

Analisi computazionale dei report di sostenibilità: la vaghezza come strategia di greenwashing

Erica Cutuli

Università degli Studi di Catania, Italia
erica.cutuli@phd.unict.it

Abstract

Oggi è possibile elaborare automaticamente enormi quantità di testo in linguaggio naturale, ma persistono difficoltà nel trattare casi di ambiguità e vaghezza linguistica. L'obiettivo del presente contributo è l'analisi dei fenomeni di vaghezza e imprecisione nei report di sostenibilità, con particolare attenzione al *greenwashing*.

Il corpus analizzato è composto da 225 report di sostenibilità in lingua italiana, pubblicati da 45 aziende, tra 2017 il 2021. I testi, in formato PDF, sono stati elaborati con Sketch Engine e un notebook in Python, partendo dalla ricerca mirata di parole chiave solitamente legate al *greenwashing*, e annotando un campione di concordanze. In particolare, per ciascun esempio estratto è stato stabilito se rispondesse ai criteri per essere considerato un'*asserzione ambientale* e se fosse vago, classificando ove possibile i casi di vaghezza in cinque categorie semantiche (*quantity, degree, time, category e softening stance-taking*).

Il lavoro offre un duplice contributo: da un lato, l'analisi preliminare del linguaggio dei report di sostenibilità fornisce una migliore comprensione delle strategie linguistiche associate alle *asserzioni ambientali*, promuovendo una maggiore trasparenza e responsabilità da parte delle aziende; dall'altro, pone le basi per l'identificazione automatica di quelle riconducibili al *greenwashing* con un dataset in lingua italiana per l'allenamento di modelli di Intelligenza Artificiale con un valore aggiunto in prospettiva del raggiungimento dell'uguaglianza linguistica digitale.

Parole Chiave: Elaborazione del Linguaggio Naturale, sostenibilità, greenwashing, vaghezza, asserzioni ambientali, AIUCD2024.

Today, the automatic processing of large volumes of natural language text is increasingly feasible. However, challenges persist in addressing linguistic ambiguity and vagueness. This study aims to analyze phenomena of vagueness and imprecision in sustainability reports, with a particular focus on greenwashing.

The corpus analyzed consists of 225 Italian-language sustainability reports in PDF format, published by 45 companies between 2017 and 2021. These documents were processed using the Sketch Engine tool and a Python notebook. The analysis began with a targeted search for keywords commonly associated with greenwashing, followed by the annotation of a sample of concordances. Specifically, each extracted example was evaluated to determine whether it met

the criteria for an environmental claim and whether it exhibited vagueness. When applicable, instances of vagueness were classified into five semantic categories (quantity, degree, time, category, and softening stance-taking).

The dual contribution of this research lies in its provision of a deeper understanding of linguistic strategies in sustainability reports, promoting corporate transparency and accountability, while simultaneously establishing a foundation for the automated identification of greenwashing-related claims through the development of an Italian-language dataset for Artificial Intelligence training, thereby advancing digital linguistic equality.

Keywords: Natural Language Processing, Sustainability, greenwashing, vagueness, environmental claims, AIUCD2024.

1. Introduzione¹

Questo contributo si colloca tra le ricerche che utilizzano il *Natural Language Processing* (NLP) e l'analisi computazionale dei testi nell'ambito della sostenibilità. Esso si pone come obiettivo l'estrazione e l'analisi di asserzioni ambientali da report di sostenibilità aziendali e l'individuazione tra queste di quelle vaghe e riconducibili alla pratica del *greenwashing*. Per raggiungere tale scopo, è stato compilato un elenco di termini strettamente legati al fenomeno, ipotizzando che l'analisi delle concordanze di questi termini all'interno del corpus di report raccolti permetta, dapprima, di individuare le asserzioni ambientali e, successivamente, di verificare e classificare le diverse forme di vaghezza presenti secondo le nozioni teoriche esplorate in letteratura.

L'articolo è strutturato come segue: dopo una revisione della letteratura sui concetti di vaghezza linguistica e *greenwashing*, si descrivono il corpus e la metodologia adottata per l'analisi; si passa quindi all'annotazione e alla discussione dei risultati; infine, si delineano alcuni sviluppi futuri e si traggono le conclusioni principali dello studio.

2.1 Stato dell'arte

2.1 La vaghezza linguistica

I concetti di vaghezza e *greenwashing* condividono una natura intrinsecamente multidisciplinare, il che implica la difficoltà di definirli in modo univoco. La loro comprensione, infatti, varia in funzione delle discipline che li analizzano e degli obiettivi specifici che si propongono. Per quanto riguarda la vaghezza, essa può essere definita, in prima approssimazione, come "L'essere vago, cioè indeterminato, incerto, poco preciso".² In questo studio, si adatterà una definizione operativa ispirata a Mannaioli [16], che descrive la vaghezza come la "caratteristica di un enunciato che lascia parte dei suoi contenuti non detti, o per indeterminatezza lessicale o per costruzione sintattica". Sebbene essa sia considerata da alcuni filosofi come Russell [24] come una caratteristica intrinseca e pervasiva di tutte le lingue umane, è anche generalmente riconosciuto che determinate espressioni linguistiche manifestino un grado di vaghezza più

¹ Ultimo accesso agli URL in data 10/04/2024.

² "Vaghezza", Vocabolario Treccani online, <https://www.treccani.it/vocabolario/ricerca/vaghezza/>.

marcato rispetto ad altre, tanto da poter essere definite prototipicamente vaghe, come sottolineato da Pinkal [23], 185: «Il fenomeno della vaghezza è pervasivo nelle lingue naturali: quasi ogni concetto di linguaggio naturale ha un certo margine di imprecisione in almeno alcuni casi di applicazione. Le espressioni vaghe prototipiche (termini di colore come rosso e blu, aggettivi scalari come alto e intelligente, avverbi e preposizioni spaziali) presentano un'area di confine sostanziale di indefinitezza semantica in qualsiasi contesto naturale (sebbene le dimensioni e la posizione di queste aree possano variare)».³ In altre parole, termini come “rosso”, “alto” o “caldo” sono prototipici di vaghezza in quanto non hanno una linea di demarcazione netta che separi ciò che rientra nel loro campo di applicazione da ciò che ne è escluso. Questi termini, quindi, operano all'interno di un continuum di precisazioni contestuali: ad esempio, ciò che è considerato “alto” in un contesto potrebbe non esserlo in un altro, evidenziando la stretta dipendenza della vaghezza da fattori extralinguistici e contestuali.

Infatti, Pinkal introduce una “semantica della precisazione” per mostrare come il contesto influenzi il significato delle espressioni indefinite, determinandone il valore di verità e l'estensione semantica ([21], [23]). Studi recenti ([16], [17], [18]) approfondiscono ulteriormente questa prospettiva, mostrando come espressioni vaghe possano, in realtà, acquisire significati precisi in specifici contesti. Queste considerazioni implicano che, se ci si pone l'obiettivo di verificare se e in che misura un'espressione sia vaga, è fondamentale definire il contesto in cui essa viene interpretata e identificare i parametri che ne determinano il significato. Tuttavia, fornire una definizione formale e assoluta di contesto, inteso come l'insieme di elementi che influenzano l'interpretazione di un'espressione, rappresenta una sfida significativa, poiché rende necessario tenere in considerazione piani molto diversi. Tra questi fattori eterogenei, è possibile citare il contesto comunicativo di uno scambio verbale, che può includere luogo, tempo, oggetti e persone coinvolte; il contenuto testuale che precede o segue un'espressione vaga all'interno di un testo scritto; e, infine, elementi multimodali, come immagini, musica e inquadrature, particolarmente rilevanti in supporti audiovisivi quali gli spot pubblicitari.

