



UNIVERSITÀ
DI TORINO

Per un'intelligenza artificiale a favore del multilinguismo europeo

*Raccomandazioni
strategiche rivolte ai
decisioni europei*

a cura di

Rachele Raus

Università di Bologna

Membro del Centro
d'eccellenza Jean Monnet

*Artificial Intelligence for
European Integration*
Università di Torino

Per un'intelligenza artificiale a favore del multilinguismo europeo

Raccomandazioni strategiche rivolte ai decisori europei

a cura di

Rachele Raus

Università di Bologna

Membro del Centro d'eccellenza Jean Monnet

Artificial Intelligence for European Integration

Università di Torino



**Artificial Intelligence
for European Integration**
Jean Monnet Centre of Excellence



**UNIVERSITÀ
DI TORINO**

Artificial Intelligence for European Integration | Report - 2023

AI4EI

www.jmcoe.unito.it

Collane@unito.it
Università di Torino

ISBN ebook: 9788875902681
ISBN cartaceo: 9791256000036



Quest'opera è distribuita con
Licenza Creative Commons Attribuzione.
Condividi allo stesso modo 4.0 Internazionale.
Copyright © 2023



**Artificial Intelligence
for European Integration**
Jean Monnet Centre of Excellence



With the support of the
Erasmus+ Programme
of the European Union

Ledizioni 
The Innovative LEDipublishing Company

Ledizioni LediPublishing
Via Antonio Boselli, 10
20136 Milano – Italia
www.ledizioni.it
info@ledizioni.it

INDICE

7 **Introduzione**

Rachele Raus | *Università di Bologna*

RACCOMANDAZIONE 1

Occorre investire su nuovi tipi di formazione critica sull'intelligenza artificiale

13 **L'intelligenza artificiale e il multilinguismo europeo**

Dardo de Vecchi | *KEDGE Business School*

17 **L'intelligenza artificiale per la professionalizzazione delle competenze multilingui in Europa**

Maria Margherita Mattioda - Ilaria Cennamo | *Università di Torino* - Silvia Domenica Zollo | *Università di Napoli "Parthenope"*

23 **Intelligenza artificiale, traduzione automatica e apprendimento linguistico**

Alessandra Molino - Lucia Cinato | *Università di Torino*

27 **Intelligenza artificiale e formazione in traduzione: nuove competenze per traduttori e revisori specialistici**

Maria Teresa Zanola - Anna Serpente - Martina Ali | *Università Cattolica del Sacro Cuore, Milano*

RACCOMANDAZIONE 2

Occorre investire su nuovi profili professionali nell'industria linguistica

31 **Investire su nuovi profili di mestiere**

Danio Maldussi | *Università di Bergamo* - Micaela Rossi | *Università di Genova*

35 **Nuovi profili professionali a cavallo tra lingue e tecnologie. Risultati dell'analisi dei bisogni condotta nel progetto Erasmus+ UPSKILLS**

Silvia Bernardini - Adriano Ferraresi - Maja Miličević Petrović | *Università di Bologna*

39 **La figura professionale dell'annotatore per garantire una maggiore affidabilità dei dati provenienti dall'Intelligenza Artificiale**

Michela Tonti | *Università di Bergamo*

45 **Bias di genere e intelligenza artificiale: una mancanza di competenze?**

Mara Floris | *Università Vita-Salute San Raffaele*

RACCOMANDAZIONE 3

Occorre investire sulla creazione di corpora multilingui elaborati a partire da materiale autentico nazionale rispettoso della diatopia linguistica

- 49 **Gestione della diversità linguistica e intelligenza artificiale: rischi e opportunità**
Giovanni Agresti | *Université Bordeaux Montaigne*

- 55 **I corpora come risorse per l'uguaglianza digitale tra le lingue ufficiali dell'Unione Europea**
Federico Gaspari | *Università di Napoli "Federico II"*

- 61 **La lingua italiana: salvaguardia delle varietà linguistiche dall'impatto con l'intelligenza artificiale**
Chiara Russo | *Università degli Studi di Catania*

- 65 **Corpora multilingui rispettosi della diatopia: il caso del Québec**
Valeria Zotti | *Università di Bologna*

- 69 **Corpora multilingui e lingue speciali: preservare la diatopia linguistica**
Marta Muscariello | *Università IULM di Milano*

