome questi modelli sono in grado di individuare, in condizioni ottimali, tanto il mutamento quanto l'assenza di mutamento, possono essere utili per la datazione dei nuovi significati, potendo stabilire quando un mutamento non si è verificato<sup>80</sup>.

Finora si sono presentate le problematiche di una semantica diacronica distribuzionale, è bene però discutere di alcuni limiti, sia linguistici che tecnici, della semantica distribuzionale nella sua versione generale, non applicata alla ricerca del mutamento semantico. Questo perché emerge abbastanza chiaramente che i problemi riscontrati nell'identificazione del mutamento semantico non siano dovuti alle metodologie d'individuazione. Infatti, il calcolo di *MSE* tra relazioni di similarità sembra molto promettente, anche per quanto riguarda il rapporto tra quantità di dati da misurare e sforzo computazionale. Le problematichità risiedono invece nella codifica distribuzionale del significato, che è affidata alla similarità semantica, la quale tuttavia - come è stato notato in 2.1 - tende a connettere «broadly related neighbors, rather than similar ones»<sup>81</sup>, oltre al fatto che il significato che essa assume dipende dalla tipologia di contesto con cui viene codificata. A questo si aggiungano i problemi ancora aperti riguardo la codifica di relazioni asimmetriche o la composizionalità dei significati, oppure dell'utilizzo di un tipo di contesto che non si limiti al co-testo lineare, ma che tenga conto di tutti quei legami che in un testo intercorrono tra le parole. A proposito si può notare come i limiti di una semantica distribuzionale risiedano anche in tutti quei fattori che essa esclude dall'analisi, come quelli attinenti al contesto extralinguistico: la referenza, gli usi pragmatici, quelli metaforici e, in generale, tutti i prodotti legati all'oralità.

---

<sup>79</sup> De Mauro T., Mancini M., *Garzanti etimologico: i grandi dizionari*. Garzanti linguistica., 2000.

<sup>80</sup> Questo sempre tenendo in considerazione l'interrogativo precedente, ovvero se esistano dei mutamenti di significato che non cambino l'estensione del significato, senza essere così individuati dal modello.

<sup>81</sup> Lenci A., *Distributional Models of Word Meaning*, cit., p. 161.

Dal punto di vista tecnico, invece, sono altrettanti i fattori che garantiscono un'adeguata codifica distribuzionale del significato. In primo luogo, la quantità dei dati linguistici deve essere sufficiente per poter approssimare la somma dei contesti di una parola all'estensione del suo significato. Si può avere un'idea di quanto sia questa quantità prendendo in considerazione il lavoro di Berardi G. et al., dove un modello distribuzionale basato su 2.222.726.367 parole ha risolto correttamente il 47.35% delle analogie presentate - il che porta a pensare che per codificare adeguatamente il significato, risolvendo correttamente le analogie, siano necessari numeri ancora superiori. Dopodiché la codifica del significato dipende anche dalla qualità dei dati linguistici, nella loro capacità di fornire dati di contesto rilevanti, e da caratteristiche del modello, come la lunghezza della finestra di contesto o la dimensione del vettore di posizione<sup>82</sup>.

La fiducia in futuri lavori di questo tipo è che la maggiore quantità di dati possa sopperire alle problematiche che abbiamo riscontrato nell'ultimo capitolo. Sarebbe interessante, infatti, confrontare spazi semantici supportati da corpora più grandi, il che probabilmente indirizza questo tipo d'analisi verso la lingua inglese, che dispone al momento della più grande quantità di dati testuali in formato elettronico al mondo. In definitiva, quindi, i risultati illustrati in questo lavoro vanno valutati nei limiti delle prestazioni dei modelli che abbiamo realizzato, poiché è chiaro che una miglioria nelle prestazioni porterebbe a una migliore codifica del mutamento semantico. Anche se, per un fatto linguistico, si verifica che il *word embedding* funzioni meglio per parole dal significato specifico, poiché riducendosi l'estensione aumenta la probabilità di registrare un numero soddisfacente di contesti per codificarla correttamente. Il che potrebbe in parte spiegare perché i casi più apprezzabili di mutamento di significato che abbiamo individuato siano significati abbastanza specifici, come *portale*, *ripartenza* e *arancione*<sup>83</sup>.

---

<sup>82</sup> Cfr. nota 29.

<sup>83</sup> Cfr. p. 31.

## Bibliografia

Arseniev-Koehler A., Foster J. G., *Sociolinguistic Properties of Word Embeddings*, 2020, SocArXiv doi:10.31235/osf.io/b8kud.

Bamler R., Mandt S., *Dynamic word embeddings* in *International conference on Machine learning*, Sydney, PMLR, 2017, pp. 380-389.

Benotto G., *Modelli distribuzionali delle relazioni semantiche: il caso dell'iperonimia* in *Animali, Umani, Macchine. Atti del convegno 2012 del CODISCO*, Roma-Messina, Corisco, 2013, pp. 85-97.