Riconoscendo la complessità che deriva da tali variabili, appare più produttivo adottare in questa sede un approccio pragmatico che si focalizzi sul funzionamento della vaghezza nella comunicazione reale, delimitando il più possibile il fenomeno da analizzare mediante definizioni operative chiare e circoscritte. A tale scopo, l'analisi della letteratura esistente si è soffermata su alcuni lavori particolarmente sistematici che offrono un quadro teorico e metodologico utile per orientare l'identificazione e la classificazione delle espressioni vaghe, sia nel caso in cui si seguano approcci manuali sia nell'ottica di applicazione di strumenti computazionali che mirano a automatizzare tale processo.

La letteratura sulle espressioni vaghe in lingua inglese si è concentrata su analisi di tipo lessicale, sintattico e pragmatico, partendo dai lavori pionieristici di Joanna Channell, in particolare *Vague Language* [3]. La sua analisi ha introdotto una categorizzazione sistematica delle espressioni vaghe, tra le principali categorie individuate vi sono:

³ Questa e le successive citazioni dall'inglese sono state tradotte in italiano dall'autrice; per completezza si riportano in nota i testi originali. Testo in lingua originale: «The vagueness phenomenon is pervasive in natural languages: Almost every natural-language concept has some margin of imprecision in at least some cases of application. Prototypical vague expressions – colour terms like red and blue, scalar adjectives like tall and clever, spatial adverbs and prepositions – come with a substantial borderline-area of semantic indefiniteness in any natural context (although the size and location of these areas may vary)».

- *Quantificatori indefiniti*: espressioni come “alcuni”, “un po’”, “molti” sono frequentemente utilizzate per esprimere quantità in modo indeterminato.
- *Approssimatori e hedges*: termini come “all’incirca”, “quasi”, “relativamente” attenuano la precisione delle affermazioni.
- *Estensioni categoriali*: espressioni come “e così via”, “e cose del genere” estendono il significato di un’affermazione senza definirne i limiti esatti.
- *Espressioni generiche o casi di open texture*: parole come “roba”, “gente”, “questioni” possono riferirsi a una molteplicità di stati di cose e la loro specificità spesso dipende molto dal contesto.
- *Pronomi non anaforici*: quando il referente non è chiaramente specificato o si riferisce a qualcosa di implicito o generico.

Successivi studi [4][7][10][25] hanno ampliato il campo di indagine, non solo includendo nuove categorie linguistiche — come aggettivi di grado (ad esempio, “straordinario” e “unico”) e figure retoriche quali metafore e iperboli —, ma anche introducendo applicazioni metodologiche diverse. Tra queste si annoverano l’analisi comparativa di corpus per lo studio delle differenze nell’uso della vaghezza tra l’inglese britannico e americano in contesti educativi [25], l’impiego della *Adaptation Theory* come quadro teorico per esplorare il legame tra scelte linguistiche nella pubblicità e comportamento dei consumatori [4], e l’approccio corpus-assisted adottato per analizzare la funzione comunicativa del linguaggio vago all’interno dei report di responsabilità sociale d’impresa (CSR) [10].

La letteratura ha così approfondito non solo gli aspetti lessicali, ma anche le finalità comunicative della vaghezza, come sintetizzato da Ruzaite [25], 28: «Il linguaggio vago è un fenomeno linguistico naturale, solitamente intenzionale e multifunzionale, che comporta imprecisioni e viene impiegato per determinate strategie comunicative». ⁴ Questa definizione mette in luce la natura funzionale e intenzionale della vaghezza, sottolineando come essa sia spesso impiegata per adattare il messaggio alle esigenze del contesto, piuttosto che per un semplice difetto di chiarezza: la vaghezza intenzionale, dunque, non va intesa come un errore o una mancanza di precisione, bensì come una scelta strategica da parte di chi trasmette il messaggio per perseguire specifici obiettivi comunicativi.

In particolare, Chen [4] osserva che nei contesti pubblicitari la vaghezza viene impiegata per evitare di fare affermazioni specifiche sui loro prodotti o servizi che potrebbero essere difficili da mantenere e per proteggere gli interessi degli inserzionisti. In particolare, la vaghezza può aiutare gli inserzionisti a evitare responsabilità legali o danni alla reputazione nel caso in cui i consumatori interpretino male le loro pubblicità: l’uso di termini vaghi come “migliore” o “più efficace” può rendere difficile per i consumatori dimostrare che un prodotto non è all’altezza delle loro aspettative. Analogamente, Li [13] analizza come la vaghezza nei testi legislativi possa conferire flessibilità interpretativa in modo tale da mitigare potenziali problemi e prevenire conflitti futuri: ad esempio, l’uso di espressioni come “dati necessari” o “misure appropriate” consente alle parti di interpretare il significato di queste in base al contesto specifico.

Accanto a questi usi strategici in ambito pubblicitario e legislativo, ulteriori ricerche si sono concentrate sul potere persuasivo della vaghezza: Mannaioli [16], ad esempio, evidenzia come il

⁴ «Vague language is a natural, usually purposeful and multi-functional linguistic phenomenon that involves im[10]precision and is employed for certain communicative strategies».

ricorso alla vaghezza consenta agli emittenti di evitare impegni concreti su contenuti rischiosi o potenzialmente non veritieri, come promesse o affermazioni difficilmente verificabili, e permetta ai destinatari di integrare il messaggio con significati personali, proiettando su di esso valori e idee proprie. Tale dinamica, come argomentato da Mannaioli [18] aumenta la probabilità che il messaggio venga accettato, poiché ciascun destinatario lo interpreta in base alle proprie opinioni e preferenze.

In altre parole, la vaghezza si dimostra persuasiva non tanto per ciò che dice esplicitamente, quanto per ciò che lascia non detto: evitando di fornire informazioni dettagliate e specifiche, il messaggio vago permette al destinatario di “riempire i vuoti” con le proprie interpretazioni. Questo meccanismo facilita la comunicazione di connotazioni generiche e universalmente accettabili, evitando allo stesso tempo di entrare nel merito di denotazioni specifiche che potrebbero risultare controverse o divisive [14][15].

Queste strategie comunicative sono ampiamente sfruttate nell’ambito del marketing o, più in generale, della comunicazione aziendale, per rendere le pubblicità più persuasive [4] o costruire un’immagine positiva dell’azienda [10]. Tra i vari contesti in cui la vaghezza gioca un ruolo cruciale, questo contributo si focalizza sull’ambito della sostenibilità, e in particolare sul fenomeno del *greenwashing*.

2.2 Il greenwashing

Il termine *greenwashing*, coniato da Jay Westerveld nel 1986, descrive inizialmente affermazioni ambientali false o fuorvianti, ma nel corso del tempo ha acquisito una portata più ampia, includendo questioni di governance, ambientali e sociali. In ambito di sostenibilità, il *greenwashing* si riferisce all’intenzionale adozione di tattiche ingannevoli volte a far credere che un’azienda, un prodotto o un servizio siano più sostenibili di quanto non siano in realtà. Secondo la definizione generale fornita dalle Nazioni Unite, il *greenwashing* consiste nell’uso di “tattiche ingannevoli dietro le dichiarazioni ambientali”.⁵ Lane [12] ha evidenziato la necessità di ridefinire il concetto per comprendere tutte le forme di comunicazione ingannevole che influenzano le percezioni ambientali delle aziende [22]. Il *greenwashing* non solo compromette la fiducia dei consumatori, ma danneggia anche le aziende autenticamente sostenibili e ostacola la canalizzazione di risorse e investimenti verso progetti realmente ecologici, rallentando così il progresso verso una sostenibilità effettiva [28]. La letteratura recente ha analizzato sistematicamente il fenomeno, proponendo diverse classificazioni. De Freitas Netto et al. [8] distinguono il *greenwashing* su due livelli:

1. *Livello aziendale*: riguarda la comunicazione complessiva dell’organizzazione sulle proprie pratiche ambientali. In questo caso, l’azienda presenta un’immagine generale di responsabilità ecologica che può essere ingannevole o non pienamente supportata dai fatti.
2. *Livello di prodotto/servizio*: riguarda affermazioni specifiche sui benefici ambientali di un singolo prodotto o servizio. In questo caso, il *greenwashing* si manifesta attraverso certificazioni o etichette ambientali fuorvianti, o attraverso dichiarazioni che esagerano o falsificano le qualità ecologiche di prodotti specifici.