RACCOMANDAZIONE 4

Occorre investire nello sviluppo di tecnologie linguistiche e computazionali fattivamente *made in EU*

- 73 **Indipendenza tecnologica e diversità culturale nell'intelligenza artificiale europea**
Moreno La Quatra | *Università degli studi di Enna "Kore"*

- 79 **Investire nello sviluppo di tecnologie "made in EU"**
Alida Maria Silletti | *Università di Bari*

- 83 **Verso un'intelligenza artificiale europea e trasparente**
Giuseppe Attanasio | *Università Bocconi*

- 87 **Proposta di creazione di un gruppo di lavoro dell'Unione europea per l'elaborazione di corpora multilingui e multimodali in risposta a situazioni di multi-crisi**
Federico Garcea | *Università di Bologna*

91 **Risorse terminologiche plurilingui nel rispetto dei principi FAIR del web semantico**

Silvia Calvi - Klara Dankova - Lucrezia Marzo - Maria Teresa Zanola | *Università Cattolica del Sacro Cuore, Milano*

93 **Per un quadro di riferimento europeo comune di valutazione delle tecnologie della traduzione basate sull'IA**

Philippe Langlais | *Université de Montréal - François Yvon | Sorbonne Université*

97 **ALLEGATO**

Lista delle Università e degli enti di ricerca che hanno collaborato in forme diverse tramite il loro personale, alle ricerche condotte dal *Panel* sui diritti linguistici e l'IA del Centro d'eccellenza *Jean Monnet Artificial Intelligence for European Integration*

99 **GLOSSARIO**

Introduzione

Questo report intende fornire ai *policy* e *decision makers* dell'Unione europea degli argomenti e dei dati utili al fine d'investire su un'intelligenza artificiale a favore del multilinguismo europeo, proponendo quattro raccomandazioni specifiche in tal senso.

Precisiamo che in questa sede, con il termine "intelligenza artificiale"¹ s'intende la "scienza che si propone di sviluppare sistemi informatici intelligenti, riproducendo i processi mentali più complessi" (IATE: scheda terminologica 3571274²). Per "industria delle lingue", invece, va inteso l'insieme dei "prodotti, tecniche, attività o servizi che richiedono il trattamento automatico della lingua naturale (IATE: scheda terminologica 921669, definizione disponibile in francese).

In un periodo di diffusione di "*large language models*", ovvero di modelli informatici che prevedono un apprendimento profondo automatizzato, e perciò non supervisionato dalla componente umana, o semiautomatizzato e basato su grandi quantità di dati generalmente recuperati da Internet, come nell'esempio molto discusso di ChatGPT, è opportuno riflettere in modo obiettivo e scientifico su tali modelli in modo da intraprendere misure e azioni comuni che possano salvaguardare uno dei valori principali dell'UE: il multilinguismo.

A tal fine, il gruppo di lavoro che ho avuto il piacere di coordinare sulle attività linguistiche e che si è costituito attorno alla rete del Centro d'eccellenza Jean Monnet *Artificial Intelligence for European Integration* (AI4EI)³ a Torino, rete confluita attualmente nell'Osservatorio⁴ inaugurato all'Università di Torino il 12 dicembre 2023, ha inteso far dialogare persone esperte nei settori della linguistica, della glottodidattica, della traduzione, dell'analisi del discorso, dell'informatica e dell'ingegneria per l'informazione al fine di produrre dati e ricerche sull'impatto dell'IA nel settore dell'industria delle lingue e quindi, in ultimo, sul multilinguismo dell'Unione europea. Il metodo adottato, presentato in fase progettuale, è stato quello della ricerca "circolare", che ha inteso incrociare il modello tradizionale *bottom-up* con un modello *top-down* per il quale è stato coinvolto personale avente funzioni molto diverse (dirigenza, docenza, ma anche componente studentesca, persone con assegni di ricerca o in dottorato, di varie età, genere e diversa esperienza).

¹ Segnaliamo la presenza al fondo di un Glossario dei termini principali utilizzati nel presente report.

² <https://iate.europa.eu>.

³ <https://www.jmcoe.unito.it/home>.