Berardi G., Esuli A., Marcheggiani D., *Word Embeddings Go to Italy: A Comparison of Models and Training Datasets*, in *6th Italian Information Retrieval Workshop*, Cagliari, 2015.

Berruto G., Cerruti M. S., *La linguistica. Un corso introduttivo*, UTET De Agostini, 2017.

Bojanowski P., Grave E., Joulin A., Mikolov, T., *Enriching word vectors with subword information* in «Transactions of the Association for Computational Linguistics», 5 (2017), pp. 135-146.

Casadei F., *Lessico e Semantica*, Roma, Carocci, 2003.

Coseriu E., *Teoria del linguaggio e linguistica generale*, Bari, Laterza, 1971, pp. 225-286.

De Mauro T., Mancini M., *Garzanti etimologico: i grandi dizionari*. Garzanti linguistica., 2000.

De Mauro T., Chiari I., *Il Nuovo vocabolario di base della lingua italiana*, 2016, Internazionale, [20/10/2021], <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana>.

Eger S., Mehler A., *On the linearity of semantic change: Investigating meaning variation via dynamic graph models*, 2017, arXiv preprint arXiv:1704.02497.

Galke L., Saleh A., Scherp, A., *Word embeddings for practical information retrieval* in «INFORMATIK», P275 (2017).

Györi G., *Semantic Change as Linguistic Interpretation of The World*, John Benjamins Publishing Company, 2000.

Hamilton W. L., Leskovec J., Jurafsky D., *Diachronic word embeddings reveal statistical laws of semantic change*, 2016, arXiv preprint arXiv:1605.09096.

Jurafsky D., Martin J. H., *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Draft of December 30, 2020.

Kim Y. et al., *Temporal Analysis of Language through Neural Language Models*, 2014, arXiv preprint arXiv:1405.3515.

Kutuzov A. et al., *Diachronic Word Embeddings and Semantic Shifts: a Survey*, 2018, arXiv preprint arXiv:1806.03537.

Labov W., *Il continuo e il discreto nel linguaggio*, Bologna: Il Mulino, 1977.

Lakoff G., Johnson M., *Metaphors We Live By*, Bompiani, 1980.

Lenci A., *Distributional semantics in linguistic and cognitive research* in «Italian journal of linguistics», 20(1) (2008), pp. 1-31.

Lenci A., *Modelli distribuzionali del lessico: metodi computazionali per l'analisi semantica* in «Informatica Umanistica», 3 (2010), pp. 57–69.

Lenci A., Montemagni S., Pirelli V., *Testo e computer. Elementi di linguistica computazionale*, Roma, Carocci, 2016.

Lenci A., *Distributional Models of Word Meaning*, in «Annual Review of Linguistics», 4 (2018), pp. 151-171.

Levin B., *English verb classes and alternations: A preliminary investigation*, Chicago, University of Chicago press, 1993.

Levy O., Goldberg Y., *Dependency-based word embeddings* in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Baltimora, Association for Computational Linguistics, 2014, pp. 302-308.

Lowe W., *Towards a theory of semantic space*, in «Proceedings of the 23<sup>rd</sup> Annual Conference of the Cognitive Science Society», 23 (2001), pp. 576-581.

Mikolov T., Chen K., Corrado G., Dean J., *Efficient estimation of word representations in vector space* in *Proceedings of Workshop at International Conference on Learning Representations*, Scottsdale, US, ICLR, 2013.

Padó S., Lapata M., *Dependency-based construction of semantic space models*, in «Computational Linguistics» 33/2 (2007), pp. 161-199.

Peng H. et al., *Incrementally Learning the Hierarchical Softmax Function for Neural Language Models*, in: «Proceedings of the AAAI Conference on Artificial Intelligence», 31 (2017), pp. 3267–3273.

Pennington J., Socher R., Manning C.D., *Glove: Global vectors for word representation* in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, QA, EMNLP, 2014, pp. 1532–1543.

Petricca P., *Semantica. Forme, Modelli e Problemi*, Milano, LED Edizioni Universitarie, 2019, pp. 29-33.

Raymond W. Gibbs Jr., *The Cambridge Handbook of Metaphor and Thought*, Cambridge University Press, 2008, pp. 2-28.

Tiberii P., *Dizionario delle collocazioni*, Bologna, Zanichelli, 2012.

Ullmann S., *Principi di Semantica*, traduzione di M. Mayer Modena e A. M. Finoli, Einaudi, 1977.

Wang S. I., Manning C. D., *Baselines and bigrams: Simple, good sentiment and topic classification* in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Jeju Island, Association for Computational Linguistics, 2012, pp. 90-94.

Wang Z. J., Liu Y., Wang Z. J., *E-mail filtration and classification based on variable weights of the Bayesian algorithm*. In «Applied Mechanics and Materials», 513 (2014), pp. 2111-2114.

Yao Z. et al., *Dynamic word embeddings for evolving semantic discovery*, in: *Proceedings of the eleventh acm international conference on web search and data mining*, 2018, pp. 673–681.