⁵ “Greenwashing – the deceptive tactics behind environmental claims”, Nazioni Unite, <https://www.un.org/en/climatechange/science/climate-issues/greenwashing>

Inoltre, identificano criteri di classificazione basati sul contenuto (orientamento al prodotto, processo, immagine, fatti ambientali) e sulla strategia utilizzata (vaghezza, omissione, falsità). Bernini e La Rosa [1] offrono una prospettiva complementare, articolata in tre pilastri:

1. L'omissione di dati negativi, o divulgazione selettiva, che si verifica quando le aziende nascondono i dati negativi sulle prestazioni ambientali e/o enfatizzano sproporzionatamente le prestazioni positive.
2. La divulgazione ingannevole, che implica affermazioni false o comunque parzialmente non veritiere sulle pratiche di sostenibilità di un'azienda.
3. La discrepanza tra affermazioni dichiarate e azioni effettive, questo accade quando le aziende non sostengono le loro affermazioni ambientali con azioni concrete.

Negli ultimi anni, il fenomeno del *greenwashing* ha ricevuto un'attenzione crescente non solo nella letteratura accademica, ma anche nel dibattito pubblico, contribuendo a una comprensione più articolata e a una classificazione sistematica delle sue molteplici manifestazioni. Tra queste, emergono tipologie sofisticate, tra cui il cosiddetto *greenlighting*, che si verifica quando le aziende enfatizzano un aspetto particolarmente "verde" delle loro attività o dei loro prodotti – spesso rappresentativo di una minima parte delle loro operazioni – per distogliere l'attenzione da pratiche dannose per l'ambiente condotte altrove; il *greenrinsing*, attraverso cui le aziende creano un'illusione di impegno verso la sostenibilità modificando costantemente i loro obiettivi ambientali, posticipandone le scadenze o alzando i parametri senza aver raggiunto quelli precedenti; e il *greenshifting*, che si distingue per lo spostamento della responsabilità sui consumatori, in questo caso, le aziende promuovono comportamenti individuali sostenibili senza affrontare il proprio ruolo nell'inquinamento o nel cambiamento climatico [9][28].

Questi contributi confermano la natura complessa e multidisciplinare del *greenwashing*, sottolineando la necessità di approcci integrati e complementari per condurre un'analisi esaustiva. Il presente studio si concentra in particolare sui casi che rientrano nel secondo pilastro relativo alla divulgazione ingannevole [1], e che si allineano alla definizione di *greenlabelling*, tattica che consiste nell'etichettare prodotti o servizi come "verdi" o "sostenibili" utilizzando termini fuorvianti o privi di una definizione chiara [9][28]. In altri termini, l'oggetto della ricerca include i casi in cui il fenomeno si presenta sotto forma di: «[...] *greenwashing* a livello di prodotto/servizio, che utilizza argomentazioni testuali che si riferiscono esplicitamente o implicitamente ai benefici ecologici di un prodotto o servizio per creare un'affermazione ambientale fuorviante» [8], 7.⁶ Più specificamente, si tratta spesso di «[...] affermazioni eccessivamente vaghe, ambigue, troppo ampie e/o prive di una chiara definizione» [8], 8,⁷ definizione che evidenzia appieno il collegamento tra la vaghezza linguistica e tale fenomeno. Questi casi specifici, intrecciandosi con il concetto di vaghezza linguistica e con la tematica della pubblicità ingannevole, si configurano come un'area di indagine promettente per un'analisi linguistico-pragmatica, che diventa fondamentale per comprendere come le aziende utilizzino intenzionalmente la vaghezza per influenzare la percezione pubblica e costruire narrazioni persuasive senza impegnarsi in azioni concrete o verificabili.

⁶ «[...] product/service-level claim greenwashing, which uses textual arguments that explicitly or implicitly refer to the ecological benefits of a product or service to create a misleading environmental claim».

⁷ «[...] claims that are overly vague, ambiguous, too broad, and/or lacking a clear definition».

L'interesse crescente delle autorità globali verso la regolamentazione del *greenwashing* sottolinea la rilevanza critica di questo fenomeno, riconosciuto per la sua capacità di ostacolare il progresso verso una sostenibilità autentica. In particolare, la Commissione Europea ha adottato misure sempre più stringenti per contrastare tali pratiche, dichiarando «[...] l'impegno a contrastare la problematica delle asserzioni ambientali false, garantendo agli acquirenti di ricevere informazioni attendibili, comparabili e verificabili, e così permettendo loro di prendere decisioni più sostenibili e ridurre il rischio di un marketing ambientale fuorviante (*greenwashing*)» [5],1; dove per asserzione ambientale si intende «[...] un messaggio o una dichiarazione [...] che asserisce o induce a ritenere che un dato prodotto o professionista ha un impatto positivo o nullo sull'ambiente oppure è meno dannoso per l'ambiente rispetto ad altri prodotti o professionisti oppure ha migliorato il proprio impatto nel corso del tempo» [5], 3.

Questa attenzione delle istituzioni non è recente: già nel 2021, la Commissione Europea aveva pubblicato i risultati di uno studio innovativo, il primo screening dei siti web focalizzato sul *greenwashing*. I risultati dell'analisi, che ha preso in esame 344 affermazioni potenzialmente ingannevoli, hanno mostrato che il 37% delle affermazioni faceva uso di termini vaghi e generici, mentre nel 59% dei casi mancavano prove facilmente accessibili a sostegno delle dichiarazioni. In generale, le autorità hanno stimato che nel 42% dei casi le affermazioni potevano essere considerate false o ingannevoli, configurando una potenziale violazione della Direttiva sulle pratiche commerciali sleali [6].

Sebbene le direttive della Commissione Europea si concentrino principalmente sulla comunicazione pubblicitaria delle aziende rivolta ai consumatori, gli stessi principi possono essere applicati anche alla comunicazione destinata agli stakeholders attraverso i report di sostenibilità. Questi documenti, che includono diverse tipologie come i report CSR (Corporate Social Responsibility), ESG (Environmental, Social, Governance), finanziari e non finanziari, offrono un terreno fertile per l'analisi delle strategie comunicative persuasive adottate dalle aziende.

I report aziendali rappresentano un oggetto di studio multidisciplinare, con approcci che spaziano dall'ambito linguistico e comunicativo a quello economico e sociale. La loro crescente diffusione, favorita anche dall'obbligatorietà di tali documenti per specifiche categorie di imprese, ha generato una quantità significativa di dati testuali, spesso analizzati attraverso tecniche di Natural Language Processing (NLP). Come evidenziato dalla rassegna di Moodaley e Telukdarie [20], l'applicazione dell'Intelligenza Artificiale (IA), del NLP e del Machine Learning ai report di sostenibilità è ormai una pratica consolidata. Tuttavia, l'utilizzo di queste tecnologie per studiare il fenomeno del *greenwashing* e gli studi sul *greenwashing* in relazione ai report di sostenibilità rappresentano ambiti di ricerca ancora emergenti.

Alcuni studi recenti stanno gettando le basi per colmare questa lacuna: gli stessi Moodaley e Telukdarie [19] hanno delineato un quadro concettuale per l'adattamento di *Large Language Models* (LLMs) al dominio specifico della sostenibilità; Webersinke et al. [27] propongono una versione del modello BERT allenato su testi inerenti all'ambito climatico, per migliorare l'accuratezza dell'analisi in questo ambito; parallelamente, Stambach et al. [26] propongono un dataset annotato e un modello allenato su questo dataset con l'obiettivo specifico di automatizzare l'individuazione di frasi legate all'ambiente. Questi contributi dimostrano il potenziale dell'IA e del NLP non solo per l'analisi dei report di sostenibilità, ma anche per identificare e contrastare il *greenwashing*. Tali tecnologie, se adeguatamente adattate, possono offrire strumenti efficaci per valutare criticamente le strategie linguistiche e comunicative delle aziende, contribuendo a promuovere una maggiore trasparenza e responsabilità nella comunicazione aziendale.

