Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Ingegneria dell'Informazione

Corso di laurea in Ingegneria Informatica

Bias di genere nei dataset utilizzati per l'addestramento di assistenti virtuali

Candidato:
Fabio Cociancich

Relatore:
Prof. Antonio Rodà
Correlatrice:
Prof.ssa Silvana Badaloni

Anno Accademico 2022/2023
Data di laurea: 28 settembre 2023

Abstract

Questa tesi si propone di esplorare il fenomeno dei bias di genere all'interno dei dataset utilizzati per l'addestramento di assistenti virtuali. Gli assistenti virtuali, come Siri, Alexa e Google Assistant, sono diventati parte integrante della nostra vita quotidiana, fornendo supporto e risposte ai nostri quesiti. Tuttavia, c'è una crescente preoccupazione che tali assistenti siano affetti da bias di genere.
Il presente studio si concentra sull'analisi del dataset MASSIVE al fine di identificare eventuali pregiudizi impliciti di genere. Le analisi eseguite sul dataset sono basate su quelle condotte dallo studio di Seaborn "Transcending the "male code": Implicit masculine biases in nlp contexts" [1], infatti sono state realizzate in maniera simile ma considerando solamente gli enunciati in lingua italiana. Oltre alle evidenze di bias impliciti maschili, un altro contributo consiste in AVA (*Ambiguity for Virtual Assistants*): un dizionario che raccoglie i termini ambigui comuni al linguaggio di genere e al linguaggio degli assistenti virtuali.
Dopo i primi due capitoli introduttivi la tesi si concentra sull'analisi del dataset MASSIVE. Il terzo e quarto capitolo si concentrano sulla descrizione dei vari dizionari usati e sul dataset MASSIVE. Il quinto e sesto capitolo sono incentrati sulle prime analisi effettuate sul dataset. I due capitoli successivi descrivono la creazione del dizionario AVA e trattano le analisi eseguite utilizzando tale dizionario. Infine è riportata una discussione sulle future analisi e ricerche, sulle limitazioni di questa ricerca e le conclusioni finali.

Indice

Capitolo 1

Introduzione

1.1 Introduzione agli assistenti virtuali

Le macchine ora riescono a comunicare in maniera naturale con noi. I progressi negli ambiti del Machine Learning (ML) e nel Natural Language Processing (NLP) hanno spianato la strada a metodi di comunicazione più naturali con i computer.
L'utilizzo sempre maggiore in questi ultimi anni di assistenti vocali dotati di intelligenza artificiale ha sicuramente modificato il nostro modo di interagire con la tecnologia. Al giorno d'oggi nelle nostre abitazioni sono presenti assistenti virtuali di ogni tipo: negli smartphones, in dispositivi come Amazon Alexa, Google Nest, Cortana, smart TV. Il maggiore utilizzo di queste nuove tecnologie ha anche evidenziato come queste ultime possano riprodurre e diffondere stereotipi negativi, i quali possono essere relativi a: genere, età, etnia ed intersezioni di queste.
L'attenzione della comunità scientifica è aumentata in questi anni e sono stati pubblicati alcuni report, come quello dell'UNESCO "I'd Blush If I Could"[2], i quali sottolineano come gli assistenti vocali possano diffondere pregiudizi di genere, come quello per cui le donne dovrebbero ricoprire ruoli subordinati.
Risulta quindi di grande interesse cercare di misurare in maniera precisa i possibili pregiudizi insiti negli assistenti virtuali per poter successivamente ridurli o eliminarli il più possibile.
Molto lavoro è stato fatto in questi ultimi anni, ponendo particolare attenzione su alcuni ambiti: la ricerca e la correzione di stereotipi nei dataset usati per allenare sistemi di machine learning, con una speciale attenzione ai bias verso le donne e la femminilità [3][4]; il tracciamento della diffusione della tossicità e del linguaggio abusivo (specialmente in relazione alla misoginia) [5] [6] ; la rimozione di bias dagli algoritmi di Machine Learning e dai dataset nel contesto dei word embeddings [7] [3] [4].
Citando lo studio che verrà approfondito nella prossima sezione "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings" [7], le analogie trovate analizzando i word embeddings mostrano come questi strumenti possano evidenziare pregiudizi di genere.

Tuttavia, per affrontare il tema degli stereotipi di genere è necessario considerare una visione più ampia dei generi, ad esempio considerando il maschile come genere dominante.
È cruciale affrontare le discriminazioni e la tossicità, ma è altrettanto importante analizzare le varietà di discriminazioni di genere più subdole che possono apparire nella scelta delle parole o nelle assunzioni di genere.
Il centro dell'attenzione è stato incentrato sulle tendenze sociali a centrare la mascolinità, in maniera conscia o meno, nel linguaggio. Questo può avvenire facendo determinate scelte nelle parole usate, nell'uso del maschile come norma, o nella sovra rappresentazione di riferimenti maschili.
La domanda che si sono posti gli autori della ricerca [1] è la seguente:
"I pregiudizi impliciti del maschile in termini del linguaggio di genere e del linguaggio maschile come norma, in particolare l'uso di pronomi e di nomi, di parole marcate in base al genere, esistono nel contesto dei dataset creati per allenare gli assistenti virtuali e in che misura?"

Per rispondere a questa domanda è stata condotta un'analisi del dataset MASSIVE [8] utilizzando dizionari basati sul linguaggio di genere.
Il contributo apportato consiste in:

- evidenza di linguaggio di genere, soprattutto bias impliciti maschili, nel dataset MASSIVE
- AVA, un nuovo dizionario che raccoglie i termini ambigui nell'ambito del linguaggio di genere e nell'ambito degli assistenti virtuali

Capitolo 2

Lavori correlati

Il linguaggio umano è affetto da pregiudizi, e i dataset basati su quest'ultimo non fanno eccezione.
È sempre più presente un interesse nell'ambito del NLP nel riconoscere e agire nei confronti degli eventuali bias di genere o di altro tipo. Alcuni lavori hanno l'obiettivo di localizzare e analizzare l'estensione dei pregiudizi nei dataset per NLP, nello sviluppare metodi per ridurre o eliminare tali bias dai dataset e dagli algoritmi e altre alternative creative.
Tuttavia sono presenti delle carenze in questi lavori, come ad esempio la mancanza di rappresentazione dei generi non binari e di mascolinità.
Quando bisogna sviluppare dataset per NLP, spesso bisogna trovare un compromesso tra il realismo dei dialoghi e l'evitare i bias di genere. In questi casi, solitamente si usano metodi di "crowdsourcing" come Amazon Mechanical Turk (AMT) [9] per raccogliere, annotare e valutare dataset.
Così facendo si possono ottenere dataset di grandi dimensioni e in un linguaggio naturale, ma possono contenere dei pregiudizi di genere.
Bisogna inoltre considerare come molti dataset provengano da nazioni occidentali, ricche, industrializzate e democratiche. Per questo motivo una varietà di lingue e culture sono sottorappresentate in questi dataset. Ultimamente sono stati fatti degli sforzi per tradurre e sviluppare dataset in altre lingue, un esempio è il dataset MASSIVE [10] [8]. Bisogna considerare come i pregiudizi presenti nei dataset possano influenzare un ampio sistema di interazioni con gli assistenti virtuali. Infatti non è chiaro quanto siano stati presi in considerazione i pregiudizi di genere durante la creazione di dataset NLP. Nel paper relativo al dataset MASSIVE [10] non c'è alcuna discussione relativa ai bias. Se non si ha nessuna discussione relativa ai pregiudizi, la possibilità di codificarli o di introdurne di nuovi è elevata.
Attualmente, la parte più consistente del lavoro ha analizzato i pregiudizi di genere come associazioni di genere stereotipate nell'ambito del NLP. Specialmente nell'ambito dei word embeddings sono stati condotti studi relativi ai ruoli sociali, alle occupazioni [7] [11] [4] e alla mascolinità tossica, caratterizzata da comportamenti misogini e

aggressività [5] [6] [4].

Sono state condotte alcune ricerche dove si è tentato di affrontare e risolvere un problema più subdolo, ovvero quello dei bias impliciti nell'ambito del machine learning.
Un campo di ricerca che è stato approfondito riguarda l'ambito degli stereotipi nei confronti delle donne, una soluzione già testata è quella di utilizzare riferimenti femminili al posto di quelli maschili e vice versa per cercare di rimuovere tali bias. [4] Questa tecnica ha comunque delle carenze in quanto esclude i generi non binari.
Quando si parla di tossicità, rimuovere i contenuti può risultare un processo delicato. Rimuovere espressioni di odio potrebbe portare ad un agente non allenato per rispondere in determinate circostanze.
Allenare gli assistenti virtuali comprende anche dare loro la possibilità di capirci.
Devono essere perciò allenati utilizzando diversi dataset per massimizzare l'inclusività.
Uno degli studi più interessanti per quanto riguarda la rimozione di bias è quello descritto nella prossima sezione.

2.0.1 Termini utili

Una parte importante del lavoro legato ai bias di genere è stato incentrato sulle forme negative di mascolinità.
La mascolinità tossica indica l'adesione a ruoli di genere maschile tradizionali in cui gli uomini trattengono le emozioni ed evitano comportamenti considerati poco mascolini e si esprimono in modo aggressivo.
La misoginia è un concetto collegato, si riferisce all'odio nei confronti delle donne e/o delle femminilità. È una forma estrema di sessimo nei confronti del genere femminile.
Il sessimo è un concetto più generico, include stereotipi secondo i quali le donne tendono ad essere più fragili emotivamente rispetto agli uomini e altre discriminazioni, come quelle nei confronti delle persone con identità non binarie.
Solitamente la tossicità e il sessimo sono analizzati assieme nei lavori nell'ambito delle NLP, anche se non tutte le forme di tossicità sono legate al genere.
La maggior parte del lavoro fatto in questo ambito si concentra sugli stereotipi, definiti come opinioni precostituite e generalizzate su persone o gruppi sociali.
Come già discusso, l'utilizzo di riferimenti femminili al posto di quelli maschili e viceversa [4] come tecnica di rimozione di bias è una strategia già ampiamente utilizzata, ma è limitata ad un contesto di generi binari.
I lavori riguardanti il NLP e ambiti simili hanno analizzato: manifestazioni di mascolinità tossica che devono essere rimosse e la svalutazione della femminilità. Tuttavia sono presenti forme più subdole di pregiudizi maschili.

