# Descrizione integrale dell’annotazione del LocoMIN Corpus

Gloria Comandini

LocoMIN Corpus © 2025 by Gloria Comandini is licensed under CC BY-NC-SA 4.0. To view a copy of this license, visit:

https://creativecommons.org/licenses/by-nc-sa/4.0/

Se usate il LocoMIN Corpus per la vostra ricerca, ricordate di citare:

Comandini, G. (2025), «Frena la locomotiva d’Europa»: il ‘linguaggio della ferrovia’ per raccontare la crisi economica tedesca. Analisi linguistica sul corpus LocoMIN, «Studi Germanici» 27, pp. 127-182.

L’indagine è stata condotta su tre database di articoli, formando tre sotto-corpora atti ad analizzare il fenomeno in ottica diacronica.

Il primo database è basato sul lavoro dell’agenzia Infojuice, la quale ha operato una scrematura preliminare, isolando tutti gli articoli di quotidiani italiani che fanno riferimento alla Germania. Infojuice ha raccolto 59.929 articoli italiani cartacei, pubblicati fra il 1° gennaio 2021 e il 31 agosto 2023, nei quali compaiono il termine *Germania* o altre parole chiave utilizzate da Infojuice per selezionare gli articoli (*Scholz*, *Merkel*, *Berlino*, *Baviera*, ecc.). Fra questi, ne sono stati isolati 203 in cui si nomina il termine *locomotiva* usato in maniera figurata[[1]](#footnote-1), formando così un primo sotto-corpus di riferimento, LocoIJ21-23, composto di 159.761 token.

Il secondo database è il corpus *La Repubblica[[2]](#footnote-2)* (circa 380 milioni di token), formato da articoli tratti dal quotidiano *La Repubblica* pubblicati fra il 1985 e il 2000 (Baroni et al., 2004). Da questo database sono stati isolati 220 articoli in cui compare il termine *locomotiva* in senso metaforico, di cui 108 pubblicati fra il 1985 e il 1986, e 112 pubblicati nel biennio 1995-96; questi 220 articoli formano il secondo sotto-corpus di riferimento, LocoRep. Per la struttura del corpus *La Repubblica*, non è stato possibile estrarre l’intero testo di ogni articolo, ma solo le frasi nelle immediatamente precedenti e successive alla comparsa del termine *locomotiva*. Pertanto, LocoRep è il sotto-corpus più piccolo, composto di 16.039 token.

Il terzo database è l’archivio storico de *La Stampa*, che contiene gli articoli pubblicati dal giornale dal 1867 al 2006. Da questo database sono stati isolati 175 articoli in cui compare, sempre in senso metaforico, il termine *locomotiva*, di cui 78 pubblicati nel biennio 1985-86 e 97 pubblicati fra il 1995 e il 1996. Questi 175 articoli formano il terzo e ultimo sotto-corpus di riferimento, LocoLS, composto di 560.294 token.

In totale, sono stati raccolti 598 articoli, di cui 186 pubblicati nel biennio 1985-86, 209 pubblicati nel biennio 1995-96 e 203 pubblicati fra il 2021 e il 2023. I 598 articoli totali formano il LocoMIN (*Locomotive as Metaphor in Italian Newspapers*) Corpus[[3]](#footnote-3), con 736.134 token e 52.239 types.

Gli articoli di LocoMIN sono stati annotati in due categorie aperte. La prima è il **target** a cui è associato il termine *locomotiva*, ossia il suo referente, che può essere un’entità singola (es: Italia, Germania, Lombardia, export italiano, ecc.) o una pluralità di entità (es: Francia e Germania; Giappone e Germania; Stet, Telecom e TIM, ecc.), così come ci si può riferire a entità astratte (es: l’esportazione italiana), a Stati o aree geografiche (es: Germania, Lombardia, Nord Italia, ecc.), ad aziende o a realtà collettive (es: un partito politico), a singole persone (es: Silvio Berlusconi, Pippo Baudo, Vladimir Smirnov, ecc.) e persino a specifici oggetti materiali (es: il Prosecco, il tartufo, ecc.). Nel caso in cui il referente fosse una persona specifica, l’annotazione del target indica in maniera generica la sua professione (es: politico, sportivo, attore, ecc.)[[4]](#footnote-4). Invece, nei casi in cui fossero rilevabili più target si è deciso di annotare usando la formula *Vari: [target 1], [target 2], [target 3]* (es: Vari: Germania, Francia), con i singoli target posti in ordine di apparizione nell’articolo. Infine, nei casi in cui non fosse possibile trovare un referente di locomotiva, poiché il termine è utilizzato in maniera generica, si è stabilito di attribuire l’etichetta, appunto, *Generico*.