2.3 Il presente studio

Il contributo di questo studio si articola su due fronti complementari: da un lato, la costruzione di un corpus composto da 225 report di sostenibilità in lingua italiana, pubblicati da 45 aziende tra il 2017 e il 2021; dall'altro, come *proof of concept*, la creazione di un dataset annotato (un campione di 100 frasi) che consente un'analisi preliminare del linguaggio utilizzato nei report aziendali e della comprensione delle strategie linguistiche associate al *greenwashing*, e che potrà successivamente essere ampliato e finalizzato all'addestramento di modelli di intelligenza artificiale per il riconoscimento delle asserzioni ambientali vaghe.

Questo dataset rappresenta un primo passo verso l'identificazione automatica del *greenwashing*, in linea con i precedenti contributi [2][19][26][27] che hanno già esplorato l'adattamento di modelli di Large Language Models (LLMs) per scopi specifici. Inoltre, poiché la maggior parte delle ricerche in questo ambito si concentra quasi esclusivamente sull'inglese, riteniamo che sviluppare strumenti specifici per la lingua italiana costituisca un ulteriore valore aggiunto, contribuendo a promuovere l'uguaglianza linguistica digitale e a colmare il divario nella rappresentanza delle diverse lingue in questo ambito.

La centralità dell'analisi linguistica nella comunicazione aziendale emerge con particolare evidenza dalle recenti ricerche che collegano l'utilizzo di dichiarazioni ambientali vaghe, imprecise o non verificabili — fenomeno indicato nello studio di Bingler et al. [2] con il termine “cheap talk” — a un aumento delle emissioni. Gli autori osservano che:

[...] abbiamo dimostrato che le aziende con livelli più elevati di “cheap talk” sono associate a una maggiore copertura mediatica negativa in materia ambientale e a una maggiore crescita delle emissioni. Queste relazioni suggeriscono che le aziende con un livello elevato di “cheap talk” diano priorità al mantenimento di una percezione pubblica positiva piuttosto che apportare cambiamenti significativi alle proprie pratiche aziendali. Di conseguenza, ciò può ostacolare il progresso verso gli obiettivi globali di riduzione delle emissioni, poiché le aziende che si impegnano in “cheap talk” non adottano le misure necessarie per mitigare il loro impatto climatico e gestire i rischi climatici.⁸

Pertanto, l'individuazione di fenomeni linguistici associati al *greenwashing* potrebbe offrire strumenti più precisi per monitorare le pratiche aziendali, attività di cui sono responsabili diversi soggetti, tra cui, a livello interno, i Sustainability Manager e i Comitati ESG, e, a livello esterno, enti pubblici come l'ISTAT, il Ministero dell'Ambiente e della Sicurezza Energetica (MASE) e il Sistema Nazionale per la Protezione dell'Ambiente (SNPA). Un monitoraggio più efficace potrebbe favorire il raggiungimento degli obiettivi di sostenibilità, contribuendo a orientare il comportamento delle aziende verso pratiche più responsabili e concrete. Lo studio, dunque, non solo intende arricchire il panorama delle ricerche sul *greenwashing*, ma si propone anche come un contributo innovativo per l'analisi linguistica, tecnologica e sociale nel contesto della sostenibilità.

⁸ «[...] we demonstrate that companies with higher levels of cheap talk are associated with higher negative environmental news coverage and with higher emissions growth. These relationships suggest that companies with a high level of cheap talk prioritize maintaining a positive public perception over making meaningful changes to their business practices. Consequently, this can impede progress toward global emissions reduction targets, as companies that engage in cheap talk fail to take the necessary steps to mitigate their climate impacts and manage their climate risks».

3. Descrizione del corpus

Il corpus analizzato è stato costruito attraverso una raccolta manuale, a partire da un elenco pubblicato dalla Commissione Nazionale per le Società e la Borsa (CONSOB),⁹ che comprendeva i soggetti che, nel periodo compreso tra il 1° gennaio e il 16 novembre 2022, avevano reso pubblici i loro bilanci non finanziari per l'anno fiscale iniziato il 1° gennaio 2021. Per ciascuna impresa individuata, sono stati scaricati dal sito ufficiale o da repository pubblici¹⁰ tutti i report di sostenibilità disponibili, per un totale iniziale di 1392 file in formato PDF. Al fine di arricchire il dataset per ulteriori analisi, i documenti raccolti sono stati integrati con dati finanziari provenienti da *FactSet Research Systems Inc.* (2023), che includevano informazioni su classificazioni settoriali (RBICS), ricavi e eleggibilità secondo la Tassonomia Europea. L'integrazione è stata effettuata tramite un'operazione di *inner join* sui nomi delle aziende, seguita da una selezione per il periodo 2017–2021 e dalla rimozione dei casi con dati mancanti. Il corpus finale risulta composto da 225 report di sostenibilità in lingua italiana, relativi a cinque anni consecutivi (2017-2021) e pubblicati da 45 aziende.¹¹

Considerando il fatto che ciascuna azienda può operare in settori diversi, si riporta in Tabella 1 un riepilogo dei settori rappresentati all'interno del corpus.

Settore	Numero di aziende che operano nel settore	Percentuale
Settore industriale	17	38%
Beni di consumo ciclici	13	29%
Finanza	9	20%
Servizi aziendali	9	20%
Servizi al consumatore	6	13%
Sanità	5	11%
Materiali non energetici	5	11%
Tecnologia	4	9%

⁹ Disponibile a questo URL: <https://www.consob.it/web/area-pubblica/storico-elenco-dnf-al-16-novembre-2022>.

¹⁰ In particolare, il database dell'“Osservatorio DNF - Osservatorio delle Dichiarazioni Non Finanziarie (DNF) e delle Pratiche Sostenibili”, accessibile all'URL: <https://www.osservatoriodnf.it/it/home/about-dnf>

¹¹ Di seguito l'elenco completo delle aziende selezionate: Amplifon Spa, Arnoldo Mondadori Editore Spa, Atlantia Spa, Autogrill Spa, Avio Spa, Banca Generali Spa, Be Shaping The Future Spa, Brembo Spa, Brunello Cucinelli Spa, Buzzi Unicem Spa, Cairo Communication Spa, Caltagirone Spa, Cembre Spa, Datalogic Spa, Diasorin Spa, Dovalue Spa, El.En. Spa, Elica Spa, Enav Spa, Enel Spa, Fiera Milano Spa, Fincantieri Spa, Gruppo Mutuonline Spa, Immsi Spa, Iren Spa, Leonardo Spa, Lu-Ve Spa, Moncler Spa, Monrif Spa, Openjobmetis Spa, Pininfarina Spa, Pirelli & C. Spa, Prima Industrie Spa, Saes Getters Spa, Safilo Group Spa, Salvatore Ferragamo Spa, Saras Spa, Servizi Italia Spa, Sol Spa, Technogym Spa, Tesmec Spa, Tod's Spa, Unipol Gruppo Spa, Webuild Spa, Zignago Vetro Spa.

Servizi pubblici	3	7%
Beni di consumo non ciclici	3	7%
Energia	2	4%

Tabella 1 Tabella riassuntiva dei settori in cui operano le aziende selezionate.

Il settore industriale, che include la produzione e la gestione di infrastrutture e tecnologie per il trasporto, emerge come il più numeroso. Segue il settore del commercio al dettaglio e della produzione, che spazia dall'abbigliamento e gli accessori alla produzione di componenti per veicoli e all'intrattenimento online. In contrasto, i settori legati alla produzione, distribuzione e vendita di energia, così come quelli relativi ai servizi per la salute, la cura personale e l'educazione, risultano meno presenti.