2.0.2 Bias maschili impliciti

Sin dall'antichità la società ha adottato un punto di vista maschile nella propria visione del mondo, marginalizzando o ignorando il punto di vista femminile. Questo fenomeno si chiama "androcentrismo", "maschile come norma", "maschile come default".
In questa ricerca useremo i termini "bias maschili impliciti" o "pregiudizi maschili impliciti" per indicare il fenomeno precedentemente descritto.
Il fenomeno dei pregiudizi maschili impliciti è il riflesso di come funzionano le gerarchie nella società, dove gli uomini e la mascolinità sono posizionati in cima. Questa dinamica è comunemente conosciuta come sistema patriarcale.
Gli uomini sono considerati come detentori del potere e la società è organizzata attorno a questa assunzione. Uno studio sui word embeddings mostra come le parole "persona" e "persone" siano più strettamente collegate ai termini "uomo" e "uomini".
Possiamo anche non usare esplicitamente parole maschili, ma tendiamo a identificare l'umanità come mascolina in altre maniere, ad esempio nella scelta dei termini quando comunichiamo.
Nelle lingue romanze, come il francese e l'italiano, si tende ad utilizzare termini maschili come se fossero neutri in relazione al genere (per es. "uomini"), questo fenomeno prende il nome di "maschile non marcato".
Questo utilizzare vocaboli maschili come termini neutri può avere delle ripercussioni: le persone tendono ad assegnare un genere o un riferimento maschile quando vengono presentati loro dei termini neutri. In contrasto, i riferimenti femminili sono usati per inserire la "donna" come "altro", come un caso speciale, distinto e opposto all'uomo.
È importante analizzare la frequenza dei riferimenti maschili, includendo nomi con genere, pronomi maschili. Per questo esaminerò i bias impliciti maschili analizzando il maschile come norma.
Un altro elemento importante è quello del linguaggio di genere, un utilizzo implicito del linguaggio virtualmente inesplorato nell'ambito del NLP. Il linguaggio di genere è l'idea che la scelta stessa delle parole sia influenzata dal genere, perciò le persone tendono a utilizzare determinati vocaboli in base alla loro identità e alle loro caratteristiche. Gli uomini, per esempio, tendono ad utilizzare un linguaggio più dominante per esprimere la mascolinità in termini di autorità e di potere.
Una ricerca di Gaucher D. [12] effettuata analizzando vari annunci di offerte di lavoro mostra come gli annunci relativi a professioni tipicamente maschili utilizzino un linguaggio di genere maschile, contrariamente a quello che accade negli annunci relativi a professioni con una maggiore percentuale femminile.
Sono presenti altre caratteristiche che possono influenzare il linguaggio che utilizziamo, come l'appartenenza politica ad esempio. Una ricerca di Roberts e Utych [13] evidenzia come i presidenti repubblicani degli Stati Uniti d'America tendano a usare un linguaggio più mascolino rispetto ai presidenti democratici.
Nell'ambito delle interazioni uomo-computer sono presenti sempre più assistenti vir-

tuali, personaggi di videogiochi e robot che utilizzano questi dataset per parlare.
Sappiamo che questi agenti possono riflettere i pregiudizi presenti nei dataset da loro utilizzati, serve quindi impegno per creare dei dataset che siano affetti il meno possibile da bias. È possibile che i dataset più nuovi, come MASSIVE, abbiamo una minore quantità di pregiudizi se paragonati a dataset meno recenti come ReDial (almeno per quanto riguarda la lingua inglese).
È fondamentale quindi agire ora per evitare di avere una diffusione di agenti affetti da pregiudizi e stereotipi.

2.1 Debiasing Word Embeddings

Per quanto riguarda la rimozione di bias dai word embeddings uno studio da citare è "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings" [7].
In questo studio si mettono in evidenza le problematiche legate al sessismo e si propone un metodo di mitigazione di tali problemi.
Un word embedding rappresenta ogni parola (o frase) w come un vettore d-dimensionale $\overrightarrow{w} \in \mathbb{R}^d$. Funziona come una sorta di dizionario per programmi, i quali hanno la necessità di essere a conoscenza del significato delle parole e dei loro legami e similitudini.
In questo spazio vettoriale i vettori relativi a termini semanticamente simili tendono ad essere raggruppati a breve distanza. Si è notato come la differenza vettoriale tra vettori nei word embeddings tenda a rappresentare le relazioni tra parole.
In un esempio con analogie, "l'uomo sta al principe come la donna sta a x" (scritto in equazione $uomo : principe :: donna : x$), un semplice calcolo vettoriale usando i vettori del word embeddings trova che $x = principessa$ è la soluzione migliore in quanto $\overrightarrow{uomo} - \overrightarrow{donna} \approx \overrightarrow{principe} - \overrightarrow{principessa}$.
In un esempio simile, si trova $x = Italia$ per l'analogia $Parigi : Francia :: Roma : x$.
Questi semplici calcoli vettoriali possono mostrare una varietà di analogie. Per questo motivo i word embeddings sono uno strumento dotato di grandi potenzialità. I word embeddings sono attualmente utilizzati e studiati in svariate applicazioni come l'ordinamento dei documenti, l'analisi dei sentimenti e il reperimento di domande. Inoltre, i word embeddings sono in grado di individuare anche la presenza di sessismo nei testi.
Per fare un esempio:

$$\overrightarrow{uomo} - \overrightarrow{donna} \approx \overrightarrow{programmatore} - \overrightarrow{casalinga}$$

Quindi il sistema risponderà $x = casalinga$ nella analogia

$$\overrightarrow{uomo} - \overrightarrow{donna} \approx \overrightarrow{programmatore} - \overrightarrow{x}$$

In un esempio simile si può ottenere come risposta:

$$\overrightarrow{padre} - \overrightarrow{madre} \approx \overrightarrow{dottore} - \overrightarrow{infermiera}$$

Nello studio sopracitato [7] si dimostra che gli stereotipi contenuti nei word embeddings riflettono quelli già presenti nella società, e possono essere addirittura amplificati. Per rimuovere i bias di genere dai word embeddings il primo passo è quello di identificare una direzione, o un sottospazio, degli embeddings che catturino i bias. Successivamente ci si assicura che non ci siano parole di genere neutre nel sottospazio di genere e si equalizzano set di parole fuori da tale sottospazio rendendo così le parole neutre equidistanti rispetto alle parole in ciascun insieme di uguaglianza.
Per fare un esempio, se {nonno, nonna} e {ragazzo, ragazza} sono due insiemi di uguaglianza, allora dopo il procedimento descritto "babysitter" dovrebbe essere equidistante da "nonno" e "nonna" ma anche da "ragazzo" e "ragazza", anche se probabilmente sarà più vicino ai primi due termini rispetto agli ultimi due.
Il metodo appena descritto è abbastanza invasivo, perciò gli autori hanno proposto un altro metodo per ridurre le differenze tra questi set, cercando di mantenere la massima similarità con l'embedding originale, utilizzando una variabile per gestire questo compromesso.

Capitolo 3

Dizionari

3.1 Dizionari utilizzati

Per ripetere le analisi eseguite nella ricerca [1] ho dovuto utilizzare gli stessi dizionari usati dagli autori adattandoli alla lingua italiana.
I dizionari descritti di seguito sono stati utilizzati per eseguire le analisi sulla frequenza dei termini nel dataset MASSIVE in lingua italiana.
I due dizionari adottati dagli autori, ovvero Gaucher [12] e Roberts & Utych [13], sono stati scelti in quanto basati sul linguaggio di genere.
Sono state fatte diverse considerazioni riguardo la rilevanza degli obiettivi di questa analisi (ad esempio: valuta il linguaggio maschile?), la qualità (ad esempio: è stato creato in maniera rigorosa?) e le recensioni di esperti e accademici.
I due dizionari sono diversi per il contesto nel quale sono stati creati, la data e il metodo di creazione, il formato e il numero di termini. Hanno in comune la distinzione binaria dei generi.

Citazione	Anno	Generi	Fonti	Metodo di categorizzazione	Formato e interpretazione	Totale
Roberts & Utych [13]	2019	maschile, femminile, neutrale	ANEW, sinonimi curati	AMT [9]	scala da 1 a 7, dove fem. ≤ 3 e masch. ≥ 5, fem. stretto < 2.5 e masch. stretto > 5.5	3450(loose), 485 (strict)
Gaucher et al. [12]	2011	maschile, femminile	annunci lavorativi	Autori (manualmente)	lista delle due categorie (maschile e femminile)	2333

Anno : pubblicazione della ricerca

Totale : numero di termini dopo la traduzione in italiano

3.2 Gaucher

Il dizionario Gaucher deriva dalla ricerca "Evidence That Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality". [12]
L'obiettivo della ricerca [12] era quello di evidenziare se un utilizzo del linguaggio di ge-

nere potesse mantenere alcune disuguaglianze di genere presenti nell'ambito lavorativo, nello specifico la sottorappresentazione femminile in professioni tipicamente maschili.
In tale ricerca, è stata analizzato se fossero presenti termini legati a stereotipi maschili (come "competitivo", "dominante", "leader") negli annunci lavorativi relativi a occupazioni tipicamente maschili.
É stato quindi valutato qualora la presenza di questi termini potesse dissuadere le donne dal candidarsi per quell'impiego, magari pensando che non fosse adatto a loro.
Per poter eseguire queste analisi sono state create due liste di termini, le quali sono presenti nell'appendice della ricerca.
La lista di termini maschili consiste in parole trovate più frequentemente in annunci relativi a lavori prettamente maschili mentre la lista di termini femminili è composta da termini più frequenti negli annunci più scelti da donne.
Ho tradotto le parole presenti nella lista di parole maschili e femminili per creare le due liste utilizzate nelle successive analisi effettuate in lingua italiana.
Per la traduzione mi sono affidato a siti come deepl.com, translate.google.it, reverso.net e a dizionari come dictionary.cambridge.org. Ho utilizzato questi siti anche per le traduzioni del dizionario descritto successivamente .

3.3 Roberts and Utych

Il dizionario Roberts & Utych deriva dalla ricerca "Linking Gender, Language, and Partisanship: Developing a Database of Masculine and Feminine Words". [13]
In questa pubblicazione si è cercato di analizzare i discorsi dei presidenti degli Stati Uniti d'America per valutare se fossero presenti delle differenze tra i presidenti appartenenti al partito politico repubblicano e quelli appartenenti al partito democratico. Si è notato come i presidenti repubblicani abbiano una tendenza a fare discorsi utilizzando un linguaggio di genere più maschile.
Per poter eseguire queste analisi è stata necessaria la creazione di una lista di termini considerati mascolini o femminili. Per questo motivo 175 partecipanti hanno dato dei voti da 1 a 7 ad alcuni termini in base alla loro percezione della "mascolinità" o "femminilità" di tale parola.
Un voto medio di 4 indica una neutralità della parola in questa misura unidimensionale della mascolinità (misura che potrebbe essere limitante). Un voto medio superiore al 5 indica una parola mascolina, mentre un voto medio superiore a 5.5 indica una parola fortemente mascolina. Un voto medio inferiore a 3 indica una parola femminile, mentre un voto medio inferiore a 2.5 indica una parola fortemente femminile.
In questa pubblicazione vengono poi create delle liste di parole maschili (63 termini), molto maschili (20 termini), femminili (31 termini) e molto femminili (11 termini) per analizzare e fare paragoni tra i discorsi dei politici statunitensi. Nelle analisi successive le liste di termini molto maschili o molto femminili saranno indicate come "strict",

"conservative" (talvolta abbreviate in "cons") oppure "strette".
Le liste maschili o femminili come "loose" o "allargate".
Similmente al dizionario Gaucher ho tradotto le parole delle quattro liste sopracitate per poter creare queste liste in lingua italiana, in modo da poter eseguire le analisi successive.