La seconda concerne la presenza o l’assenza di un **paragone** fra due o più entità (es: Francia-Germania, USA-UE, Nord Italia-Sud Italia, ecc.), riportati in ordine di comparsa. Le linee guida per l’annotazione specificano che il paragone può coinvolgere il target di *locomotiva* (es: “A fine 2022 la locomotiva americana, dice l'Fmi, registrerà un Pil superiore di 6 punti a quello del 2019, il convoglio europeo soltanto 1 punto in più”), oppure no (es: “Un abisso tra Nord e Sud. È il Triveneto la nuova locomotiva”); si è deciso di rendere l’annotazione del paragone indipendente dal target per poter cogliere anche i casi in cui quest’ultimo avesse una relazione indiretta col paragone. Durante l’annotazione, si è specificato che in questa categoria sarebbero ricaduti non solo i paragoni espressi in maniera esplicita, ma anche quelli veicolati in maniera implicita. Qualora invece nel testo non si delineassero paragoni, si è scelta l’etichetta *No*.

Infine, gli articoli di LocoMIN sono stati annotati anche in due categorie chiuse: la macro-tematica in cui è usata la metafora e la polarità dell’articolo nei confronti del target nel contesto in cui viene utilizzato il termine *locomotiva*. Con **macro-tematica** si intende l’argomento generale trattato dal testo, la cui scelta doveva ricadere fra cinque opzioni possibili:

1. Economia, a cui sono ascrivibili tematiche quali crisi o successi economici, oscillazioni del PIL, guadagni, fatturazione e andamento dell’industria (es: “L'Italimpianti, che ha ottenuto importanti commesse di lavoro dall'estero, può diventare la «locomotiva» delle industrie metalmeccaniche”);
2. Politica, che include i testi che parlano di elezioni, partiti politici, riforme, attività legislative et similia (es: “«Agnelli ha ragione: Dilli può essere il premier di una maggioranza di centro-destra dopo le elezioni, con Berlusconi nel ruolo chiave di ‘locomotiva’ politica del polo»”);
3. Sport, comprensiva degli articoli che parlano, appunto, di sport, in tutte le sue forme (es: “Fra questi c'è «locomotiva», Tellini, il generoso difensore che fu già uno del punti di forza del Rapallo nei campionati passati.”);
4. Spettacolo, che include tutti i testi relativi a cinema, televisione, teatro e simili (es: “‘Una stupenda locomotiva’ così Patroni Griffi definisce l'interprete del suo nuovo spettacolo, Valeria Marini, che stasera debutta a Fabriano in Nata ieri, la commedia di Garson Kanin scritta nel 1945”);
5. Altro, in cui sono inseriti tutti i testi che non possono essere inclusi nelle macro-tematiche precedenti, come nel caso di articoli relativi a sanità, istruzione o cucina (es: “«La locomotiva sui vaccini adesso è la Lombardia. Roma? Cerchino un altro»”).

Questa annotazione deve confrontarsi con un ovvio ostacolo: le macro-tematiche scelte non possono essere considerate degli ambienti a chiusura stagna; dunque, casi in cui un articolo possa presentare elementi appartenenti a più di una macro-tematica sono inevitabili. Tuttavia, per semplificare l’analisi si è deciso che ogni articolo potesse essere inserito in una sola macro-categoria; in caso di co-presenza di più di una macro-categoria, si è dato istruzioni di scegliere quella maggiormente presente in termini di spazio e/o di enfasi, mentre in casi particolarmente ostici si è scelta la categoria Altro.

Con **polarità** si intende l’impressione che si ha del target di *locomotiva* alla fine della lettura dell’estratto. La polarità è associata a due elementi, i quali possono comparire da soli o insieme:

1. Un **giudizio** dell’articolista dei confronti del target, che può essere espresso in maniera esplicita, quindi individuabile nel testo grazie a un qualche tipo di elemento sintattico o lessicale, o implicita, quindi inferibile dal lettore solo indirettamente, grazie al suo bagaglio di conoscenze generali sul mondo. Un giudizio positivo (es: *Attendevate il ritorno della locomotiva tedesca per tirare la ripresa europea. Gli ultimi segnali da Bonn sono incoraggianti*) porta a un’annotazione positiva; un giudizio negativo (es: *Perché l'Ajax è il futuro, è una danza, è il computer contro la locomotiva Real [Madrid] che sbuffa, stride, prende velocità e potenza ma sempre su binari dove non corre la fantasia*) porta a un’annotazione negativa.
2. La descrizione di una **situazione** relativa al target[[5]](#footnote-5). Tale situazione può essere favorevole (annotazione positiva), dipingendo uno scenario di crescita economica (es: […] *la locomotiva Germania sta riprendendo a camminare, e l'anno prossimo il suo prodotto interno lordo potrebbe crescere del 2,4 per cento*), vittoria o successo (es: *Ai Mondiali di sci nordico vince la 30 km a tecnica classica Smirnov, una locomotiva*), anche quando sono espresse numericamente (*Rispetto alla locomotiva Veneto (+3,4%),* […]). Altrimenti, la situazione descritta può essere sfavorevole (annotazione negativa), trattandosi di una crisi, una recessione economica (es: *Ma ad aver frenato il dato complessivo è certamente il -0,3% della Germania, la ormai ex "locomotiva d'Europa" dove è entrato in crisi il modello basato su forniture di gas a basso costo*) o una sconfitta, anche quando sono espresse numericamente (es: *Il primo dato è quello della Germania in stallo dove il Pil di marzo-giugno è allo 0%. È un avvertimento alla manifattura e ai servizi italiani, dipendenti dalla «locomotiva d'Europa»*).

Qualora un articolo non riportasse né un giudizio positivo o negativo, né una situazione positiva o negativa, ma stesse semplicemente definendo il target *locomotiva* senza ulteriori specificazioni (es: *Risultato? In Europa è la Norvegia la locomotiva dell'auto elettrica*), o stesse usando il termine senza riferirsi a un target specifico, ma parlando solo di un generico *rango*/*funzione* *di* *locomotiva* (es: *Non credo — ha aggiunto — che servano molto gli inviti europei all'America a ribassare i tassi di interesse, né gli inviti americani all'Europa ad assumere il ruolo di "locomotiva economica" in una fase in cui gli Usa conoscono l'attuale rallentamento*), l’articolo va invece annotato come neutro.

Nel caso in cui un articolo riportasse giudizi o situazioni sia positive, sia negative, come nel caso in cui si descrivessero situazioni complesse o si sottolineassero sia i punti di forza, che i punti di debolezza del target, si sono applicate due strategie. La prima consiste nell’individuare quale delle polarità sia prevalente, in termini di enfasi, di temporalità (la polarità più legata al presente sarà quindi privilegiata) e di spazio dedicato. Per esempio, (1) è annotato con polarità positiva, poiché l’articolo sottolinea che i problemi del target sono inferiori ai suoi punti di forza; similmente anche (2) è positivo, poiché le qualità positive del target sono enfatizzate di più, rispetto a quelle negative. La seconda strategia si attiva quando non è possibile evincere una gerarchia tra le tre polarità possibili, poiché l’articolo ne presenta almeno due in maniera equivalente; in questa eventualità, si è applicata una scala di impatto, chiamata *Regola del Neg>Po>Neu*: la polarità negativa surclassa quelle positiva e neutra, mentre la polarità positiva prevale solo su quella neutra.

1. E un esperto della Banca Mondiale concorda: "Tutto il mondo riconosce che la Russia ha un enorme potenziale, e che può diventare una delle locomotive dell'economia mondiale del prossimo secolo", dice Aleksandr Morozov, "ha anche molti problemi, molti ostacoli da superare, ma nessuno appare insormontabile".
2. «Malgrado le difficoltà che lei ha elencato, le imprese italiane hanno messo in luce una straordinaria capacità di reazione che sta portando il nostro Paese [l’Italia] a essere la vera locomotiva d'Europa facendo meglio anche di Francia e Germania, tradizionali Paesi guida».

Nei casi in cui, invece, si notino due polarità distinte, una sul target e l’altra su un’altra entità a cui il target viene paragonato, l’annotazione tiene conto solo della polarità relativa al target, come si vede in (3), che è positiva.

1. È vero che il Nord Est tira la locomotiva con un tasso di crescita dell'1,43 per cento, ma anche il Mezzogiorno, con l'1,35, si colloca al di sopra della media nazionale.

Tutti i livelli dell’annotazione sono stati sottoposti a test di *inter-annotator agreement* (IAA), così da verificare il grado di oggettività e di affidabilità dell’annotazione. Infatti, nel corpus non mancano casi incerti, o in cui in generale l’annotazione può essere soggetta a dubbi e/o possa essere influenzata dalla sensibilità personale degli annotatori.