4. Metodologia

I report di sostenibilità sono stati scaricati in formato pdf dai siti ufficiali delle aziende e sono stati analizzati con il tool Sketch Engine [11], che permette facilmente l'estrazione del testo e il suo processamento (come, ad esempio, *tokenizzazione* e *Part Of Speech tagging*). L'estrazione di testo dai PDF rappresenta ancora una sfida, spesso caratterizzata da un certo grado di imprecisione, soprattutto in presenza di immagini e layout complessi. Nonostante tali difficoltà, Sketch Engine ha raggiunto un livello di accuratezza ritenuto adeguato a soddisfare gli obiettivi del presente studio: la maggior parte del contenuto testuale è stata correttamente estratta, permettendo un'analisi efficace senza la necessità di un *pre-processing* esteso dell'intero corpus. Ciò ha comportato un significativo risparmio di tempo e risorse che altrimenti sarebbero stati necessari per preparare i documenti prima dell'analisi, migliorando l'efficienza complessiva del processo. Tuttavia, va notato che la segmentazione delle frasi, indicate nel tool come *sentences*, è risultata poco precisa, tendendo a includere segmenti di testo estranei (come titoli o didascalie) all'interno delle frasi estratte, oppure a segmentare in modo irregolare blocchi testuali continui. Ciò è probabilmente dovuto alla struttura grafica dei report di sostenibilità — caratterizzata dall'uso frequente di titoli, sottotitoli, elenchi puntati ed elementi visivi spesso privi di punteggiatura convenzionale — che può compromettere la corretta identificazione dei confini fraseologici da parte degli algoritmi di *tokenizzazione* e divisione in frasi.

Per valutare e gestire questo problema, è stato effettuato un controllo manuale sul campione di frasi selezionato per l'annotazione successiva. Durante questa fase, è stata realizzata una normalizzazione preliminare consistente principalmente nella rimozione di segmenti testuali in eccesso (come titoli isolati, elenchi puntati, o brevi didascalie) al fine di garantire la coerenza sintattica delle frasi. Non è stata effettuata una stima quantitativa formale degli errori sull'intero corpus, ma la verifica manuale sul campione ha confermato che, nonostante le criticità nella segmentazione, il testo estratto risultava generalmente idoneo agli obiettivi dell'analisi.

Una volta caricato e processato, il corpus contiene 14.279.823 *tokens* e 304.698 *sentences*, con una media di 63.466 *tokens* per report. La distribuzione dei *tokens* per anno (Figura 1) mostra un andamento di crescita costante passando da quasi due milioni (1.917.907) per i report del 2017 a quasi quattro milioni (3.858.273) per il 2021. Questo dimostra una chiara tendenza per le aziende a scrivere report di sostenibilità in media sempre più lunghi, con un contenuto testuale che arriva a raddoppiare nei cinque anni presi in esame.

Tale aumento può essere attribuito non solo al crescente interesse e ai requisiti regolatori, ma anche alla maggiore inclusione di report integrati (documenti che combinano in un unico file sia le informazioni finanziarie sia quelle di sostenibilità) nel corpus analizzato. Inoltre, la crescente lunghezza dei report di sostenibilità solleva domande sulla qualità e la densità delle informazioni contenute: mentre un aumento del volume dei report può indicare una maggiore trasparenza e dettaglio, è anche possibile che la quantità di testo non sempre corrisponda a un reale incremento del contenuto informativo. Pertanto, le analisi condotte in questo studio possono essere rilevanti poiché forniscono una valutazione qualitativa del contenuto, aiutando a discernere tra l'espansione puramente quantitativa e l'effettivo valore aggiunto delle informazioni presentate.

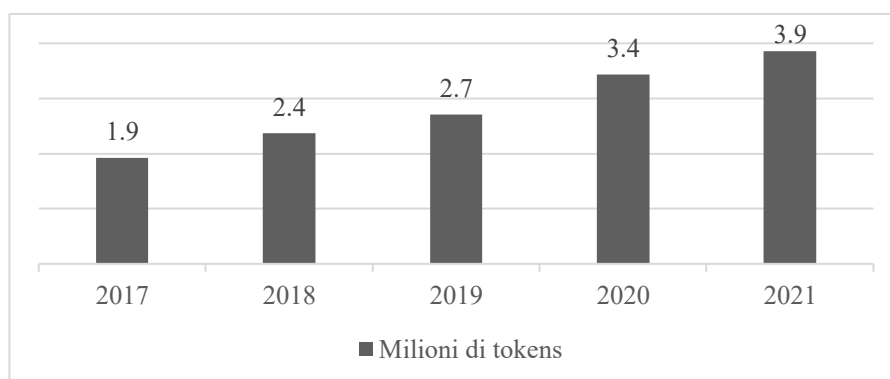


Figura 1 Distribuzione di tokens per anno

Per quanto riguarda la distribuzione dei *tokens* per azienda si passa da un massimo di 1.286.124 (Banca Generali Spa, 9% del corpus) ad un minimo di 60.291 (Gruppo Mutuonline Spa, 0,4% del corpus). Questa disparità, rappresentata graficamente in Figura 2, indica che alcune aziende hanno un impatto maggiore sull'analisi rispetto ad altre, suggerendo la necessità di considerare l'influenza delle singole aziende sui risultati complessivi dell'analisi.

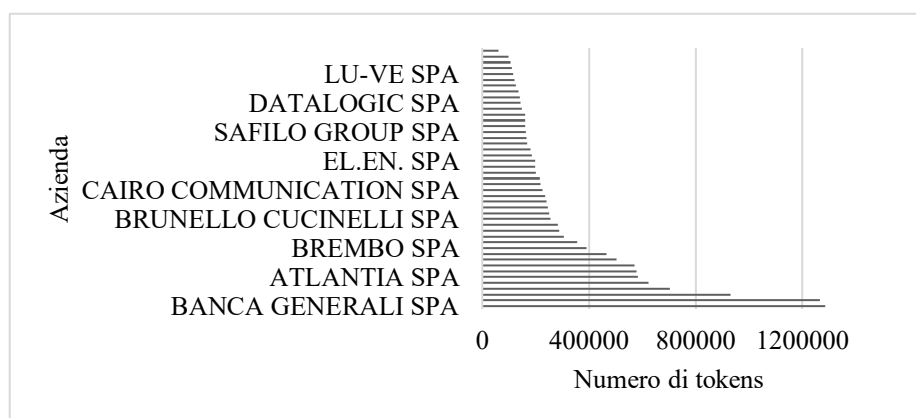


Figura 2 Distribuzione di tokens per azienda

Data l'assenza di dataset annotati o corpora composti esclusivamente da esempi di *greenwashing*, l'implementazione di un processo di estrazione automatica delle parole chiave non è stata possibile. Pertanto, si è scelto di fare riferimento alla metodologia utilizzata da Li [13]: come primo passo della sua analisi, l'autore ha selezionato casualmente un sample di 100 documenti dal corpus e ha stilato un elenco degli elementi lessicali associati all'uso vago incontrati durante la lettura dei documenti. Analogamente, sono stati letti integralmente 15 report redatti da tre aziende (Enel SpA, Fincantieri SpA, Zignago Vetro SpA)¹² con una media di 96996 *tokens* per report, insieme alla documentazione ufficiale della Commissione Europea [6],[5] che offre definizioni ed esempi concreti, ed è stata stilata una lista di 20 termini ritenuti facilmente associabili al fenomeno del *greenwashing*, cercando di mantenere una giusta rappresentazione della tematica ambientale e della vaghezza linguistica. La maggior parte di questi termini presenta caratteristiche riconducibili alle principali categorie di vaghezza individuate in letteratura: termini come "sostenibile", "ecologico", "naturale", "green", "ecosostenibile" o "verde" si prestano a essere considerati *espressioni generiche/open texture*, in quanto privi di definizioni condivise e fortemente dipendenti dal contesto; mentre espressioni come "riduzione", "ridurre" o "minore impatto" possono agire come *quantificatori indefiniti o approssimatori*, se evocano una variazione di quantità non specificata. L'ipotesi di partenza è che la ricerca di questi termini consenta di individuare *asserzioni ambientali* delle quali verificare l'eventuale vaghezza leggendone le concordanze.