3.4 Morph-it!

Morph-It! [14] è un una lista di parole italiane liberamente scaricabile e utilizzabile. Consiste in un file txt contenente 500mila vocaboli. Ogni riga contiene la parola indicata, la parola dalla quale deriva (lemma) e dei tag che indicano le caratteristiche di tale termine.
In seguito sono inserite alcune righe di esempio del file txt:

```
statuetta statuetta NOUN-F:s
statuette statuetta NOUN-F:p
belli bello ADJ:pos+m+p
bello bello ADJ:pos+m+s
fui essere VER:ind+past+1+s
stato essere VER:part+past+s+m
```

Come si può notare negli esempi le varie parole sono catalogate in base alla parte del discorso a cui fanno parte (articolo, sostantivo, predicato, ...), in base al genere e se sono al singolare o al plurale.
Per avere altre informazioni sui tag delle parole e sul funzionamento di Morph-It! consultare il sito [15].

3.4.1 Utilizzo di Morph-it!

Ho utilizzato le espressioni regolari (regex) nei vari notebook Python per ottenere liste di parole adatte alle mie analisi.
Ad esempio, ho cercato tutte le righe contenenti il testo "NOUN" per poter creare un dizionario di sostantivi al fine di analizzare i sostantivi contenuti nel database MASSIVE. Inoltre ho creato alcune ricerche automatizzate per trovare tutte le parole derivate partendo da quella base (ad esempio cercando *bello* trovare *bello, bella, belli, belle*), per fare questa ricerca ho usato un semplice comando regex, ovvero

```
HT + parola_da_cercare + HT
```

dove HT corrisponde al carattere ASCII 9, che sta per "horizontal tab". Questo perchè nel dizionario Morph-it! le colonne vengono separate da tab orizzontali e la parola non derivata è presente nella colonna centrale.
Usando la stringa regex `HT + bello + HT` si trovano tutte le righe contenenti parole derivate da "bello".

Alcune righe sono:
`bella bello ADJ:pos+f+s`
`belle bello ADJ:pos+f+p`
Trovate le righe corrispondenti alle parole derivate rimane da isolare la parola derivata, è necessario perciò selezionare tutti i caratteri partendo dal primo della riga fino al carattere prima del "tab orizzontale".
Applicando questo procedimento alla parola "bello" si trovano i seguenti termini:
`['bei', "bell'", "bell'", 'bella', 'belle', 'belli', 'bellissima', 'bellissime', 'bellissimi', 'bellissimo', 'bello']`

Capitolo 4

Dataset MASSIVE

Con il termine Natural Language Understanding (NLU) si intende l'abilità di una macchina di comprendere il significato e le relative entità di un testo.
Dato l'enunciato "qual è la temperatura a Padova?" un modello NLU potrebbe classificare l'intent come `weather_query` e gli slot come
`weather_descriptor : temperature, place_name : Padova`

Una componente importante nell'ambito della NLU è la SLU (Spoken Language Understanding), nella quale una traccia audio viene convertita in testo prima che venga eseguita la NLU.
La SLU è fondamentale per gli assistenti virtuali come Alexa, Siri, Google Assistant.
Nonostante in questi ultimi anni gli assistenti virtuali siano stati migliorati e potenziati notevolmente, essi supportano comunque solamente una frazione delle più di 7000 lingue attualmente usate al mondo.
Una difficoltà nel creare modelli NLU che supportano più lingue consiste nella mancanza di dati per allenare e testare i modelli, specialmente enunciati che siano realistici per l'utilizzo finale del modello NLU.

Per sopperire a questa problematica è stato creato MASSIVE (Multilingual Amazon Slu resource package (SLURP) for Slot-filling, Intent classification, and Virtual assistant Evaluation), un dataset di enunciati in 52 lingue, utile per sviluppare modelli NLU in una varietà di linguaggi. [10]
È stato creato da traduttori professionisti, i quali hanno tradotto gli enunciati in inglese del dataset SLURP [16] in 51 lingue (inizialmente erano 50 ma nella versione 1.1 è stata aggiunta la lingua catalana).
MASSIVE contiene 1 milione di enunciati in 52 lingue [10].
Ho utilizzato questo dataset per analizzare l'eventuale presenza di pregiudizi di genere impliciti in quanto l'italiano è una delle lingue presenti.
Nell'ambito del NLP molti dataset sono composti di dati con vari tag.
Una foto di un gatto potrebbe avere una o più etichette, ad esempio "gatto", "felino", "adorabile" e così via.
Questi set di dati sono usati per allenare modelli di intelligenza artificiale. L'obiettivo

è quello di creare dei modelli capaci di riconoscere e catalogare dati nuovi, simili a quelli visti in allenamento ma non identici.
Per ottenere ciò si utilizzano tre partizioni: la partizione "train", quella "dev" e quella "test". La partizione train viene usata per allenare i modelli, la partizione dev tiene utilizzata per fare delle analisi dopo ogni ciclo di apprendimento sugli errori del modello, la partizione test contiene dati mai visti nelle partizioni precendenti che vengono utilizzati per testare il modello finale e calcolare l'affidabilità. La partizione "test" solitamente contiene anche dei casi limite per osservare il comportamento del modello di fronte a queste evenienze.

4.1 Dataset SLURP

La creazione del dataset MASSIVE è strettamente legata al dataset SLURP.
Pubblicato nel 2020, SLURP (Spoken Language Understanding Resource Package) [16] è un dataset per SLU (Spoken Language Understanding), una collezione di 72 mila registrazioni audio riguardanti interazioni tra un essere umano e un assistente vocale.
SLURP [16] è stato creato chiedendo a lavoratori AMT (Amazon Mechanical Turk) [9] di formulare comandi per un assistente virtuale, usando 200 prompt predefiniti come "Come chiederesti l'ora?", "Come imposteresti l'allarme?", "Come chiederesti di riprodurre la tua canzone preferita?".
Le varie registrazioni audio sono state registrate in varie condizioni, ad esempio con i microfoni in varie configurazioni, con un diverso tono della voce e con vari livelli di rumore di sottofondo. Questo per avere una varietà nella qualità delle registrazioni, in modo da poter allenare un agente SLU capace di riconoscere il linguaggio parlato in una varietà di situazioni.
Dopo aver filtrato le registrazioni non consone hanno ricavato 58 ore di materiale.
SLURP possiede livelli di complessità lessicale più alti rispetto a dataset simili, oltre ad essere un dataset notevolmente più grande rispetto ai precendenti.

4.2 Creazione di MASSIVE

MASSIVE è stato creato traducendo i vari enunciati presenti in SLURP (solo in lingua inglese) nelle diverse lingue supportate.
I linguaggi di MASSIVE sono stati scelti inizialmente considerando il costo per la produzione del database e la disponibilità di lavoratori per il lavoro di traduzione. Dopo questa prima valutazione è stata considerata la disponibilità di tali linguaggi nei principali assistenti virtuali.
La terza considerazione che è stata fatta era riguardante la diversità tipologica delle varie lingue, mentre la quarta circa la centralità degli autovettori [17] degli articoli Wikipedia, dei tweets e dei libri, usati come indicatori dell'influenza di una lingua in

Internet. [18]
L'ultima considerazione era riguardante gli alfabeti di ogni lingua, si è cercato di massimizzare la diversità di questi.
Per quando riguarda le varie lingue, la maggior parte (28) usano l'alfabeto latino, 3 lingue usano l'alfabeto arabo, 2 usano l'alfabeto cirillico e le rimanenti 18 lingue utilizzano alfabeti unici.

MASSIVE è composto da enunciati diretti a un device, e non ad una persona. Il database consiste principalmente in interrogativi e imperativi. Ci sono pochi enunciati dichiarativi. Questo è in contrasto con altri database, i quali ne contengono una frazione maggiore, solitamente perché gli enunciati sono stati estrapolati da contesti dove gli umani comunicano con altri umani.
Nell'ambito degli assistenti vocali, un utilizzatore solitamente chiede al device di eseguire un'azione o di rispondere a una domanda.
Per questo motivo gli enunciati dichiarativi sono meno frequenti e anche quando sono presenti enunciati dichiarativi come "Fa freddo" l'assistente virtuale potrebbe rispondere a tale frase aumentando la temperatura, trattando quindi un enunciato dichiarativo come uno imperativo.
Il database MASSIVE dà l'opportunità di studiare queste tipologie di enunciati nelle varie lingue supportate grazie a dei database paralleli nelle varie lingue.

4.2.1 Struttura delle frasi

Le varie lingue tendono ad avere regole diverse rispetto la costruzione delle frasi. Indicando il soggetto con S, il verbo con V e l'oggetto con O sono possibili 6 ordini. Nelle varie lingue del database sono state osservati tutti i possibili ordinamenti, anche se la grande maggioranza delle 51 lingue hanno il soggetto come primo elemento. 24 lingue sono di tipo SVO, 15 sono SOV e solo 3 lingue hanno il verbo come primo elemento (3 VSO). Non sono presenti lingue con l'oggetto come primo elemento della frase, 5 lingue non hanno un ordine preferito e 4 non hanno dati relativi all'ordine delle parole.

4.2.2 Livelli di cortesia

Varie lingue hanno differenti livelli di cortesia attraverso l'utilizzo dei pronomi. Nella lingua italiana, ad esempio, il pronome "tu" (seconda persona singolare) è utilizzato nei dialoghi informali, mentre il pronome della terza persona singolare "lei" viene usato quando ci si rivolge indirettamente a una persona di genere femminile, oppure nei dialoghi più formali.
Questi sistemi di cortesia sono pesantemente influenzati dal contesto sociale e il dataset MASSIVE permette di analizzare come le persone adattino il linguaggio quando si riferiscono ad un assistente virtuale.
21 lingue in MASSIVE attuano una distinzione binaria tra linguaggio formale e informa-

le usando due pronomi distinti. Questo comportamento è influenzato dalla provenienza europea di molte lingue presenti nel dataset.
Altre 8 lingue hanno più di due livelli di formalità, come informale, formale e onorifico. 7 lingue invece tendono a omettere completamente i pronomi in una situazione formale. Infine 11 lingue non hanno dati relativi a situazioni formali o di cortesia.