⁴ <https://www.observatory.unito.it/>

Per un'intelligenza artificiale a favore del multilinguismo europeo

Raccomandazioni strategiche rivolte ai decisori europei

Rachele Raus

Università di Bologna

Membro del Centro d'eccellenza Jean Monnet

Artificial Intelligence for European Integration

Università di Torino

Tali ricerche hanno contribuito a far emergere in modo evidente gli attuali *bias*, ovvero distorsioni della realtà, errori, che possono produrre veri e propri pregiudizi, di tipo linguistico, sociale e culturale diffusi dall'industria linguistica basata sull'apprendimento profondo e perciò hanno condotto a stilare quattro raccomandazioni strategiche, rivolte ai decisori dell'UE, per evitare gli impatti negativi che l'IA potrebbe produrre nei prossimi anni sul multilinguismo europeo.

Questo report è il frutto di un lavoro cominciato dal 6 ottobre 2020 con l'invito a partecipare rivolto a Laurent Romary, direttore del Comitato tecnico 37 dell'International Organization for Standardization (ISO), nell'ambito del convegno di apertura⁵ del progetto del Centro d'eccellenza Jean Monnet AI4EI, con il quale abbiamo posto la questione, segnalata da più parti (Vetere 2023, Kim *et al.*), di un utilizzo problematico dell'inglese non solo come lingua veicolare europea maggiormente utilizzata sia nella redazione sia nella traduzione dei testi europei, sui quali peraltro sono spesso addestrati gli algoritmi d'IA, ma anche dell'esistenza della maggior parte dei prodotti legati alla tecnologia linguistica in inglese.

A queste prime riflessioni ha fatto seguito un convegno internazionale nelle giornate del 23 e 24 aprile 2021⁶ nelle quali si è riflettuto da un lato sull'impatto dell'IA rispetto al multilinguismo, riflessioni confluite nel volume *Multilinguismo e variazioni linguistiche in Europa nell'era dell'intelligenza artificiale (Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence)*⁷, e dall'altro sono stati presentati a Torino i risultati di una prima *survey*, effettuata durante l'anno accademico 2020-2021, alla quale ne sono seguite due analoghe nei due anni accademici successivi, presso la componente studentesca di diverse università italiane e francesi (sono state coinvolte 18 Università di cui 10 italiane e 8 francesi per un totale, sui tre anni, di 3.328 questionari raccolti) sull'utilizzo di dispositivi di traduzione automatica basati sull'apprendimento profondo (*deep learning*) prima e dopo aver seguito corsi di formazione critica sull'uso dell'IA nell'industria linguistica.

Nel corso dei tre anni, inoltre, la prima rete di Università italiane e francesi coinvolte si è ampliata con la partecipazione a diversi eventi internazionali (es. le *Assises de la Francophonie scientifique*⁸) o di scambi di ricerca con altri Centri Jean Monnet (ad

⁵ <https://www.jmcoe.unito.it/content/kick-conference-ai4ei>.

⁶ <https://www.jmcoe.unito.it/content/linguistic-rights-and-language-varieties-europe-age-ai>.

⁷ <https://www.collane.unito.it/oa/items/show/132#?c=0&m=0&s=0&cv=0>.

⁸ <https://www.auf.org>.

esempio, l'*Hawke EU Center Jean Monnet dell'University of South Australia*⁹) o altre Università estere (ad esempio, l'*Universidade Estadual de Campinas*¹⁰ a San Paolo in Brasile), arrivando a coinvolgere oltre trenta enti e istituti di ricerca basati in Italia, Belgio, Francia, Spagna, Australia, Canada e Brasile¹¹.

Una seconda pubblicazione della rete, inerente alle attività sull'IA e i linguaggi naturali svolte nel corso del 2021-2023, è in attesa di pubblicazione presso il sito del Centro d'eccellenza AI4EI.

Questi scambi e riflessioni tra esperti universitari e non (hanno partecipato anche alcune rappresentanze di realtà private) hanno permesso di approfondire le criticità legate a un utilizzo acritico dell'intelligenza artificiale e messo in luce la necessità di investire su nuovi tipi di formazione critica all'IA (raccomandazione 1), anche funzionalmente alla necessità di investire su nuovi tipi di professione e di professionalizzazione richieste dal mercato del lavoro nell'industria linguistica (raccomandazione 2).