Per l'estrazione delle frasi contenenti i termini selezionati è stata adottata una metodologia di ricerca lessicale basata sull'analisi automatizzata dei testi. Il procedimento ha previsto la ricerca dei lemmi di interesse senza distinzione tra maiuscole e minuscole, nonché l'estrazione di intere frasi contenenti tali lemmi, al fine di preservare il contesto d'uso delle espressioni individuate. Per gestire la variabilità linguistica e ottimizzare la copertura dei risultati, sono stati utilizzati pattern di ricerca che includessero varianti ortografiche e sintattiche: ad esempio, la ricerca di "eco--sostenibile" ha permesso di includere le varianti "ecosostenibile", "eco sostenibile" e "eco-sostenibile". Analogamente, la combinazione "emissioni zero|zero emissioni" ha permesso di raggruppare le due varianti nella stessa ricerca.

Questo approccio è stato implementato attraverso uno strumento di analisi testuale che consente l'esplorazione delle concordanze nei corpora, nello specifico, Sketch Engine.¹³

I risultati ottenuti dalle ricerche delle concordanze dei 20 termini selezionati sono riassunti nella Tabella 2.

Ricerca	Numero di occorrenze totali	Numero di frasi distinte
sostenibile	7.831	6.951
riduzione	7.419	6.038
ridurre	4.149	3.426
co2	3.782	3.155

¹² Queste aziende erano state selezionate per ulteriori studi correlati secondo parametri come emissioni, ricavi e numero di dipendenti.

¹³ I parametri utilizzati sono "simple" per la ricerca per lemma e "sentences" che consente di estrarre le frasi complete contenenti le parole ricercate.

naturale	3.098	2.607
impatto ambientale	2.857	2.351
green	1.785	1.540
riciclare	1.595	1.341
biodiversità	1.020	905
verde	478	433
riciclato	357	327
ecologico	305	277
biologico	287	247
eco--sostenibile	118	106
emissioni zero zero emissioni	112	106
riciclabile	99	79
minore impatto	82	72
eco--compatibile	70	59
biodegradabile	57	55
impatto zero	14	10
Totali	35.515	27.888¹⁴

Tabella 2 Tabella riassuntiva dei risultati di ricerca delle concordanze.

Questi dati sono stati scaricati in formato tabellare (.xlsx) e successivamente elaborati utilizzando un notebook in Python. Il *dataframe* iniziale, che combinava tutti i risultati delle ricerche, contava 35.515 righe. Dopo l'eliminazione delle righe duplicate, che includevano casi in cui le parole ricercate si ripetevano all'interno della stessa frase o in frasi adiacenti non correttamente separate durante il processamento, il numero è stato ridotto a 33.455 righe. Considerando che alcune aziende potrebbero ripetere le stesse frasi in report diversi o che una singola frase potrebbe essere ripetuta in più righe a causa della presenza di due o più termini ricercati, è stato verificato che il numero di frasi distinte ammonta a 27.888.

4.1 Annotazione e analisi

Dato l'elevato numero di frasi estratte, si è scelto di selezionarne casualmente 100 per proseguire con l'analisi e l'annotazione; tale scelta ha permesso inoltre di verificare manualmente la corretta segmentazione delle frasi rimuovendo eventuali segmenti testuali in eccesso. Per ciascuna di

¹⁴ Questa cifra si riferisce al numero totale di frasi distinte, rimuovendo tutte quelle duplicate per le diverse ragioni sopraindicate. Non corrisponde dunque alla somma delle cifre in colonna che rappresentano invece le frasi senza duplicati per ciascuna parola chiave ricercata.

queste frasi, è stato annotato se essa possa essere considerata un'asserzione ambientale, secondo la definizione data dalla Commissione Europea, citata precedentemente, e se presenti caratteristiche di vaghezza. In particolare, per l'annotazione della caratteristica della vaghezza è stata adottata una definizione operativa ispirata a quella proposta da Mannaioli [16]:

Caratteristica di un enunciato che lascia parte dei suoi contenuti non detti, o per indeterminazione lessicale (espressioni indefinite, approssimatori...) o per costruzione sintattica (omissioni di costituenti).

Questa caratteristica sarà attribuita, per necessità pratiche, prendendo in considerazione l'enunciato fuori dal suo contesto più ampio, cercando eventuali precisazioni contestuali all'interno dell'enunciato stesso.

Per la discussione dei diversi casi, si riportano di seguito tre esempi in cui sono state sottolineate le parti di interesse.

- **Esempio 1:** «Un modello di business basato sulla circolarità implica la massima collaborazione tra tutti gli attori; per questo riteniamo fondamentale aprirci al confronto con i soggetti che condividono la nostra visione, coinvolgendo le filiere e promuovendo iniziative comuni per salvaguardare le risorse naturali e accrescere la competitività del Paese.» - ENEL SPA, 2019
- **Esempio 2:** «Quella di Montepacini è un'esperienza pluriennale di collaborazione pubblico-privato (TOD'S), finalizzata al pieno esercizio dei diritti delle persone disabili e fragili, che coinvolge volontari, associazioni e persone impegnate nelle tematiche relative a biodiversità, filiera corta, sostenibilità, solidarietà e valorizzazione delle produzioni locali "buone, pulite, giuste e per tutti".» - TOD'S SPA, 2020
- **Esempio 3:** «Scatole e vassoi: anche per tale materiale di imballaggio, il Gruppo ha avviato una politica di acquisto orientata all'utilizzo di materiali riciclati, raggiungendo percentuali di assoluto interesse: infatti, nel 2020 oltre il 72% delle scatole e vassoi utilizzati sono stati prodotti con materiale riciclato.» - ZIGNAGO VETRO SPA, 2020

Tutti e tre gli esempi possono essere considerati come *asserzioni ambientali*, poiché ciascuno collega l'azienda a pratiche che suggeriscono un impatto positivo sull'ambiente: nel primo esempio, si fa riferimento alla circolarità e alla promozione di iniziative per la salvaguardia delle risorse naturali; il secondo esempio menziona l'impegno in tematiche come la biodiversità e la sostenibilità; infine, il terzo esempio focalizza l'attenzione sull'adozione di materiali riciclati da parte dell'azienda. Per arricchire l'annotazione, sono state identificate, ove possibile, alcune caratteristiche linguistiche associate all'uso della vaghezza in relazione a cinque categorie semantiche: *degree*, *quantity*, *category*, *time* e *softening stance-taking*. Queste categorie sono state selezionate combinando le analisi di Li [13], che esplora le categorie *degree*, *quantity*, *category*, *time* in testi legislativi, e di Jin [10], che identifica le categorie *degree*, *quantity*, *time* e *softening stance-taking* su CSR reports. Di seguito vengono fornite descrizioni per ciascuna categoria, illustrate attraverso gli esempi sopracitati.

1. *Degree*: vengono inseriti in questo gruppo gli utilizzi di aggettivi e avverbi ambigui o di quei termini il cui significato dipende fortemente dal contesto e che necessiterebbero quindi di essere spiegati o esplicitati. Esempi come “massima” e “salvaguardare” (Esempio 1), “buone, pulite, giuste” (Esempio 2), “di assoluto interesse” (Esempio 3),

rientrano in questa categoria poiché non rispondono chiaramente alle domande “come? / in che senso?”.