4.3 Raccolta dati per il dataset

Il dataset MASSIVE è stato creato con un flusso di lavoro gestito grazie ad Amazon Mechanical Turk [9] [19]. É stato creato da un insieme di fornitori con le capacità e risorse per creare un grande dataset in varie lingue.
I fornitori sono stati selezionati in base al costo e alla disponibilità di risorse.
I creatori del dataset hanno utilizzato due meccanismi per valutare i lavoratori.
Il primo meccanismo è un test sulla conoscenza della lingua nel quale i lavoratori ascoltano delle domande e degli enunciati nella lingua target, successivamente rispondono a delle domande a crocette.
Il secondo meccanismo consiste in un test di valutazione per osservare se i lavoratori siano riusciti a notare degli errori in alcuni enunciati tradotti.
Per migliorare ulteriormente la qualità della selezione dei lavoratori è stato creato un terzo quiz di traduzione dove ai lavoratori veniva richiesto di aver compreso le istruzioni del progetto rispondendo ad alcune domande.
Prima di cominciare i lavori è stato fatto un primo piccolo test, creando un dataset di dimensioni ridotte in tre lingue, per migliorare la chiarezza delle istruzioni e risolvere eventuali problematiche.

4.3.1 Task per la raccolta

Il primo task consiste nella traduzione o localizzazione degli slot. I lavoratori trovano l'enunciato con alcune parole colorate che corrispondono agli slot, in seguito viene richiesto loro di tradurre o localizzare lo slot.
Ad esempio, nella frase "I would like to listen to pop music please" la parola pop (genere musicale) non ha bisogno di una traduzione in italiano, la parola "music" invece deve essere tradotta in "musica". Localizzare uno slot significa cambiare un termine o un intero slot con qualche altro termine o insieme di parole più idonee alla cultura o alla lingua in cui si sta traducendo. Per esempio "La La Land" in francese verrebbe localizzato in "Pour l'amour d'Hollywood". Anche i nomi di autori, cantanti e attori possono essere localizzati in altri nomi più conosciuti nella lingua di destinazione. É presente anche l'opzione di tenere lo slot inalterato per parole uguali nelle varie lingue (es. "pop") oppure per nomi di artisti, personaggi o opere di fama internazionale (es. "Taylor Swift").

Nel file JSONL è presente l'opzione usata per ogni slot (traduzione, localizzazione, invariato).

"annot_utt": "olly spegni le luci della [house_place : camera da letto]",
"worker_id": "31", "slot_method": [{"slot": "house_place", "method": "translation"}]

In questo caso, "camera da letto" è un termine tradotto dall'inglese (verosimilmente "bedroom").
Dopo questo primo lavoro di localizzazione degli slot, ad un secondo lavoratore viene chiesto di tradurre o localizzare l'intero enunciato usando il lavoro fatto dal primo lavoratore. Egli può decidere se mantenere lo slot com'era stato tradotto, modificarlo o eliminarlo se non rilevante. É inoltre responsabile della gestione dei generi grammaticali e delle preposizioni. L'obiettivo è quello di tradurre l'enunciato nella lingua target in maniera fluida e naturale.
Questa divisione del lavoro di traduzione tra due persone è stata scelta in quanto aiuta a rendere il lavoro più veloce e a ridurre la mole di lavoro per lavoratore.

4.3.2 Controllo qualità traduzioni

La qualità dell'enunciato completamente tradotto viene giudicata da tre lavoratori, i quali controllano e giudicano: la corrispondenza dell'intento semantico tra l'enunciato originale e quello tradotto, la corrispondenza semantica degli slot, la grammatica della frase tradotta e la naturalezza, l'ortografia e l'identificazione della lingua.
Questi giudizi sono inclusi nel file JSONL.

"judgments": [{"worker_id": "40", "intent_score": 1, "slots_score": 0, "grammar_score": 4,
"spelling_score": 2, "language_identification": "target"}, ...]

I lavoratori sono stati controllati per verificare se le traduzioni fossero principalmente frutto di traduttori automatici. I compiti assegnati ai lavoratori che hanno usufruito di traduttori sono stati riassegnati ad altri.
Gli autori hanno inoltre analizzato attentamente gli enunciati tradotti nelle lingue di cui avevano padronanza per controllare la qualità delle traduzioni e per identificare eventuali problematiche.

4.4 Caratteristiche dataset MASSIVE

Essendo MASSIVE un dataset utilizzabile liberamente è possibile osservare come sia stato strutturato nella pagina GitHub del progetto. [8]
Il dataset è organizzato in file JSON. Ogni file contiene tutti gli enunciati di tutte le 3 partizioni in un linguaggio supportato da MASSIVE.

Ogni riga del file json contiene varie informazioni come:

- `id` : identificatore originale nel database SLURP

- `locale` : codice del linguaggio e della nazione
- `partition` : partizione dell'enunciato: può essere "train", "dev" o "test"
- `scenario` : il dominio generale dell'enunciato, definito "scenario" in SLURP. Alcuni esempi sono "iot", "music", "play", "weather".
- `intent` : l'intento specifico di un tale enunciato. Alcuni esempi sono "weather_query", "iot_coffee", "audio_volume_mute".
- `utt` : enunciato originale senza annotazioni testuali
- `annot_utt` : enunciato con annotazioni testuali
- `worker_id` : identificatore dell'utente che ha completato il lavoro di localizzazione dell'enunciato
- `slot_method` : per ogni slot nell'enunciato indica se sia stata una traduzione (stessa espressione tradotta nella lingua target), una localizzazione (non la stessa espressione bensì una più adatta al contesto e alla lingua) o se sia rimasto invariato.
- `judgments` : ogni giudizio per un enunciato consiste in 6 chiavi
  - `worker_id` : indica l'ID dell'utente che ha giudicato l'enunciato
  - `intent_score` : punteggio all'intento semantico dell'enunciato
  - `slots_score` : punteggio alla corrispondenza tra slot ed etichette (labels)
  - `grammar_score` : punteggio alla grammatica e alla naturalezza dell'enunciato
  - `spelling_score` : punteggio all'ortografia
  - `language_identification` : identificazione della lingua

4.5 Uilizzo del dataset MASSIVE

Ho scaricato il database MASSIVE [8] e ho utilizzato solamente il file JSON relativo agli enunciati in lingua italiana.
Con una ricerca regex ho creato altri tre file JSON (train.json, dev.json, test.json) i quali contengono solamente gli enunciati relativi a una partizione.
Per esempio, per controllare che una riga del file JSON fosse della partizione train ho fatto una ricerca regex col comando `"partition": "train"` in modo da selezionare tutte le righe contenenti tale sottostringa e copiarle nel file train.json.
Successivamente ho creato altri file (train.txt, dev.txt, test.txt, utt_only.txt) i quali contengono solamente gli enunciati (uno per riga) divisi per partizione, oppure tutti

gli enunciati presenti nel database nel caso del file utt_only.txt.
A partire da questi 4 file txt creati ho cominciato le mie successive analisi relative al database.

Capitolo 5

Prime analisi del dataset MASSIVE

Per effettuare le varie analisi ho utilizzato Python nella versione 3.11.4.
Ho condotto analisi sull'intero dataset, analizzando anche le singole partizioni train, dev e test.
Per ogni analisi ho calcolato la frequenza delle parole (il conteggio delle occorrenze totali dei termini di un tale dizionario nel database MASSIVE o in una sua partizione) e ho effettuato il conteggio delle parole diverse trovate.

5.1 Analisi Gaucher

Nella prima analisi ho calcolato l'utilizzo delle parole nella lista maschile e femminile Gaucher [12]. I risultati relativi alla frequenza dei termini e al numero di parole diverse sono i seguenti :

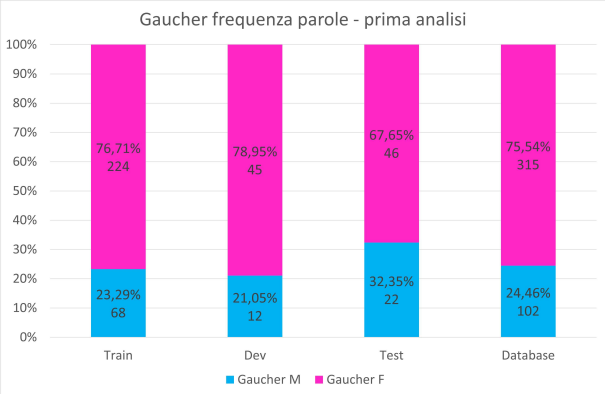


Figura 5.1: Grafico analisi Gaucher

Lista	Train	Dev	Test	Tot
M	68	12	22	102
F	224	45	46	315
M diz.	19	7	10	24
F diz.	47	16	21	53

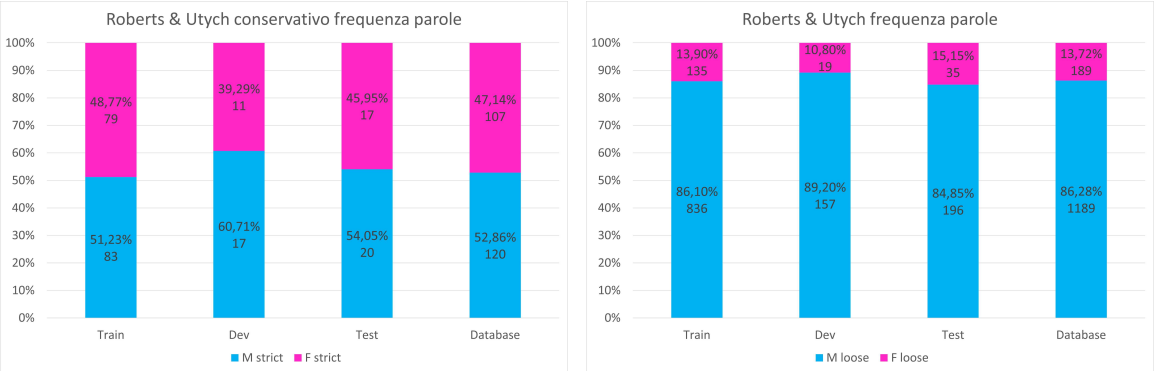
Diz : termini diversi apparsi nell'analisi

Tabella 5.1: Tabella analisi Gaucher

Si può notare come ci sia una netta prevalenza di termini appartenenti alla lista femminile (75.54 % delle occorrenze totali), anche la quantità di vocaboli diversi rilevati risulta maggiore per i termini femminili (68.83 % dei termini diversi totali).
I termini appartenenti alla lista femminile con più occorrenze sono: "consegna" (51 volte), "caldo" (42), "rendi" (28), "impegni" (23), "insieme" (16). I termini maschili più frequenti sono: "attiva" (42), "attive" (14), "opinione" (6), "superiore" (5).