È emerso, inoltre, l'imperativo di investire nello sviluppo di corpora multilingui (raccomandazione 3). È importante anzitutto capire che, sebbene a livello di apprendimento l'IA impari da dati, i dati presenti nel *web* o predisposti per l'apprendimento sono di fatto dei "corpora" (Rastier 2021), cioè materiale che va opportunamente contestualizzato e selezionato ai fini della ricerca e dell'apprendimento dell'IA. I dati, infatti, danno la percezione di un'informatività "neutra" che i testi utilizzati per l'addestramento delle reti neurali, e che sono strettamente legati all'interpretazione umana, non hanno. In tal senso, l'uso consapevole dei dati, considerati come corpora, e soprattutto lo sviluppo di corpora nelle lingue europee permetterebbe di difendere la loro diatopia costitutiva, ovvero la loro specificità legata a una data cultura e a un dato territorio, ivi comprese le lingue "minoritarie", riequilibrando peraltro la disuguaglianza di risorse informatiche disponibili in lingua inglese rispetto alle altre lingue europee.

Questo tipo di questione è strettamente legata ad altre due problematiche: da un lato, la carenza di un'IA fattivamente "*made in Europe*", dato che la maggior parte dei modelli e delle tecnologie informatiche legate al *deep learning* proviene da paesi extraUE; dall'altro, la diffusione sempre più massiva di modelli non supervisionati, maggioritariamente prodotti extraUE, generano *bias* di tipo linguistico, sociale e culturale di cui l'UE non può te-

⁹<https://www.unisa.edu.au/research/hawke-eu-centre-for-mobilities-migrations-and-cultural-transformations/>.

¹⁰ <https://www.unicamp.br/unicamp/>.

¹¹ Nell'annesso finale riportiamo le Università di afferenza delle persone e dei Centri di ricerca che hanno contribuito alle ricerche linguistiche del Centro d'eccellenza durante i 3 anni del finanziamento Erasmus+.

neri in conto in un'ottica di promozione dell'inclusione sociale e del multilinguismo. In tal senso, diventa quindi fondamentale investire su un'IA realmente *made in Europe* (raccomandazione 4), come peraltro richiesto dalla Commissione europea sin dal Piano coordinato sull'intelligenza artificiale del 2018 (COM 2018(795): 1). Per un'IA che sia di fatto *made in Europe* non è sufficiente "assemblare" dispositivi in Europa ma occorre sviluppare ricerca e tecnologia informatica effettuate in Europa che tengano conto della necessità di adottare approcci e modelli informatici supervisionati dalla componente umana, che possano evitare proprio i *bias* presenti nei *large language models*.

In tal senso, questo report si allinea ad altre iniziative consimili che chiedono un'attenzione maggiore a livello legislativo ed etico in merito all'IA prodotta in Europa, come ad esempio l'iniziativa della petizione portata avanti dal progetto LAION¹², che chiede all'Unione europea di creare strutture di supercalcolo su larga scala per la ricerca sull'IA, consentendo alla comunità di ricerca europea di studiare modelli di fondazione *open-source* in condizioni controllate con una supervisione pubblica¹³, o i progetti promossi dalla rete *Human AI Net*¹⁴.

Ognuna delle quattro raccomandazioni che qui riportiamo in sintesi:

1. Investire su nuovi tipi di formazione critica sull'intelligenza artificiale, ovvero in grado di promuovere un utilizzo consapevole di dispositivi dell'industria linguistica supportati da algoritmi di apprendimento profondo;
2. Investire su nuovi profili professionali nell'industria linguistica;
3. Investire sulla creazione di corpora multilingui elaborati a partire da materiale autentico nazionale rispettoso della diatopia linguistica;
4. Investire nello sviluppo di tecnologie linguistiche e computazionali fattivamente *made in EU*

sono argomentate da schede scientifiche divulgative redatte da alcune delle persone esperte che hanno contribuito alle attività del Centro d'eccellenza AI4EI in questi tre anni e che hanno inteso supportare tali raccomandazioni sulla base dei dati e delle ricerche raccolte.

Nel ringraziare tutte le persone che hanno contribuito al progetto per aver permesso un confronto costruttivo su tali tematiche, ci auguriamo che questo report consenta di inaugurare politiche eu-

¹² <https://laion.ai/notes/letter-to-the-eu-parliament/>.

¹³ «The EU should establish large-scale supercomputing facilities for AI research, enabling the European research community to study open-source foundation models under controlled conditions with public oversight». <https://www.unite.ai/laion-and-a-group-of-27/>.