2. *Quantity*: questa categoria comprende espressioni che indicano approssimazione o quantità non specificate. Esempi tipici includono “oltre il 72%” (Esempio 3), “circa” oppure espressioni come “percentuali di assoluto interesse” (Esempio 3). Inoltre, frasi come “che coinvolge volontari” (Esempio 2) o espressioni generiche come “ridurre/riduzione/minore impatto” senza specificare la quantità rientrano in questa categoria, poiché non rispondono alle domande “quanto? / quanti? / di quanto?”.
3. *Category*: espressioni come “le risorse naturali” (Esempio 1), “persone disabili e fragili” (Esempio 2) e “materiali/e riciclati/o” (Esempio 3) sono associati ad un senso di vaghezza di questo tipo in quanto sembrano riferirsi ad una categoria di entità piuttosto che un oggetto specifico.
4. *Time*: fanno parte di questa categoria espressioni temporali generiche come “pluriennale” (Esempio 2), “nel tempo” e “da anni”, che mancano di precisione lasciando non detta la durata o la frequenza esatta delle azioni descritte.
5. *Softening stancetaking*: in questa categoria rientrano espressioni che attenuano la presa di posizione, rendendo l’affermazione meno diretta e più negoziabile, ad esempio l’espressione “riteniamo fondamentale aprirci al confronto” (Esempio 1) rappresenta un modo meno assertivo di affermare che l’azienda “si confronta”, lasciando spazio a dubbi sull’effettiva azione intrapresa.

5. Risultati

La Tabella 3 riporta i risultati dell’annotazione di un *sample* di 100 frasi, proveniente da 65 report distinti, prodotti da 33 aziende diverse. Va notato che il campione è leggermente sbilanciato, poiché quattro aziende (Salvatore Ferragamo Spa, Pirelli & C. Spa, Enel Spa e Iren Spa) contribuiscono da sole a un terzo delle frasi analizzate.

		vaghezza	quantity	degree	time	category	softening stance-taking
<i>enunciato</i>	20	9	7	3	1	1	1
<i>asserzione ambientale</i>	80	69	32	36	6	15	4
Totale complessivo	100	78	39	39	7	16	5

Tabella 3 Tabella riassuntiva dei risultati dell’annotazione di un campione di frasi.

L’ipotesi che i termini ricercati riuscissero ad individuare molti casi di asserzioni ambientali vaghe è stata confermata: l’80% delle frasi è stato annotato come asserzioni ambientali, complessivamente il 78% delle frasi ha mostrato espressioni vaghe, mentre il 69% delle frasi è stato annotato come potenziale caso di *greenwashing*. La distribuzione disomogenea tra le diverse categorie semantiche può essere attribuita alla selezione delle parole chiave durante la fase di estrazione delle frasi. In particolare, si è osservata una prevalenza di casi riconducibili alle

categorie *quantity* e *degree* (39% e 39%), seguiti da alcuni esempi appartenenti alla categoria *category* (16%) e, infine, pochissimi casi rientranti nelle categorie *time* (7%) e *softening stance-taking* (5%).

6. Sviluppi futuri

Poiché la condivisione dei dati è essenziale per facilitare la riproducibilità dei risultati e stimolare ulteriori ricerche, il dataset annotato utilizzato in questo studio è stato pubblicato sulla piattaforma Zenodo.¹⁵

Tra le prospettive future, ci si propone prima di tutto di ampliare la lista dei termini ricercati. L'aggiunta di pattern linguistici e termini di natura vaga, come quelli che hanno determinato il grado e la tipologia di vaghezza negli esempi citati (ad esempio “di assoluto interesse”, “circa”, “pluriennale”), permetterà di ottenere una panoramica più estesa sulle strategie di vaghezza impiegate all'interno dei report. Inoltre, questa espansione consentirà di identificare eventuali termini legati all'ambiente che potrebbero essere stati esclusi nelle ricerche precedenti.

Un altro aspetto cruciale riguarda il miglioramento del processamento del testo nei file PDF per consentire un'analisi linguistica più precisa. Attualmente, l'uso di Sketch Engine presenta alcune limitazioni nel trattare i report aziendali, che spesso si distinguono per una struttura complessa, ricca di titoli, didascalie ed elementi grafici che separano i blocchi di testo: tali complessità portano ad una suddivisione imprecisa delle frasi, specialmente in assenza di punteggiatura, e ad una gestione non sempre ottimale di colonne di testo separate. Inoltre, gli strumenti e la metodologia impiegati non permettono di contestualizzare adeguatamente le espressioni vaghe all'interno dei documenti, né di distinguere efficacemente tra frasi inserite in paragrafi ampi e quelle isolate in titoli o didascalie. Per affrontare queste sfide, sarà necessario considerare l'adozione di tecniche di *pre-processing* del testo più avanzate, capaci di migliorare la gestione delle strutture complesse tipiche dei report aziendali. L'integrazione di tali tecniche potrebbe non solo affinare la suddivisione del testo, ma anche permettere un'analisi più approfondita e accurata, che tenga conto delle differenti funzioni comunicative ricoperte dalle frasi all'interno dei documenti.

Un'ulteriore limitazione del presente studio risiede nell'annotazione eseguita da un solo annotatore e nella dimensione ridotta del dataset, fattori che possono introdurre un potenziale *bias* e limitare la generalizzabilità dei risultati. Per migliorare la qualità e l'affidabilità del dataset, sarà previsto il coinvolgimento di un numero maggiore di annotatori, insieme all'implementazione di un protocollo di annotazione più dettagliato e all'ampliamento del numero di frasi selezionate.

7. Conclusioni

L'analisi condotta ha evidenziato come le parole chiave selezionate per la ricerca di esempi di *asserzioni ambientali* vaghe siano utilizzate frequentemente all'interno del corpus come strategia comunicativa. Tale pratica si manifesta attraverso l'adozione di termini generici come “sostenibile” o “ecologico”, che raramente vengono supportati da dettagli concreti e verificabili

¹⁵ Cutuli, Erica. «Dataset di dichiarazioni ambientali vaghe da report di sostenibilità». Zenodo, 21 gennaio 2025. <https://doi.org/10.5281/zenodo.14712579>.

all'interno dell'immediato contesto discorsivo. Questo approccio lascia poco chiare le dichiarazioni aziendali, rendendo difficile identificare con precisione le posizioni, le attività e i progressi effettivi in ambito ambientale, con potenziali conseguenze sulla trasparenza delle dichiarazioni aziendali e sulla fiducia di consumatori e stakeholder.

Una delle strategie più diffuse all'interno del campione analizzato consiste nell'utilizzo di verbi come "ridurre", "aumentare" o aggettivi come "minore", "maggiore", "ridotto", spesso privi di riferimenti quantitativi chiari; anche quando sono fornite delle quantità o cifre, queste vengono spesso accompagnate da approssimazioni come "circa", "fino a", "oltre", che riducono comunque la chiarezza e la verificabilità delle informazioni. In tal modo, le aziende possono presentare i propri (ipotetici) impatti positivi sull'ambiente, senza però offrire una reale misura di tali impatti, rendendo le informazioni difficilmente comparabili e verificabili. Inoltre, questo meccanismo lascia ampi spazi interpretativi e consente alle aziende di proiettare un'immagine positiva senza esporsi a responsabilità concrete.

I risultati dello studio contribuiscono non solo a una comprensione più approfondita delle pratiche linguistiche associate al *greenwashing*, ma pongono anche le basi per interventi futuri. La creazione di un dataset annotato in lingua italiana rappresenta un progresso significativo verso l'automatizzazione dell'individuazione di tali fenomeni attraverso modelli di intelligenza artificiale. Questo non solo colma un vuoto nella ricerca focalizzata sul contesto italiano, ma contribuisce anche al dibattito più ampio sull'uguaglianza linguistica digitale.

Infine, il presente lavoro si propone come uno strumento utile per promuovere pratiche comunicative più trasparenti e in linea con le normative europee, offrendo un supporto critico alle autorità regolatorie e ai consumatori nella lotta contro il *greenwashing*. In prospettiva, la standardizzazione di metodologie per l'identificazione e l'analisi della vaghezza nei report di sostenibilità potrà favorire una maggiore *accountability* delle aziende, contribuendo a orientare le loro attività verso una sostenibilità concreta e misurabile.