5.2 Analisi Roberts and Utych

Nella seconda analisi ho utilizzato le liste Roberts and Utych maschile, femminile, maschile conservativo e femminile conservativo.
I risultati sono riportati di seguito :



(a) Grafico analisi Roberts & Utych conservativo

(b) Grafico analisi Roberts & Utych allargato

Figura 5.2: Grafici analisi Roberts & Utych

Tabella riassuntiva con tutti i risultati:

Lista	Train	Dev	Test	Tot
M cons	83	17	20	120
F cons	79	11	17	107
M	836	157	196	1189
F	135	19	35	189
M cons diz	14	7	8	15
F cons diz	15	6	8	16
M diz	70	34	36	84
F diz	33	11	16	40

Diz : termini diversi apparsi nell'analisi

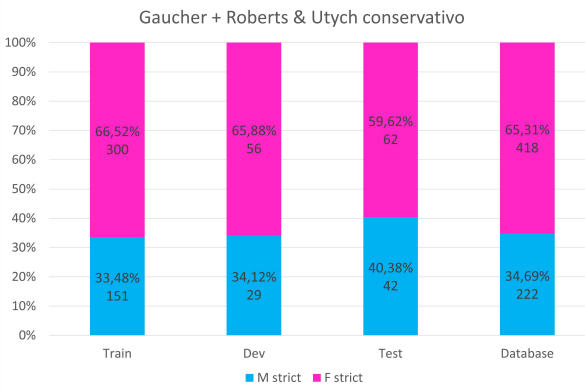
Tabella 5.2: Tabella analisi Roberts & Utych

Come si può notare è presente un leggero pregiudizio maschile nella prima analisi del dizionario Roberts & Utych, ovvero quella delle liste maschile e femminile conservativo. Nell'analisi delle liste maschile e femminile è presente un evidente bias maschile. I termini presenti nella lista femminile ma non in quella femminile conservativa appaiono solamente 82 volte nell'intero database. I termini maschili ma non maschili stretti invece appaiono ben 1069 volte.
Un termine presente molto spesso nella lista maschile è "puoi" e altre coniugazioni del verbo "potere".
La situazione è simile per il numero di termini diversi che appaiono: se per le liste

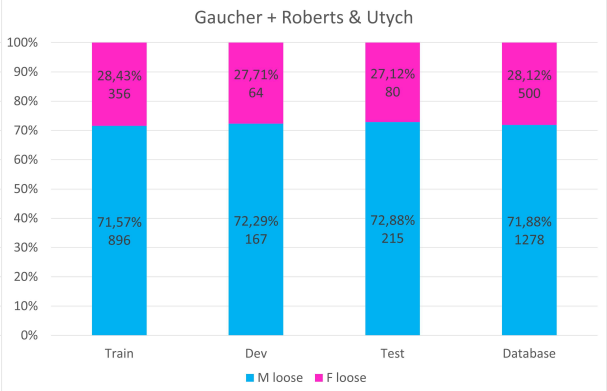
maschile e femminile strict c'è una sostanziale parità, nelle liste maschile e femminile si ha un numero di termini maschili doppio rispetto alla varietà di termini femminili. I termini maschili non conservativi più ricorrenti sono: "puoi" (527 occorrenze), "posso" (172), "potresti" (50). I termini femminili non conservativi più frequenti sono: "bella" (45 volte), "sentire" (39 volte), "amore" (12).

5.3 Analisi Gaucher + Roberts & Utych

Qui sono riportati i risultati dell'unione delle liste Gaucher e Roberts & Utych.



(a) Grafico analisi Gaucher + Roberts & Utych conservativo



(b) Grafico analisi Gaucher + Roberts & Utych allargato

Figura 5.3: Grafici analisi Gaucher + Roberts & Utych

Lista	Train	Dev	Test	Tot
M cons	151	29	42	222
F cons	300	56	62	418
M	896	167	215	1278
F	356	64	80	500
M cons diz	33	14	18	39
F cons diz	60	22	28	67
M diz	85	39	44	103
F diz	78	27	36	91

Diz : termini diversi apparsi nell'analisi

Tabella 5.3: Tabella analisi Gaucher + Roberts & Utych

Si può notare come nelle analisi della lista Gaucher e la lista maschile o femminile stretta di Roberts sia presente un bias femminile (65.31% termini femminili nei linguaggi di genere dell'intero database), mentre nell'analisi della lista Gaucher + Roberts allargata (M/F loose) si nota un pregiudizio maschile con un 72% di termini di genere maschile.

Per la diversità di termini nell'analisi della lista Gaucher + Roberts strict si ha una maggiore diversità femminile (63.2% dei diversi termini apparsi sono femminili), mentre usando la lista Gaucher + Roberts non conservativa si nota una quasi uguaglianza rispetto la diversità di vocaboli (53.1 % di termini diversi maschili rispetto al totale).

Capitolo 6

Altre analisi

Oltre alle analisi della frequenza dei termini presenti nei dizionari Gaucher e Roberts & Utych ho anche eseguito diverse altre analisi qui riportate.

6.1 Analisi sostantivi

Sfruttando le caratteristiche del dizionario italiano Morph-It! [14] [15] ho potuto creare una lista di sostantivi presenti nella lingua italiana.
A tal fine ho usato una ricerca regex cercando tutte le righe che avessero
`HT + NOUN + altri caratteri`
dove HT è il carattere horizontal tab.
Utilizzando le due liste di sostantivi maschili e femminili ho potuto analizzare la loro frequenza nel dataset MASSIVE e nelle relative partizioni:

Lista	Train	Dev	Test	Tot
M	13324	2308	3193	18825
F	10168	1749	2663	14580
M diz	1362	588	672	1576
F diz	989	400	516	1139

Diz : termini diversi apparsi nell'analisi

Tabella 6.1: Tabella analisi sostantivi

È presente un numero maggiore di sostantivi maschili (56.35 % dei sostantivi nel database sono maschili), così come il numero di sostantivi diversi apparsi è maggiore per quelli di genere maschile.
I sostantivi maschili più ricorrenti sono: "favore" (939 occorrenze), "oggi" (876), "calendario" (517), "domani" (378), "evento" (344). I sostantivi femminili più frequenti sono invece: "email" (983 occorrenze), "lista" (685), "cosa" (495), "settimana" (419), "luci" (389), "canzone" (369).

6.2 Analisi nomi di persona

Successivamente ho analizzato la frequenza di nomi di persona maschili e femminili.
Per creare una lista il più possibile completa ho unito quattro diverse liste di nomi maschili [20] [21] [22] [23] e altre quattro liste di nomi femminili [24] [25] [26] [27].
Dopo aver effettuato l'unione di queste liste, la lista maschile conteneva circa 3700 nomi, mentre quella femminile circa 7200 nomi. In seguito, ho dovuto analizzare i nomi presenti nelle due liste sopracitate e rimuovere eventuali stop words in quanto i primi risultati mostravano molti termini utilizzati non come nomi di persona (esempi includono "Che", "Mai" , "Ma" e altri). Per eliminare le stop words dall'elenco di nomi di persona ho utilizzato il sito [28]. Ho inoltre effettuato delle analisi contestuali osservando se i nomi di persona presenti nel dataset corrispondevano a stop words e in caso affermativo ho rimosso tali termini.
In un secondo momento ho rimosso dalle liste i nomi "Alexa", "Siri" e "Olly" in quanto questi nomi femminili sono utilizzati per riferirsi all'assistente virtuale e non ad una persona.

Lista	Train	Dev	Test	Tot
M	743	128	203	1074
F	826	145	195	1166
F no agenti	388	68	99	555
M diz	160	59	89	191
F diz	122	45	54	132
F no agenti diz	120	43	52	130

Diz : termini diversi apparsi nell'analisi
No agenti : analisi senza i nomi "Alexa", "Siri", "Olly"

Tabella 6.2: Tabella analisi nomi di persona

Si nota come più della metà delle occorrenze di nomi femminili nel database fossero in realtà riferimenti all'assistente virtuale (per esempio "olly spegni le luci della camera da letto").
Togliendo quindi i nomi degli assistenti vocali si nota come ci sia un numero ben maggiore di riferimenti a persone di genere maschile. Anche la varietà di nomi di persona è maggiore per i nomi maschili (59.5 % di termini diversi maschili sul totale).
I nomi maschili più ricorrenti sono: "Giovanni" (116 volte), "Marco" (58 volte), "Mario" (48), "Michele" (36), "Giacomo" (34). I nomi femminili più frequenti sono: "Sara" (48 volte), "Anna" (31 volte), "Laura" (30), "Maria" (28), "Giulia" (22).
Questa presenza di nomi principalmente italiani è da attribuirsi al lavoro di localizzazione eseguito durante la creazione del dataset MASSIVE. Durante questo processo i nomi originali, presenti nel dataset inglese, sono stati sostituiti con dei nomi più comuni nella lingua italiana.

Non tutti i nomi presenti sono tipicamente italiani, in quanto alcune persone o personaggi di fama internazionale non sono stati sostituiti nella traduzione italiana. Alcuni esempi sono "Michael Jackson" ed "Harry Potter".

6.3 Analisi professioni

Per verificare l'eventuale presenza di pregiudizi di genere e l'utilizzo di un maggior numero di riferimenti maschili ho pensato di creare delle liste con vari nomi di occupazioni maschili e femminili per poi analizzare la frequenza di tali termini nel database MASSIVE.
Per formare un elenco di professioni ho utilizzato la lista disponibile nel sito WeCanJob [29], partendo da questa lista iniziale ho creato un elenco di professioni al maschile e al femminile.
Analizzando le occorrenze nell'intero database si trovano 87 occorrenze per la lista maschile, 75 per la lista femminile. Per quanto riguarda la diversità di vocaboli si contano 35 diverse professioni maschili e 22 femminili.
Dopo aver fatto questa prima analisi ho notato come le due liste avessero molte professioni in comune, in quanto varie professioni non hanno una versione femminile diversa da quella maschile.
Per ovviare a questo problema ho creato una lista di professioni "neutre", ovvero di professioni comuni tra le liste maschile e femminile. Successivamente ho creato due liste di nomi di occupazioni solamente femminili e solamente maschili.