¹⁴ <https://www.humane-ai.eu/research-roadmap>

ropee incisive in merito all'intelligenza artificiale in un momento di grandi cambiamenti tecnologici e sociali che chiedono la massima attenzione e fanno riflettere sul futuro dell'Europa e sui modelli informatici, culturali e linguistici che intenderà adottare.

Bibliografia

Commissione europea (2018). *Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al comitato delle regioni. Piano coordinato sull'intelligenza artificiale*. Bruxelles: Commissione Europea, 795 finale.

Kim Yunsu, Petrov Petre, Petrushkov Pavel, Khadivi Shahram, Ney H. (2019). "Pivot-based Transfer Learning for Neural Machine Translation between Non-English Languages". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, China, November 3-7, 2019*. Association for Computational Linguistics, 866-876. DOI: 10.18653/v1/D19-1080.

Rastier Francois (2021). "Data vs Corpora". In: Damon Mayaffre, Laurent Vanni (a cura di) *L'intelligence artificielle des textes : des algorithmes à l'interprétation*. Parigi: Champion, 203-245.

Vetere Guido (2023). "Elaborazione automatica dei linguaggi diversi dall'inglese: introduzione, stato dell'arte e prospettive". In: Rachele Raus (cur.) et al. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l'aune de l'intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell'era dell'intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 69-87. <https://www.collane.unito.it/oa/items/show/13>

RACCOMANDAZIONE 1

Occorre investire su nuovi tipi di formazione critica sull'intelligenza artificiale, ovvero in grado di promuovere un utilizzo consapevole di dispositivi dell'industria linguistica supportati da algoritmi di apprendimento profondo.

L'IA fa parte della realtà quotidiana dell'UE al di là delle sue stesse frontiere. Un tale choc tecnologico è ormai incorporato nella nostra realtà quotidiana. Ad esempio, nel caso della traduzione automatica, non si può non constatare che essa ha fatto dei progressi innegabili al punto che normalmente ci si fida dei risultati ottenuti da tale operazione. D'altronde, contestare un tale risultato non è semplice. Infatti, chi è in grado di argomentare l'affidabilità di una traduzione automatica disponibile immediatamente e, per di più, gratuitamente? Il numero di persone capaci di farlo è ridottissimo mentre quello di chi usa la traduzione automatica è incalcolabile. Nella massa di persone che affidano i propri documenti ai traduttori automatici ci sono non solo studenti di numerose discipline, per i quali la lingua non è oggetto di riflessione (de Vecchi 2022), ma anche un numero considerevole di aziende per le quali le lingue sono (solo) uno strumento necessario al proprio funzionamento e non una materia prima. Le Scuole di management formano alla gestione aziendale ma non sembra che uno sguardo critico sull'IA sia all'orizzonte nelle loro priorità. L'entusiasmo attuale per l'utilizzo di queste nuove tecnologie mostra semmai il contrario (de Vecchi 2022). In entrambi i casi, sarebbe necessario che la cittadinanza comprendesse da un lato quali sono le capacità umane a effettuare un certo numero di mansioni e dall'altro i vantaggi di utilizzare l'IA per svolgerle.

Una visione strategica dell'IA non può limitarsi a considerare la sua funzionalità, ivi compresa la velocità nel trattare i dati, ma dovrebbe tener conto anche delle conseguenze dell'eliminare il fattore umano in un'Europa che, paradossalmente, vorrebbe porlo al centro delle proprie preoccupazioni (Commissione europea 2019).

Ne consegue che la parola d'ordine dovrebbe essere la comprensione di ciò di cui stiamo trattando dal punto di vista tecnico e sociale. Qualsiasi politica concernente l'IA dovrebbe quindi porre al centro un accompagnamento al suo utilizzo che, conseguentemente, non dovrebbe mai essere assente nei programmi educativi, soprattutto in un'Europa multilingue.

Dal momento che si torna al multilinguismo, è l'umano che deve essere protagonista degli applicativi basati sull'apprendimento profondo nella gestione delle lingue. L'IA non ha la "capacità", ad esempio, di decidere quali siano le ragioni che fanno sì che una lingua debba essere salvaguardata, promossa, tradotta o insegnata. Ancor meno, possiede la facoltà di stabilire strategie educative o di gestione delle lingue "maggioritarie" o "minoritarie"¹.

