

Premessa e stato dell'arte

La presente proposta si inserisce fra due ambiti di ricerca che caratterizzano gli studi linguistici attuali e sono soggetti a continui aggiornamenti. Da un lato, la crescente attenzione per la comunicazione inclusiva, intesa come quella che usa “un linguaggio inclusivo e libero da pregiudizi, non veicola stereotipi ed evita riferimenti a caratteristiche personali irrilevanti. Riconosce invece le qualità di ognuno indipendentemente dal genere, dall'orientamento sessuale, da condizioni di disabilità, dall'età, dall'appartenenza a un determinato gruppo sociale e da convinzioni religiose o di altro tipo”¹. Dall'altro lato, la traduzione automatica e il pre-editing, ovvero, l'applicazione di un insieme di raccomandazioni formali che stabiliscono parole o strutture specifiche da utilizzare in un testo dato per evitarne altre; questo sistema, chiamato anche *Controlled language*² (O'Brien 2003) è di particolare utilità nell'elaborazione di un testo di partenza nella traduzione multilingue dove i benefici del pre-editing si possono manifestare in ciascuno dei testi prodotti nella lingua di arrivo.

Alla confluenza della linguistica e del *Natural language processing*, il progetto E-MIMIC (*Empowering Multilingual Inclusive Communication*), attualmente finanziato da un PRIN del Ministero italiano, è nato per l'iniziativa congiunta del Politecnico di Torino e dell'Università di Bologna e del quale è partner il Centro d'eccellenza Jean Monnet *Artificial Intelligence for European Integration* dell'Università di Torino (<https://www.jmcoe.unito.it/home>) al fine di sviluppare un applicativo multilingue (IT, FR, ES) di supporto alla scrittura inclusiva per le pubbliche amministrazioni. Il progetto intende elaborare e inserire un insieme di criteri linguistici (elaborazione di dati sintetici) e discorsivi che contribuiscano, tramite l'annotazione di materiali autentici e la selezione di corpora ad hoc, all'addestramento delle reti neurali alla base della traduzione automatica (intralinguistica) per migliorarne le performance. Attualmente, l'applicazione è in fase di completamento per la lingua italiana (d'Italia) e in fase di elaborazione per il francese (di Francia).

Pertanto il presente progetto, finalizzato all'adattamento metodologico delle premesse del progetto E-MIMIC alla lingua francese, si inserisce in un ambito di ricerca – quello della comunicazione inclusiva e della traduzione automatica – certamente vasto, ma che negli ultimi decenni è stato protagonista di un nuovo filone molto promettente e fino ad ora solo parzialmente esplorato: la diffusione e la promozione della comunicazione inclusiva attraverso lo sviluppo del *Deep Learning*.

¹ Consiglio dell'Unione Europea. 2018. *La comunicazione inclusiva*. https://www.consilium.europa.eu/media/35431/it_brochure-inclusive-communication-in-the-gsc.pdf

² O'Brien, Sharon. 2003. *Controlling controlled English. An analysis of several controlled language rule sets*. In *Controlled language translation*. Dublin City University. 15-17 May 2003. EAMT/CLAW. <https://aclanthology.org/2003.eamt-1.12.pdf>

Proposta di progetto

Alla luce delle possibilità di adattamento a più lingue del progetto E-MIMIC³ brevemente descritto nella sezione precedente, si propone l'obiettivo di estendere alla lingua francese la metodologia sinora applicata alla lingua italiana, partendo dalla premessa che in termini d'inclusività linguistica le principali problematiche sono comuni alla maggior parte delle lingue romanze, anche tenendo conto dell'affinità che esiste tra loro.

Gli obiettivi generali alla base della proposta sono:

1. Favorire la comunicazione inclusiva della lingua francese in scenari reali di ambito amministrativo.
2. Individuare i meccanismi linguistici discriminatori nella lingua francese di natura:
 - grammaticale (ad esempio attraverso il silenziamento della forma femminile o in espressioni nominali al plurale che includono entrambi i generi);
 - semantica (come la presenza di stereotipi).
3. Riformulare il testo non inclusivo in modo inclusivo attraverso l'identificazione di meccanismi di inclusione linguistica sia di natura grammaticale che semantica, tenendo conto anche delle politiche linguistiche francofone e delle raccomandazioni già esistenti a livello amministrativo.
4. Sfruttare il dispositivo di *Deep Learning* sviluppato all'interno del progetto E-MIMIC per generalizzare e automatizzare il processo di riformulazione inclusiva su corpora specifici addestrandolo a rilevare le espressioni non inclusive e a suggerire alternative inclusive in lingua francese.

Fasi della ricerca e compiti dell'assegnista

Le fasi della ricerca riprendono in parte le fasi analitiche previste per il progetto E-MIMIC adattato alla lingua francese e in particolare:

1. Formazione iniziale:

La formazione iniziale avverrà tramite un corso di c.a. 16-20 ore durante il primo mese e mezzo di attività del progetto.

2. Selezione di corpora:

Creazione di un corpus in lingua francese utile all'ottimizzazione dell'apprendimento di algoritmi d'intelligenza artificiale preposti alla traduzione

³ Attanasio, G., Cagliero, L., Cerquitelli, Greco, S., La Quatra, M., Raus, R., Tonti, M. (2021), "E-MIMIC: Empowering Multilingual Inclusive Communication" in: *2021 IEEE International Conference on Big Data (Big Data)*, Orlando (USA), 15-18 dicembre 2021, pp. 4227-4234.

intralinguistica di testi amministrativi francesi in chiave inclusiva. Dovranno essere presi in considerazione diversi generi testuali come comunicazioni, notifiche, regolamenti, avvisi ecc., documenti provenienti dall'ambito dell'amministrazione francese.

3. Etichettatura dei testi:

Annotazione dei corpora selezionati utilizzando criteri linguistico-discorsivi predisposti dal team di ricerca e proposta di possibili riformulazioni adoperando un linguaggio inclusivo (parole generiche, epiceni, sostantivi collettivi, pronomi indefiniti...).

4. Predisposizione di dati sintetici

Redazione di paradigmi linguistici predisposti ad hoc sulla base di quanto constatato nella fase di etichettatura, che siano utili al miglioramento delle performance delle reti neurali di apprendimento profondo.

5. Test dell'applicativo

Partecipazione alle fasi di test dell'applicativo che sarà elaborato e messo nel frattempo a disposizione dal team informatico al fine di integrare i criteri linguistico-discorsivi e di migliorare la qualità del testo riformulato.

6. Disseminazione risultati della ricerca

Partecipazione al convegno di presentazione finale dei lavori effettuati, confrontandoli con quelli del team di ricerca sulla lingua italiana, ed eventuale pubblicazione del proprio intervento.

Profilo in entrata

Dati i compiti previsti per questo progetto, l'assegnista dovrà possedere la laurea magistrale o laurea equipollente in Lingue straniere.

Lingua richiesta: francese (almeno C1)

Costituiscono requisiti preferenziali:

- Dottorato di ricerca nelle discipline linguistiche e in particolare nella linguistica francese o nella linguistica computazionale;
- Pubblicazioni pertinenti rispetto alla tematica della ricerca;
- Competenze di tipo informatico (Python).

Formazione

La formazione verrà affidata alla tutor Rachele Raus (Università di Bologna), coadiuvata dal gruppo di ricerca di automatica e informatica coordinato da Tania Cerquitelli (Politecnico di Torino) e dal gruppo di ricerca d'italianistica e glottodidattica coordinato da Francesca Dragotto (Università di Roma 2 – Tor Vergata).