L’IAA della polarità è stato testato su un campione di 50 articoli di giornale bilanciati fra i vari sotto-corpora di LocoMIT; il campione è stato annotato da tre annotatrici (A, B e C) con esperienza nella linguistica di corpora. Per determinare l’*agreement* si è calcolato l’*α* di Krippendorff (2004), ottenendo un punteggio di 0.601. L’IAA nell’annotazione di target, paragone e macro-tematica è stato testato su un altro campione di 50 articoli di giornale bilanciati, annotati in questo caso da due annotatrici (A e D) con esperienza nell’analisi di prosa giornalistica. L’*agreement* è stato calcolato dunque sul *k* di Cohen (Carletta, 1996), ottenendo un punteggio di 0.875 per il target, 0.4 per il paragone e 0.805 per la macro-tematica[[6]](#footnote-6).

L’accordo più basso si ha dunque nell’annotazione del paragone; secondo la griglia di valutazione di Landis & Koch (1977), 0.4 indica un accordo sufficiente. Analizzando nel dettaglio il disaccordo tra le annotatrici, la categoria del paragone ha due caratteristiche principali che la rendono problematica da annotare. In primo luogo, quando si annotano paragoni che coinvolgono più elementi, non sempre c’è accordo su questi ultimi: in casi come (4), per esempio, l’annotatrice D ha inserito fra gli elementi paragonati anche la Francia, mentre l’annotatrice A ha inserito solo Italia e Germania. In secondo luogo, si sono avuti anche molti casi, in cui un’annotatrice individuava la presenza di un paragone, mentre l’altra non ne segnalava nessuno: in (5) l’annotatrice A ha rilevato un paragone tra Lombardia e Roma, mentre l’annotatrice D non ha riconosciuto alcun paragone. In entrambi i casi, si può attribuire il disaccordo al fatto che chi annota riconosca o meno un paragone implicito la cui presenza, in quanto tale, è *de facto* incerta e soggetta all’interpretazione personale: in (4), l’annotatrice A legge la prima frase come una semplice esposizione della situazione economica di tre Stati, mentre l’annotatrice D la legge come un paragone; in (5), al contrario, l’annotatrice A legge un paragone implicito tra la Lombardia e Roma, mentre l’annotatrice D rileva solo un’esposizione dei fatti.

1. La distanza del Pil italiano rispetto ai livelli pre-Covid (fine 2019) è di -3,8%, contro il -3,4% della Germania e il -3,3% della Francia. Se si prende a riferimento la Germania, che da sempre è considerata la vera locomotiva d'Europa, è vero che nel 2021 la sua crescita è risultata, finora, meno robusta di quella registrata in Italia, ma bisogna anche considerare che la caduta del Pil tedesco nell'anno della pandemia è stata decisamente più contenuta di quella registrata in Italia (-4,9% contro -8,9%).
2. «La locomotiva sui vaccini adesso è la Lombardia. Roma? Cerchino un altro» | […] Oggi sono convinto che la Lombardia sia la locomotiva

Per quanto l’accordo di 0.601 per la polarità sia inferiore rispetto a quello che Krippendorff (2004) considera un accordo alto, ossia 0.8, nella comunità scientifica il punteggio qui ottenuto è considerato un risultato accettabile (Artstein, 2017), specialmente in contesti di analisi in cui si richiede di riconoscere una caratteristica semantico-discorsiva di un testo (Gagliardi, 2018); secondo la griglia di Landis & Koch (1977) per l’interpretazione dei valori *k*, 0.601 rientra nell’accordo moderato. Anche in questo caso, l’analisi del disaccordo permette di comprendere quali siano le maggiori criticità di questa annotazione: infatti, su 19 casi di *disagreement*, quelli in cui si rileva un’indecisione tra positivo/negativo e neutro sono ben 15, mentre i casi di indecisione tra positivo e negativo sono solo due, così come i casi in cui le tre annotatrici hanno dato tre valori diversi. Questo fa comprendere come la regola del *Neg>Po>Neu* non risolva tutte le situazioni ambigue e come l’interpretazione della polarità di un articolo sia molto spesso dovuta alla sensibilità personale di chi annota. Per esempio, in (6) l’annotatrice A è stata l’unica ad annotare l’estratto come negativo, poiché ha interpretato *gorgo* come un termine sottilmente dispregiativo; in (7), invece, solo l’annotatrice C ha segnalato una polarità positiva, poiché ha associato il termine *film-locomotiva* a *capolavoro*.