¹ Per una definizione di lingua "minoritaria" e una discussione su tali nozioni, cfr. Agresti in questo report.

S'impone quindi una riflessione sul patrimonio linguistico. Questione che l'IA non può risolvere con gli strumenti attuali.

Una tale riflessione è necessaria a livello europeo quando si tratta di pensare a collaborazioni interdisciplinari e a strategie a lungo termine che considerano l'inglese come lingua imprescindibile di cui sembra impossibile fare a meno. E tuttavia, la diglossia, ovvero l'utilizzo di due lingue all'interno della stessa comunità di parlanti, insegna che lingue diverse possono svolgere funzioni specifiche in seno allo stesso gruppo sociale, se non addirittura in una società. Le lingue possono diventare ponti invece che ergere barriere. L'IA può gestire questa dicotomia linguistica? È molto probabile se ben gestita.

Sebbene le negoziazioni o il commercio si basino spesso sull'inglese, ci si prepara a entrambe utilizzando normalmente la propria lingua materna, che sola permette di creare un legame di fiducia. L'UE dovrebbe perciò investire non solo nella comprensione dell'IA, ma anche nella formazione a riflettere sul suo utilizzo per evitare, ad esempio, che la cittadinanza possa essere fuorviata in negoziazioni che non avvengano nella propria lingua. La storia di Internet ci ricorda quanto un tale strumento sia sfuggito dalle mani di chi lo ha creato, superando le loro attese.

Da un punto di vista linguistico, è quindi fondamentale insegnare le differenze che esistono tra linguaggio, lingua e discorso. Occorre anzitutto ricordare la natura stessa del linguaggio inteso come sistema di rappresentazione del pensiero di cui è strumento. La necessità di un tale distinguo è avvertita dagli specialisti e dai tecnici che non possono prescindere dal momento che l'IA tratta o, meglio, manipola, realtà diverse. Tale distinzione deve essere insegnata e ricordata perché è indispensabile nel momento in cui si "alimenta", la metafora merita di essere segnalata, l'IA con dei dati. Il modello *No Language Left Behind* o NLLB-200 parla di 200 lingue trattate, cosa che non può che confortarci; tuttavia, il multilinguismo, soprattutto quello europeo, non deve lasciarsi sedurre dal canto di queste sirene ma pensare, semmai, a formare alla presa di coscienza del patrimonio linguistico, alla sua importanza e impatto nella società in cui viviamo. Si pone quindi legittimamente la questione di sapere chi deciderà le lingue da "trattare".

Per ultimo, la questione fondamentale ruota attorno ai valori alla base delle nostre azioni e per i quali vorremmo costruire un multilinguismo europeo all'altezza delle nostre aspettative. Vogliamo farlo noi o vogliamo lasciarlo fare alle macchine, eliminando ogni intervento umano dal processo? Diego Marani ha scritto, nel suo *L'ultimo dei Vostiacchi* (2001) che occorrono tutte le lin-

gue del mondo per far sopravvivere l'umanità. È un multilinguismo europeo ragionato e coscienzioso che potrà far sopravvivere il cittadino europeo.

Bibliografia

Commissione europea (2019). *Creare fiducia nell'intelligenza artificiale antropocentrica*. Bruxelles: Commissione Europea, 168

De Vecchi Dardo (2022). « Le multilinguisme européen et l'IA. Enquête auprès des futurs décideurs ». In: Rachele Raus (cur.) et al. De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l'aune de l'intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell'era dell'intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 215-245. <https://www.collane.unito.it/oa/items/show/132>.

Marani Diego (2009). *L'ultimo dei Vostiacchi*. Milano: Bompiani.

Meta AI alla pagina: <https://ai.facebook.com/research/no-language-left-behind/>

Il progresso tecnologico indotto dall'Intelligenza Artificiale (IA) multilingue (Yvon 2022) ha un impatto sociale ed economico esteso¹, che ingloba sia la sfera educativa² sia il mercato del lavoro (ELIS 2022: 25; Zilner *et al.* 2021).