Lista	Train	Dev	Test	Tot
M	13	6	3	22
F	6	0	4	10
M diz	12	6	3	18
F diz	5	0	3	5
M contesto	6	6	2	14
F contesto	1	0	0	1
M contesto diz	6	6	2	12
F contesto diz	1	0	0	1

Diz : termini diversi apparsi nell'analisi
Contesto : analisi fatta togliendo i termini non usati come professioni

Tabella 6.3: Tabella analisi professioni

Rifacendo le analisi si può notare come sia presente una certa prevalenza di occupazioni maschili (22 occorrenze maschili contro 10 femminili). Dopo il completamento della seconda analisi ho fatto un'analisi contestuale e ho rimosso i termini che non venivano utilizzati per riferirsi a professioni (come "medici", utilizzata come aggettivo nella frase "cancella gli impegni medici questo fine settimana").

Dopo l'analisi contestuale risulta evidente come siano presenti molti più termini relativi a professioni maschili che femminili, l'unico termine femminile presente è "segretaria" (1 occorrenza). Le professioni maschili più presenti sono: "direttore" (3 occorrenze), "analisti", "veterinario", "professori", "medico", "venditore" e altri (sempre 1 occorrenza).

6.4 Analisi pronomi

Un'altra analisi interessante è quella relativa ai pronomi.
Ho creato una lista di tutti i pronomi utilizzando Wikipedia [30] e siti di approfondimento della lingua italiana [31].
Analizzando velocemente la presenza di questi pronomi nel database si trovano 6054 occorrenze di 29 pronomi femminili e 2280 occorrenze di 32 pronomi maschili. Una semplice analisi contestuale mostra come molte occorrenze siano relative ad articoli (come "la" o "le") che coincidono con pronomi.
Per ovviare a questo problema nella successiva analisi mi sono limitato a considerare i pronomi personali e possessivi. Il risultato mostra un utilizzo leggermente maggiore di pronomi personali e possessivi femminili (53.66 % del totale).
I pronomi possessivi e personali maschili più usati sono: "mio" (716 occorrenze), "miei" (205), "suo" (18). Mentre per quanto riguarda i pronomi possessivi o personali femminili i più presenti sono: "mia" (916 occorrenze), "mie" (177), "tua" (17).

Lista	Train	Dev	Test	Tot
M	673	123	193	989
F	792	158	195	1145
M diz	10	6	6	10
F diz	10	6	4	10

Diz : termini diversi apparsi nell'analisi

Tabella 6.4: Tabella analisi pronomi personali e possessivi

6.5 Analisi verbi e particelle

6.5.1 Analisi verbi

Sfruttando le caratteristiche del dizionario Morph-It! [14] ho analizzato i verbi per cercare se fossero presenti delle disparità nell'utilizzo di verbi coniugati al femminile o al maschile.
Per creare le due liste di verbi ho eseguito una ricerca regex utilizzando il dizionario Morph-It!. Ho ricercato tutte le righe contenenti la sottostringa "VER" per creare una lista contenente tutti i verbi e le varie coniugazioni.

Successivamente ho eseguito varie ricerche per cercare tutte le coniugazioni di verbi maschili o femminili.
Le ricerche regex per trovare i verbi coniugati al femminile sono le seguenti: `[+]f[+]`, `[+]f` in quanto m,f indicano rispettivamente una coniugazione al maschile o al femminile e i vari tag sono separati da "+". Una ricerca analoga è stata fatta per le coniugazioni maschili.
Similmente all'analisi delle professioni, ho creato una lista di coniugazioni comuni tra la lista maschile e femminile, così da creare una lista di verbi coniugati solamente maschili e femminili.
Dopo questo lavoro di preparazione ho potuto iniziare le analisi le quali rilevano una leggera maggioranza di termini maschili (1745) rispetto alle voci verbali femminili (1653).
Un'analisi contestuale rivela come sia necessario rifare le analisi escludendo alcuni termini in quanto non utilizzati come verbi. Alcuni esempi possono essere: "stato", "punto", "spesa", "posta". Questi termini sono stati utilizzati come sostantivi.
I risultati della seconda analisi sono riportati nella figura 6.5.

Lista	Train	Dev	Test	Tot
M contesto	1021	177	283	1481
F contesto	744	134	195	1073
M contesto diz	282	98	137	338
F contesto diz	170	58	82	207

Diz : termini diversi apparsi nell'analisi
Contesto : analisi fatta togliendo i termini non usati come voci verbali

Tabella 6.5: Tabella analisi voci verbali

Si può notare una maggioranza di verbi maschili (57.99 % delle occorrenze totali) e una maggiore varietà di voci maschili (62.02 % di termini diversi maschili rispetto al totale).

6.5.2 Analisi particelle

Usando sempre le ricerce regex e il dizionario Morph-It! ho potuto creare una lista di coniugazioni con particelle (come "creatosi", "scatenatasi" e simili) per eseguire una analisi simile alla precedente, ma non ho trovato occorrenze di tali termini nel database.

Capitolo 7

Sviluppo dizionario AVA

I risultati delle prime analisi del database (riportate nel capitolo 9) non erano convincenti e talvolta contraddittori.
Per tale motivo ho cercato di analizzare nel dettaglio i termini trovati dai dizionari Gaucher e Roberts & Utych per osservare gli enunciati e i contesti nei quali questi termini venivano utilizzati.
Per eliminare dalle analisi i termini è stato necessario creare un dizionario di termini ambigui (AVA, Ambiguity for Virtual Assistants).
Un termine viene definito ambiguo e inserito nel dizionario AVA se è presente nei dizionari di linguaggio di genere usati (Gaucher e Roberts & Utych), quindi se viene considerato come un'istanza di linguaggio implicito di genere, ma è anche presente in contesti legati agli assistenti virtuali o a contesti di linguaggio non di genere. Ovvero se tale termine può essere interpretato in molteplici modi in relazione al contesto dell'enunciato nel quale viene utilizzato.
Riassumendo, i criteri usati per determinare l'ambiguità di una parola:

- Il termine era presente in un dizionario di linguaggio implicito di genere
- Il termine viene trovato in almeno un'istanza nel dataset MASSIVE
- Il termine è ambiguo in uno dei seguenti modi nel dataset
  - Il termine non viene usato con le caratteristiche tipiche del linguaggio implicito di genere, ma possiede un significato speciale relativo al contesto degli assistenti virtuali. Ad esempio, il verbo "potere" e le sue coniugazioni vengono usate per dare comandi all'assistente virtuale.
  - Se il termine è presente più volte nel database: almeno in un'istanza il termine viene utilizzato con il significato indicato per considerarla un'istanza di linguaggio di genere e in almeno un'istanza il termine viene utilizzato con riferimenti al contesto degli assistenti virtuali e non è presente un'interpretazione legata al linguaggio di genere.
    Ad esempio, l'aggettivo "bella" viene utilizzato per riferirsi alla bellezza

femminile come nella frase “qual è l’attrice più bella del mondo in questo secolo” ma viene soprattutto utilizzato per chiedere all’assistente virtuale una “bella barzelletta” o una “bella canzone”.

Gli autori della ricerca [1] hanno sviluppato un dizionario AVA contenente 44 termini definiti ambigui nella lingua inglese.
Inizialmente ho analizzato il loro dizionario AVA, tradotto i termini in italiano e ho analizzato i contesti degli enunciati nei quali apparivano tali termini per capire se fosse il caso di inserire questi vocaboli nel dizionario AVA italiano.
Alcuni termini presenti nel dizionario AVA originale sono stati tradotti e inseriti nel dizionario AVA italiano, come le parole “figlio” e “bambino” e i relativi termini derivati come “bambini”, “bambina”, “bambine”. La parola nel dizionario AVA originale era “child” e veniva usata in contesti come “as a child”, “an only child”.
Essendo questi termini usati principalmente per interrogare l’assistente virtuale riguardo ai figli di persone famose o in contesti simili a quelli del database inglese ho deciso di aggiungere tali termini al dizionario.
Altri termini presenti nel dizionario AVA inglese, come “independent”, non sono stati aggiunti al dizionario italiano in quanto non appaiono in contesti ambigui o simili ai contesti segnalati dagli autori della ricerca originale. [1]
Nella creazione del dizionario AVA in italiano ho utilizzato due liste distinte, la prima consiste in parole non derivate o coniugate (come dei nomi primitivi o forme all’infinito di un verbo), come il sostantivo “autore” o il verbo “amare”. Per ogni parola presente in questa lista ho cercato, attraverso Morph-It!, tutte le parole derivate (come “autori” oppure tutte le coniugazioni del verbo “amare”).
La seconda lista utilizzata per creare il dizionario AVA consiste in termini derivati, in questo caso i termini vengono utilizzati senza l’ausilio di Morph-It!. Ad esempio, ho inserito in questa lista tutte le coniugazioni del verbo “sentire” usate in enunciati non col significato di linguaggio di genere, mentre non ho inserito le coniugazioni “senti” e “sentirmi” in quanto vengono usate riferendosi alle sensazioni che può provare una persona in un determinato momento. Le altre coniugazioni del verbo “sentire” sono state usate col significato di “ascoltare” o con altri significati non collegati al linguaggio di genere.
In conclusione, la lista AVA contiene più di 1000 termini, di cui 51 termini già derivati e più di 40 termini dei quali si sono cercate le parole derivate.
Questa lista è stata utilizzata successivamente per rifare le analisi usando le liste Gaucher e Roberts & Utych togliendo i termini in grado di creare ambiguità.