1. C' è quanto basta per prevedere che la partita tra Fiat e sindacati s'annuncia già complessa almeno per due ragioni: perché s'innesta sulle polemiche circa la ripresa dell'inflazione e perché, come da prassi collaudata, quando si muove la locomotiva contrattuale Fiat risucchia nel suo gorgo molte altre categorie di lavoratori.
2. Ma che cosa succede quando un grande film americano come Jurassic Park, L'attimo fuggente arriva in Europa? Il distributore americano detta le condizioni: "Ti do Jurassic Park per dieci settimane, ma insieme devi prendere altri cinque film che proietti per due settimane ciascuno". È quello che si chiama "treno": un film-locomotiva con tanti vagoncini attaccati. […] Nel "treno" americano il capolavoro si accompagna alla spazzatura e il potere della distribuzione fa sì che quella sia più importante della nostra. Il mercato americano accoglie i capolavori europei, non la spazzatura.

Invece, l’accordo della macro-tematica (0.805) si classifica come considerevole, mentre quello del target (0.875) è quasi perfetto (Landis & Koch, 1977). I casi di disaccordo sulla macro-tematica riguardano soprattutto situazioni in incertezza sul fatto che un articolo vada classificato come economico o meno; per esempio, l’annotatrice D ha annotato il testo (8) come economico, poiché alla fine si nomina un generico sviluppo, mentre l’annotatrice A lo ha annotato come appartenente alla macro-tematica Altro. Infine, l’accordo sul target è molto alto, nonostante si tratti di una categoria aperta; questo fa capire che, in generale, sembra molto semplice riconoscere nel testo il referente della metafora.

1. Da questa centralità che proiettava l'America alla testa del progresso nel mondo, i Rockefeller furono i primi a rendersi conto che, senza coltivare in modo adeguato le scienze, l'istruzione superiore e universitaria americana, più che una locomotiva, avrebbe finito col rappresentare un freno allo sviluppo.

1. Da questo numero sono stati esclusi i casi in cui con *locomotiva* ci si riferisca letteralmente alla vettura motore del treno e gli articoli erroneamente riportati due volte da Infojuice. [↑](#footnote-ref-1)
2. Il corpus *La Repubblica* è liberamente consultabile al seguente link: https://bellatrix.sslmit.unibo.it/noske/public/#dashboard? corpname=repubblica [↑](#footnote-ref-2)
3. Per quanto il corpus *La Repubblica* e l’archivio storico de *La Stampa* non siano rappresentativi della comunicazione giornalistica italiana nel suo insieme, sono due archivi particolarmente utili per condurre ricerche esplorative come quella presentata in questo articolo, poiché liberamente accessibili e annotati nei metadati. LocoMIN è liberamente consultabile e scaricabile al seguente link: https://github.com/GloriaComandini/Corpora. [↑](#footnote-ref-3)
4. Nome e cognome di questo genere di target (es: Silvio Berlusconi, Massimo D’Alema, Marco Pantani, Silvio Martinello, Fabiana Luperini, Valeria Marini, Vladimir Smirnov, ecc.) sono invece annotati a parte, poiché identificare chi siano i politici o gli sportivi che più spesso sono chiamati *locomotiva* non è lo scopo di questa ricerca. [↑](#footnote-ref-4)
5. In tal senso, è bene sottolineare che l’annotazione svolta sul LocoMIN Corpus ha molti punti in comune con quella presentata in Collaoni (in stampa): entrambe sono state applicate agli articoli di InfoJuice e hanno le medesime categorie (positivo, negativo, neutro). Tuttavia, l’annotazione presentata in Collaoni (in stampa) riconosce la polarità solo in relazione all’eventuale giudizio esplicito espresso nei confronti del target. Nell’annotazione di LocoMIN, invece, il riconoscimento della polarità tiene conto anche dei giudizi impliciti e delle situazioni riportate. Pertanto, sebbene le due annotazioni possano assomigliarsi, i loro risultati non sono direttamente paragonabili. [↑](#footnote-ref-5)
6. L’α di Krippendorff e il *k* di Cohen misurano i risultati in un intervallo fra 1 e -1; 1 rappresenta il totale accordo fra annotatori, mentre 0 rappresenta l’accordo ottenuto qualora l’annotazione venisse fatta in maniera totalmente casuale, e i valori inferiori allo 0 indicano un disaccordo sistematico. Si è scelto di utilizzare l’*α* di Krippendorff per la polarità, poiché ideale per misurare l’accordo fra tre o più annotatori, mentre per l’accordo di target, paragone e macro-tematica si è scelto il *k* di Cohen perché ideale se applicato all’accordo fra due annotatori (Artstein, 2017). I valori di *α* e *k* sono paragonabili. [↑](#footnote-ref-6)