La diffusione massiva di tecnologie multilingui basate sul *Deep Learning* invita a riflettere, in particolare, sulla definizione di pratiche didattiche innovative (Cennamo, De Faria Pires 2022) che meritano di essere promosse nel panorama della formazione universitaria in lingue straniere (LS) allo scopo di rispondere alle esigenze di professionalizzazione delle competenze multilingui nel mercato digitale europeo. Divenuta progressivamente parte integrante dei servizi linguistici e di traduzione in Europa (EMT 2022: 7), l'IA multilingue suscita un interesse crescente anche nel settore del business internazionale³: le ricerche scientifiche applicate a questo settore (Lecomte *et al.* 2023) sottolineano, infatti, il ruolo strategico delle competenze multilingui e traduttive nei processi di internazionalizzazione delle organizzazioni europee, nonché l'interesse crescente per le nuove tecnologie la cui integrazione nei flussi di lavoro risulta, tuttavia, ancora a uno stadio sperimentale (Wilmot 2022: 86).

In tale contesto, l'Unione Europea (UE) riconosce il multilinguismo e l'IA come due aree di importanza strategica, visto il potenziale di entrambi i settori per lo sviluppo economico e sociale della stessa. Numerose sono le azioni intraprese dalla Commissione europea al fine di promuovere e incoraggiare gli Stati membri a sviluppare strategie nazionali di IA negli ambiti dell'istruzione e della formazione (vedasi i documenti sull'IA per l'Europa 2018; il Piano Coordinato sull'IA 2018; la Revisione del Piano Coordinato sull'IA 2021; il Libro bianco sull'IA 2020). Più nello specifico, l'UE rileva come l'IA e le tecnologie linguistiche ad essa correlate possano essere utilizzate nello sviluppo o nell'applicazione di nuovi metodi di insegnamento in settori quali l'insegnamento-apprendimento delle LS, salvaguardando di conseguenza il multiculturalismo, il multilinguismo e, più in generale, la diversità culturale e linguistica europea e nazionale anche nel mondo accademico.

¹ Come evidenziato dal seminario *"Impacts sociétaux de l'intelligence artificielle"* tenutosi il 10 novembre 2022 presso l'*Université Bordeaux Montaigne*. <https://www.u-bordeaux-montaigne.fr/fr/actualites/recherche/impacts-sociaux-de-l-intelligence-artificielle.html>.

² Come emerso dal *Translating Europe Workshop* "L'intelligenza artificiale per la traduzione: verso una nuova progettazione didattica?" organizzato dalla DGT della Commissione europea e dall'Università degli Studi di Torino il 3 dicembre 2021. https://italy.representation.ec.europa.eu/system/files/2021-11/TEW_Torino%20-%20Programma.pdf.

³ Come si evince dallo sviluppo di applicativi generativi e servizi intelligenti destinati alle aziende, tra cui: <https://www.oneai.com/>.

L'intelligenza artificiale per la professionalizzazione delle competenze multilingui in Europa

**Maria Margherita Mattioda,
Ilaria Cennamo**
Università di Torino

Silvia Domenica Zollo
Università di Napoli "Parthenope"

Tuttavia, affinché l'impiego dell'IA sia vantaggioso per il settore della formazione multilingue in contesto universitario, è necessario che l'UE promuova maggiormente l'educazione linguistica e digitale per sostenere lo sviluppo delle competenze del futuro attraverso un approccio etico, responsabile e trasversale verso le tecnologie basate sull'IA.

Da un punto di vista pedagogico, sosteniamo che tale formazione interdisciplinare non debba essere circoscritta alla componente studentesca delle filiere tecnico-scientifiche (si pensi alle classi di laurea in informatica e ingegneria), ma al contrario debba essere rivolta al maggior numero possibile di settori disciplinari, coinvolgendo visibilmente anche le scienze umane, politiche e sociali, economico-statistiche e aziendali, sì da rispondere alle esigenze di un mercato lavorativo in rapida evoluzione e sempre più digitale e multilingue.