Riferimenti

- [1] Bernini, Francesca, e Fabio La Rosa, 2024. «Research in the Greenwashing Field: Concepts, Theories, and Potential Impacts on Economic and Social Value». *Journal of Management and Governance* 28, fasc. 2 (giugno 2024): 405–44. <https://doi.org/10.1007/s10997-023-09686-5>.
- [2] Bingler, Julia Anna, Mathias Kraus, Markus Leippold, e Nicolas Webersinke, 2024. «How Cheap Talk in Climate Disclosures Relates to Climate Initiatives, Corporate Emissions, and Reputation Risk». *Journal of Banking & Finance* 164 (luglio 2024): 107191. <https://doi.org/10.1016/j.jbankfin.2024.107191>.
- [3] Channell, Joanna, 1994. *Vague Language*. Describing English Language. Oxford New York Toronto: Oxford University Press, 1994.
- [4] Chen, Minghui, 2018. «Analysis of Vagueness in English Advertisement from the perspective of Adaption Theory.» *International Journal of Social Science and Economic Research* 3, n. 3 (2018): 1068-1086.
- [5] Commissione europea, 2023. «Proposta di direttiva del Parlamento europeo e del Consiglio sull'attestazione e sulla comunicazione delle asserzioni ambientali

- esplicite (direttiva sulle asserzioni ambientali).» *COM/2023/166 final*. 22 marzo 2023.
- [6] Commissione europea, 2021. «Screening of websites for ‘greenwashing’: half of green claims lack evidence.» *Press release*. Brussels, 28 gennaio 2021.
- [7] Cutting, Joan, a c. di. *Vague language explored*, 2007. Basingstoke [England] ; New York: Palgrave Macmillan, 2007.
- [8] De Freitas Netto, Sebastião Vieira, Marcos Felipe Falcão Sobral, Ana Regina Bezerra Ribeiro, e Gleibson Robert Da Luz Soares, 2020. «Concepts and Forms of Greenwashing: A Systematic Review». *Environmental Sciences Europe* 32, fasc. 1 (dicembre 2020): 19. <https://doi.org/10.1186/s12302-020-0300-3>.
- [9] Hoepner, Andreas, 2024. «Too good to be true: the greenwashers’ box of tricks.» *Financial Times*. 16 maggio 2024. <https://www.ft.com/content/78b3c741-1ab8-48f5-92a8-4e98dfa230ab> (consultato il giorno giugno 22, 2024).
- [10] Jin, Bixi, 2022. «A corpus-assisted study of vague language in corporate responsibility reports of the cosmetics industry.» *Iberica* 43 (2022): 77-102. <https://doi.org/10.17398/2340-2784.43.77>.
- [11] Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, e Vít Suchomel, 2014. «The Sketch Engine». *Lexicography* 1, fasc. 1 (1 luglio 2014): 7–36. <https://doi.org/10.1007/s40607-014-0009-9>.
- [12] Lane, Eric L., 2013. «Greenwashing 2.0.» *Columbia Journal of Environmental Law* 38, n. 2 (2013). <https://ssrn.com/abstract=2089078>.
- [13] Li, Shuangling, 2017 «A Corpus-Based Study of Vague Language in Legislative Texts: Strategic Use of Vague Terms». *English for Specific Purposes* 45 (gennaio 2017): 98–109. <https://doi.org/10.1016/j.esp.2016.10.001>.
- [14] Lombardi Vallauri, Edoardo, 2022. «Implicite e persuasione.» Cap. 11 in *Pragmatica sperimentale*. Manuali Linguistica, a cura di Filippo Domaneschi e Valentina Bambini, 157-170. Bologna: Il Mulino, 2022.
- [15] Lombardi Vallauri, Edoardo, Federica Cominetti, e Viviana Masia, 2022. «The Persuasive and Manipulative Power of Implicit Communication». *Journal of Pragmatics* 197 (agosto 2022): 1–7. <https://doi.org/10.1016/j.pragma.2022.04.015>.
- [16] Mannaioli, Giorgia, 2023. «LA VAGHEZZA LINGUISTICA COME STRATEGIA IMPLICITA PERSUASIVA E FENOMENO DI INTERFACCIA TRA SEMANTICA, SINTASSI E PRAGMATICA. ESEMPI DAL DISCORSO PUBBLICO CONTEMPORANEO IN DIVERSE LINGUE». *Italiano LinguaDue* 15, fasc. 2 (15 dicembre 2023): 206–21. <https://doi.org/10.54103/2037-3597/22028>.
- [17] Mannaioli, Giorgia, 2025. *Vagueness As an Implicitating Persuasive Strategy*. 1st ed. Pragmatics and Beyond New Series, v. 350. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2025.

- [18] Mannaioli, Giorgia, Alessandro Ansani, Claudia Coppola, e Edoardo Lombardi Vallauri, 2024. «Vagueness as an Implicit-Encoding Persuasive Strategy: An Experimental Approach». *Cognitive Processing* 25, fasc. 2 (maggio 2024): 205–27. <https://doi.org/10.1007/s10339-023-01171-z>.
- [19] Moodaley, Wayne, e Arnesh Telukdarie, 2023. «A Conceptual Framework for Subdomain Specific Pre-Training of Large Language Models for Green Claim Detection». *European Journal of Sustainable Development* 12, fasc. 4 (1 ottobre 2023): 319. <https://doi.org/10.14207/ejsd.2023.v12n4p319>.
- [20] Moodaley, Wayne, e Arnesh Telukdarie, 2023. «Greenwashing, Sustainability Reporting, and Artificial Intelligence: A Systematic Literature Review». *Sustainability* 15, fasc. 2 (12 gennaio 2023): 1481. <https://doi.org/10.3390/su15021481>.
- [21] Moruzzi, Sebastiano, 2012. *Vaghezza: confini, cumuli e paradossi*. Filosofia / [Laterza]. Roma ; Bari: Laterza, 2012.
- [22] Nemes, Noémi, Stephen J. Scanlan, Pete Smith, Tone Smith, Melissa Aronczyk, Stephanie Hill, Simon L. Lewis, A. Wren Montgomery, Francesco N. Tubiello, e Doreen Stabinsky, 2022. «An Integrated Framework to Assess Greenwashing». *Sustainability* 14, fasc. 8 (8 aprile 2022): 4431. <https://doi.org/10.3390/su14084431>.
- [23] Pinkal, Manfred, 1996. «Vagueness, Ambiguity, and Underspecification». *Semantics and Linguistic Theory* 6 (15 agosto 1996): 185. <https://doi.org/10.3765/salt.v6i0.2775>.
- [24] Russell, Bertrand, 1923. «Vagueness». *Australasian Journal of Psychology and Philosophy* 1, fasc. 2 (giugno 1923): 84–92. <https://doi.org/10.1080/00048402308540623>.
- [25] Ruzaitė, Jūratė, 2007. *Vague language in educational settings: quantifiers and approximators in British and American English*. Europäische Hochschulschriften. Reihe XIV, Angelsächsische Sprache und Literatur ; European university studies. Series XIV, Anglo-Saxon language and literature ; Publications universitaires européennes. Serie XIV, Langue et littérature anglo-saxonnes, Bd. 433 = v. 433 = v. 433. Frankfurt am Main ; New York: Lang, 2007.
- [26] Stammbach, Dominik, Nicolas Webersinke, Julia Bingler, Mathias Kraus, e Markus Leippold, 2023. «Environmental Claim Detection». In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 1051–66. Toronto, Canada: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.acl-short.91>.
- [27] Webersinke, Nicolas, Mathias Kraus, Julia Bingler, e Markus Leippold, 2022. «CLIMATEBERT: A Pretrained Language Model for Climate-Related Text». *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4229146>.
- [28] Willis, J., 2023. «The Greenwashing Hydra». *Planet Tracker*. January 2023.