Capitolo 8

Analisi con AVA

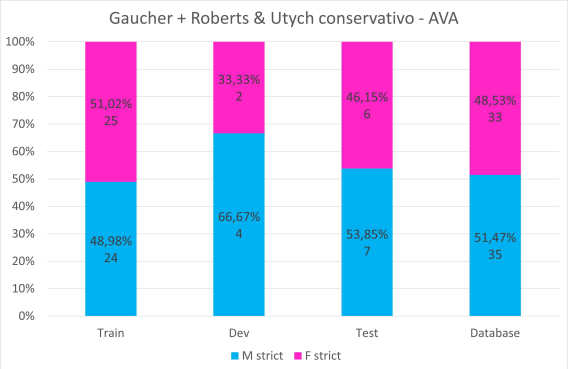
8.1 Analisi Gaucher + Roberts con AVA

Rifacendo le analisi relative alla frequenza ho ottenuto i seguenti risultati:

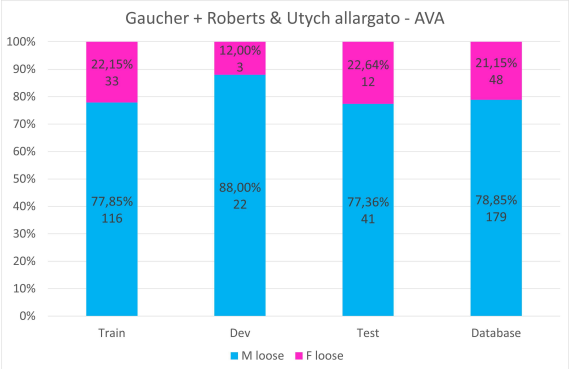
Lista	Train	Dev	Test	Tot
M cons	24	4	7	35
F cons	25	2	6	33
M	116	22	41	179
F	33	3	12	48
M cons diz	15	4	5	17
F cons diz	14	2	5	15
M diz	35	10	16	42
F diz	20	3	9	25

Diz : termini diversi apparsi nell'analisi

Tabella 8.1: Tabella analisi Gaucher + Roberts & Utych con AVA



(a) Grafico analisi Gaucher + Roberts & Utych conservativo con AVA



(b) Grafico analisi Gaucher + Roberts & Utych allargato con AVA

Figura 8.1: Grafici analisi Gaucher + Roberts & Utych con AVA

Si può notare come ci sia un leggero bias maschile nell'analisi legata al dizionario Gaucher unito al Roberts & Utych conservativo, per quanto sia da considerare come

un'analisi limitata in quanto avendo tolto molti termini ambigui ci si ritrova con un numero esiguo di occorrenze.
Per quanto riguarda i risultati usando la lista Gaucher unita alla Roberts & Utych allargata si nota una importante frequenza di termini maschili (78.85 % del totale). Anche la diversità di termini è certamente maggiore per i termini maschili (42 vocaboli diversi maschili contro 25 femminili).

8.2 Analisi Gaucher con AVA

Rifacendo le analisi fatte nel capitolo 5.1 eliminando i termini presenti nel dizionario AVA si trovano i seguenti risultati:

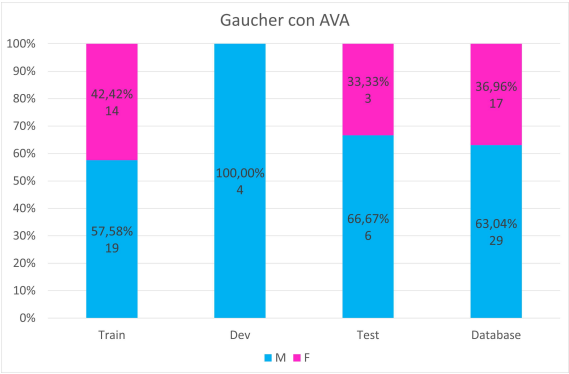


Figura 8.2: Grafico analisi Gaucher con AVA

Lista	Train	Dev	Test	Tot
M	19	4	6	29
F	14	0	3	17
M diz	12	4	4	14
F diz	11	0	3	12

Diz : termini diversi apparsi nell'analisi

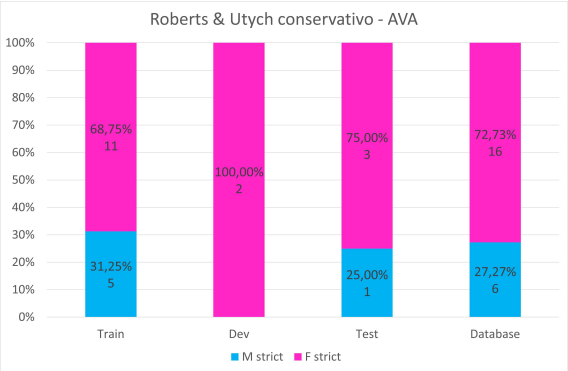
Tabella 8.2: Tabella analisi Gaucher con AVA

Si nota un maggior numero di riferimenti presenti nella lista Gaucher [12] maschile (63.04 % di termini maschili rispetto al totale). La lista maschile ha inoltre una leggera maggioranza nel numero di termini diversi apparsi (53.84 % di termini diversi maschili apparsi rispetto al totale).
I termini maschili più presenti sono: "opinione" (6 occorrenze), "sportivi" (4), "sportivo" (4), "logica" (3). I termini femminili con più occorrenze sono invece: "dolci" (3 occorrenze), "allegra" (2), "capire" (2).

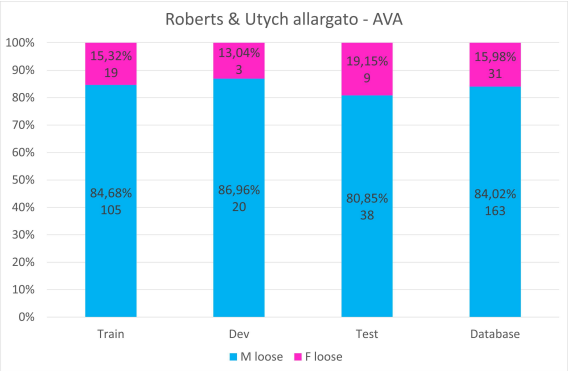
8.3 Analisi Roberts & Utych con AVA

Rifacendo le analisi relative al dizionario Roberts & Utych [13] dopo aver rimosso le parole presenti nel dizionario AVA ho ottenuto i seguenti risultati:

Si può notare come ci sia un bias femminile nell'analisi effettuata con il dizionario Roberts & Utych conservativo (72.73 % di termini femminili), per quanto sia da considerare come un'analisi limitata in quanto avendo rimosso una grande quantità di termini ambigui ci si ritrova con un numero esiguo di occorrenze. Appaiono solamente 3 termini diversi per la lista maschile conservativa, come anche per la lista femminile.



(a) Grafico analisi Roberts & Utych conservativo con AVA



(b) Grafico analisi Roberts & Utych allargato con AVA

Figura 8.3: Grafici analisi Roberts & Utych con AVA

Lista	Train	Dev	Test	Tot
M cons	5	0	1	6
F cons	11	2	3	16
M	105	20	38	163
F	19	3	9	31
M cons diz	3	0	1	3
F cons diz	3	2	2	3
M diz	27	8	14	33
F diz	9	3	6	13

Diz : termini diversi apparsi nell'analisi

Tabella 8.3: Tabella analisi Roberts & Utych con AVA

Le parole più frequenti per la lista maschile sono: "ragazzo" (3 volte), "uomini" (2), "uomo" (1). Per quanto riguarda la lista femminile sono: "donne" (7 occorrenze), "donna" (5), "signora" (4).

Per quanto riguarda i risultati ottenuti usando la lista Gaucher unita alla Roberts & Utych allargata si nota una netta maggioranza di termini maschili (84.02 % del totale). Anche la diversità di termini è certamente maggiore per i termini maschili (33 vocaboli diversi maschili contro 13 femminili).

I termini maschili più frequenti sono: "capo" (40 occorrenze), "amico" (28), "amici" (22), "cane" (19), "cani" (5). I termini femminili con più occorrenze invece sono: "donne" (7 occorrenze), "donna" (5), "piante" (4).

Capitolo 9

Discussione

9.1 Subdoli bias maschili

Nelle varie analisi ho trovato forme di bias di genere nel dataset italiano MASSIVE. Sono state rivelate anche forme di bias femminile, ma un numero maggiore di forme di pregiudizio maschile.
Un ruolo importante in queste analisi l'ha avuto il linguaggio ambiguo. Spesso il linguaggio legato all'ambito degli assistenti virtuali è stato rilevato nelle analisi relative al linguaggio di genere. Per questo motivo è stato fondamentale cambiare approccio e rianalizzare il dataset in maniera più critica e ammettendo che l'analisi finale sia limitata da questo fattore.
Il dizionario AVA è da considerare come un punto di partenza per analisi più rigorose in futuro.

9.2 Future analisi e ricerche

Fare valutazioni su grandi dataset NLP usando procedure automatiche è complesso.
Il linguaggio naturale è: strettamente legato al contesto, dinamico e ha varie sfaccettature. Termini colloquiali, nomi utilizzati anche come termini comuni e l'influenza del contesto sono alcune delle difficoltà in cui mi sono imbattuto durante la ricerca di bias subdoli di genere all'interno del dataset.
Una problematica interessante da affrontare in futuro sarà il distinguere i pronomi "essi" e "loro" quando sono usati per riferirsi a una pluralità di persone, oppure quando sono usati per riferirsi a una singola persona con un'identità non binaria.
Sarà necessario inoltre essere attenti a bias di genere femminile: bisogna evitare di creare dataset o algoritmi con pregiudizi femminili in lavori futuri di rimozione di pregiudizi di genere maschile.

9.3 Discordie tra dizionari

I dizionari creati per un particolare contesto sono molto dipendenti: non soltanto dal contesto, ma anche dal linguaggio, dalla cultura, dalle comunità locali, dalla religione, etnicità, orientamento sessuale e dal genere o sesso delle persone che hanno partecipato alla creazione del dizionario oltre che dagli autori stessi.
Un dizionario potrebbe non essere applicabile a tutti i dataset. Le differenze tra i risultati con i dizionari Gaucher[12] e Roberts & Utych [13] mostrano come il risultato finale possa variare molto usando dizionari diversi.
Per via di queste incongruenze, molti termini di un dizionario possono non avere occorrenze in un dataset, oppure possono essere utilizzati in contesti diversi dove portano significati che non erano stati presi in considerazione dal dizionario. In questo modo, questi termini possono non essere associati al linguaggio di genere nella modalità prevista originariamente dal dizionario.
Termini presenti nei dizionari di linguaggio di genere come “figlio” o “insegnante”, presenti nelle liste di termini femminili, non sono stati utilizzati come linguaggio di genere nel dataset MASSIVE.
Questi risultati indicano una necessità di creare dizionari più adatti a queste analisi. Il dizionario AVA è un punto di partenza.

9.4 AVA

L’ambiguità è onnipresente nel linguaggio. Va intesa non solamente come la mancanza di chiarezza ma anche come la presenza di più significati.
Durante questa ricerca ho creato AVA, un dizionario contenente più di mille parole ambigue, parole che originariamente erano considerate linguaggio di genere dai due dizionari consultati (Gaucher [12] e Roberts & Utych [13]), ma che sono risultate successivamente ambigue dopo un’analisi contestuale.
Rimuovere tali parole dalle successive analisi è stato cruciale per poter avere dei risultati più precisi ed affidabili.
Alcuni esempi di tali parole possono essere “potere” e le sue coniugazioni (“puoi ordinare del sushi per la cena di stasera”, “fai partire il prossimo episodio del podcast quarto potere”) oppure “caldo” (“quanto fa caldo oggi a Trapani?”, “farà caldo venerdì sera?”).
Per migliorare ulteriormente il dizionario AVA e renderlo uno strumento utile alla valutazione dei pregiudizi presenti nei dataset serviranno importati contributi da persone non binarie, transgender, genderqueer e altri gruppi sottorappresentati.

9.5 Limiti e lavori futuri

Molte delle mie analisi sono state inevitabilmente limitate dai metodi utilizzati e dagli strumenti e materiali a mia disposizione. I dizionari Gaucher e Roberts & Utych non sono stati creati per misurare il linguaggio di genere nei dataset incentrati sul NLP (Natural Language Processing), inoltre si limitano a considerare due generi.
Questi aspetti portano a pensare che sia possibile in futuro creare dei nuovi dizionari basati sul linguaggio di genere nell'ambito degli assistenti virtuali, un simile strumento potrebbe verosimilmente migliorare l'affidabilità della misura di eventuali bias di genere.
L'analisi automatizzata del contenuto del dataset MASSIVE ha le sue limitazioni.
Per cercare di migliorare i risultati ho quindi dovuto fare delle analisi del contesto manuali andando a cercare le occorrenze di certi termini e controllare in che contesti venissero usati. Questo è stato il primo passo che ha portato alla creazione del dizionario AVA.
Durante la creazione di tale dizionario ho avuto alcune difficoltà nel capire quali termini inserire o meno, in quanto non sempre era chiaro se un termine venisse utilizzato come linguaggio di genere. Per questo motivo ritengo che il dizionario sia migliorabile, magari con il contributo di una persona più competente in materia di bias di genere.
Tuttavia, è probabile che non siano stati notati alcuni bias impliciti.
In futuro servirà fare altre ricerche sui bias impliciti, magari analizzando questi pregiudizi impliciti nei word embeddings. Per eseguire queste analisi serviranno dataset ben più grandi: contenenti anche miliardi di enunciati.
Il dataset inoltre rappresenta solo una frazione delle diverse interazioni possibili con un assistente virtuale, questo aspetto ha certamente inficiato l'analisi. Il dizionario AVA è relativamente piccolo e limitato alla partizione italiana del dataset MASSIVE e ai dizionari usati, in futuro sarà perciò utile espanderlo.

Capitolo 10

Conclusioni

Viviamo in un mondo dove gli assistenti virtuali hanno un ruolo sempre più centrale nelle nostre vite, l'IoT e nello specifico gli assistenti virtuali con i quali possiamo comunicare in maniera naturale sono sempre più sviluppati e complessi.
In questo contesto sono sempre più numerose le aziende pronte a mettere a disposizione di questi assistenti virtuali dei dataset per migliorarne le abilità nell'uso del linguaggio naturale.
Di fronte a questo sviluppo così frenetico, non sono assenti delle problematiche.
Come dimostra questo studio, nel dataset italiano di MASSIVE sono presenti dei pregiudizi di genere. Il rilevamento e l'analisi dei pregiudizi impliciti di genere è un primo passo, ma sarà necessario in futuro migliorare le analisi e creare un dizionario AVA più completo di quello attuale. Successivamente sarà necessario discutere quali tipologie di linguaggio di genere devono essere rappresentate in maniera rigorosa discutendone con le varie comunità interessate.
Lavorando assieme sarà possibile creare degli agenti dotati di intelligenza artificiale caratterizzati dall'assenza di bias di genere.

Elenco delle figure

Elenco delle tabelle

Ringraziamenti

Desidero infine ringraziare tutte le persone che hanno condiviso con me questi tre anni, senza di voi questa esperienza non sarebbe stata la stessa!
In primo luogo, vorrei ringraziare il mio relatore Antonio Rodà e la mia correlatrice Silvana Badaloni per la loro disponibilità e per i preziosi consigli.
Ringrazio poi la mia famiglia: Dario, Stefania, Elisa, i miei zii e i miei nonni che mi hanno sempre supportato e che hanno sempre creduto in me. Grazie infinite per il vostro affetto!
Ringrazio calorosamente tutti gli amici che ho avuto il piacere di conoscere durante questo percorso e che hanno riempito di momenti gioiosi le mie giornate a Padova: Elisa, Samuele, Giulia B., Giulia G., Giulio, Tommaso, Damiano, Luca, Nicolò, Ledia, Leonardo, Cristian, Rachele, Marius, Riccardo, Ludovico.
Un ringraziamento particolare lo devo a Giulia B. per essere stata la mia correttrice di bozze di fiducia, oltre che per avermi supportato e sostenuto in questi ultimi mesi.
Infine, i miei ringraziamenti vanno ai miei amici di Vicenza per essere stati al mio fianco per tutto questo tempo: Altea, Anna, Leonardo, Giovanni, Alberto, Leonardo, Silvia, Rebecca, Rachele, Anna, Angelica.

Vi ringrazio di cuore e vi auguro il meglio!
Fabio

Bibliografia

[1] K. Seaborn, S. Chandra, and T. Fabre, "Transcending the "male code": Implicit masculine biases in nlp contexts," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3544548.3581017

[2] UNESCO, EQUALS Skills Coalition, "I'd blush if i could: closing gender divides in digital skills through education," p. 147, 2019. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000367416

[3] Y. Chen, C. Mahoney, I. Grasso, E. Wali, A. Matthews, T. Middleton, M. Njie, and J. Matthews, "Gender bias and under-representation in natural language processing across human languages," in *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, ser. AIES '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 24–34. [Online]. Available: https://doi.org/10.1145/3461702.3462530

[4] T. Sun, A. Gaut, S. Tang, Y. Huang, M. ElSherief, J. Zhao, D. Mirza, E. Belding, K.-W. Chang, and W. Y. Wang, "Mitigating gender bias in natural language processing: Literature review," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 1630–1640. [Online]. Available: https://aclanthology.org/P19-1159

[5] T. Farrell, M. Fernandez, J. Novotny, and H. Alani, "Exploring misogyny across the manosphere in reddit," in *Proceedings of the 10th ACM Conference on Web Science*, ser. WebSci '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 87–96. [Online]. Available: https://doi.org/10.1145/3292522.3326045

[6] M. Horta Ribeiro, J. Blackburn, B. Bradlyn, E. De Cristofaro, G. Stringhini, S. Long, S. Greenberg, and S. Zannettou, "The evolution of the manosphere across the web," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 15, no. 1, pp. 196–207, May 2021. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/18053

[7] T. Bolukbasi, K.-W. Chang, J. Zou, V. Saligrama, and A. Kalai, "Man is to computer programmer as woman is to homemaker? debiasing word embeddings," 2016.

[8] J. FitzGerald, C. Hench, C. Peris, S. Mackie, K. Rottmann, A. Sanchez, A. Nash, L. Urbach, V. Kakarala, R. Singh, S. Ranganath, L. Crist, M. Britan, W. Leeuwis, G. Tur, and P. Natarajan, "Massive dataset," 2022. [Online]. Available: https://github.com/alexa/massive

[9] Amazon, "Amazon mechanical turk." [Online]. Available: https://www.mturk.com/

[10] J. FitzGerald, C. Hench, and C. P. et al, "Massive: A 1m-example multilingual natural language understanding dataset with 51 typologically-diverse languages," 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2204.08582

[11] B. Ghai, M. N. Hoque, and K. Mueller, "Wordbias: An interactive visual tool for discovering intersectional biases encoded in word embeddings," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3411763.3451587

[12] K. A. Gaucher D, Friesen J, "Evidence that gendered wording in job advertisements exists and sustains gender inequality," 2011. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/21381851/

[13] D. C. Roberts and S. M. Utych, "Linking gender, language, and partisanship: Developing a database of masculine and feminine words," *Political Research Quarterly*, vol. 73, no. 1, pp. 40–50, 2020. [Online]. Available: https://doi.org/10.1177/1065912919874883

[14] RosaeNLG, "Morph-it!, italian words dict," GitHub. [Online]. Available: https://github.com/RosaeNLG/rosaenlg/tree/master/packages/italian-words-dict

[15] UniBo, "Morph-it!" UniBo.it. [Online]. Available: https://docs.sslmit.unibo.it/doku.php?id=resources:morph-it

[16] E. Bastianelli, A. Vanzo, P. Swietojanski, and V. Rieser, "Slurp: A spoken language understanding resource package," 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2011.13205

[17] Wikipedia, "Eigenvector centrality." [Online]. Available: https://en.wikipedia.org/wiki/Eigenvector_centrality

[18] S. Ronen, B. Gonçalves, K. Z. Hu, A. Vespignani, S. Pinker, and C. A. Hidalgo, "Links that speak: The global language network and its association with global fame," 2014. [Online]. Available: https://doi.org/10.1073/pnas.1410931111

[19] Wikipedia, "Amazon mechanical turk." [Online]. Available: https://it.wikipedia.org/wiki/Amazon_Mechanical_Turk

[20] Libero, "Lista nomi maschili." [Online]. Available: https://digilander.libero.it/gioer/nomimaschili.html

[21] L. spose di Erika, "Lista nomi maschili." [Online]. Available: https://www.lesposedierika.com/old/Matrimonio_Idee_e_Consigli/Nome_Bimbo.htm

[22] cs.cmu.edu, "Lista nomi maschili." [Online]. Available: https://www.cs.cmu.edu/Groups/AI/util/areas/nlp/corpora/names/male.txt

[23] arineng, "Lista nomi maschili." [Online]. Available: https://github.com/arineng/arincli/blob/master/lib/male-first-names.txt

[24] Libero, "Lista nomi femminili." [Online]. Available: https://digilander.libero.it/gioer/nomifemminili.html

[25] L. spose di Erika, "Lista nomi femminili." [Online]. Available: https://www.lesposedierika.com/old/Matrimonio_Idee_e_Consigli/Nome_Bimba.htm

[26] cs.cmu.edu, "Lista nomi femminili." [Online]. Available: https://www.cs.cmu.edu/Groups/AI/util/areas/nlp/corpora/names/female.txt

[27] arineng, "Lista nomi femminili." [Online]. Available: https://github.com/arineng/arincli/blob/master/lib/female-first-names.txt

[28] culturitalia.info, "Italian stop words - culturitalia." [Online]. Available: http://www.culturitalia.info/stop_ita.htm

[29] WeCanJob, "Elenco professioni." [Online]. Available: https://www.wecanjob.it/pagina9_elenco-professioni.html

[30] Wikipedia, "Pronomi." [Online]. Available: https://it.wikipedia.org/wiki/Pronome

[31] Europassitalian, "Pronomi." [Online]. Available: https://www.europassitalian.com/it/risorse-gratuite/grammatica/pronomi-e-particelle-pronominali/