Emerge dunque l'urgenza di valutare in che modo integrare in maniera strategica e mirata le tecnologie digitali nelle nuove pratiche didattiche per l'insegnamento-apprendimento delle LS in tutti i contesti universitari e come l'UE possa sostenere gli sforzi degli Stati membri in quanto responsabili delle politiche del lavoro e dell'istruzione. Un prerequisito fondamentale per affrontare questa profonda trasformazione riguarda la capacità di saper formare figure ibride con competenze al contempo trasversali e strumentali (Zollo 2022) e di investire nell'alfabetizzazione digitale (meglio conosciuta come *digital literacy*), nonché nella più specifica *machine translation literacy* (Bowker, Ciro 2019; Bowker 2020; Bowker 2021; Loock, Léchauguette 2021) anche attraverso meccanismi di *lifelong learning*. L'acquisizione di tali competenze passa attraverso un approccio didattico ragionato e critico che, grazie all'analisi del potenziale e dei limiti che si possono osservare all'interno delle prestazioni automatiche delle tecnologie neurali, favorisce la valorizzazione delle competenze linguistiche, traduttive e comunicative umane e, quindi, un'interazione più consapevole con la macchina (Cennamo e Mattioda 2022). Tale approccio didattico risponde in modo efficace alla necessità di formare profili professionali dotati di competenze plurilingui modulabili (Miličević *et al.* 2021), che permettano di operare nei diversi contesti della cooperazione internazionale. Inoltre, alla luce dei molteplici usi a cui si prestano le tecnologie multilingui, come la traduzione automatica (Monti 2019: 20) — che possono essere concepite e adottate come ausilio alla traduzione, alla scrittura, alla comprensione e all'interazione in lingua straniera, e come risorsa per l'apprendimento di una nuova lingua — la formazione universitaria dispone di una pluralità di strumenti e re-

lative modalità di integrazione che si rivelano attuabili nei diversi contesti di insegnamento-apprendimento delle LS applicate ai contesti professionali poc'anzi citati, siano essi previsti come insegnamenti caratterizzanti, integrativi e affini. L'insegnamento delle LS dovrebbe quindi evolvere verso obiettivi interdisciplinari secondo cui l'acquisizione di competenze linguistiche e digitali (*AI-based*) venga intesa in una prospettiva applicata, ad esempio, alla comunicazione aziendale, ai linguaggi settoriali, all'ambito istituzionale, alla traduzione, all'interazione orale e alla creazione di contenuti multilingui. L'interdisciplinarietà dovrebbe altresì fondarsi sullo sviluppo di *soft skills* quali creatività, managerialità e capacità di risolvere problemi in contesti multilingui e tecnologici sempre più complessi. Di particolare rilevanza potrebbe essere l'attivazione di percorsi di formazione e tirocini da svolgere presso aziende, centri di ricerca ed enti pubblici europei dove la componente studentesca sia chiamata a risolvere problemi reali, tramite l'applicazione di metodi e tecnologie linguistiche proprie dell'IA, come emerso dai risultati delle ricerche condotte nell'ambito del progetto *Artificial Intelligence for European Integration* (AI4Ei) (Raus et al. 2022).

Su un piano politico, l'orizzonte strategico dell'"Università europea"⁴ delineato per il 2024, che comprende la definizione di criteri comuni per l'istituzione di un diploma europeo congiunto, necessiterebbe di un'esplicita valorizzazione del multilinguismo e della *digital literacy* applicata alle tecnologie multilingui nel quadro delle università "europee" che, in quanto tali, intendano operare di concerto per favorire lo sviluppo di competenze e di conoscenze rispondenti alle specifiche esigenze di un mercato unico europeo, multilingue e digitale. Nello specifico, nel quadro della strategia europea per le università appare necessario chiarire quale ruolo strategico venga attribuito all'insegnamento-apprendimento delle LS (diverse dall'inglese), in particolare, nell'ambito di corsi di studio universitari che promuovono un'apertura internazionale essenzialmente mediante un'offerta formativa in lingua inglese (lingua franca). In tal senso, si sottolinea come le politiche europee giochino un ruolo determinante nell'orientamento delle future strategie di internazionalizzazione in quanto occorre incentivare le università europee ad elaborare percorsi formativi e progetti di ricerca internazionali che siano rappresentativi della diversità e dell'innovazione europee nel mondo.

Per concludere, si sottolinea quindi, *in primis*, l'importanza legata all'investimento in una didattica delle lingue innovativa che

⁴ European Universities Initiative. Cfr. <https://education.ec.europa.eu/education-levels/higher-education/european-universities-initiative>

attualmente non può più prescindere dalla diffusione capillare dell'IA e delle sue innumerevoli applicazioni multilingui. *In secundis*, si ribadisce la rilevanza di una formazione universitaria multilingue incentrata sull'apprendimento di lingue europee diverse dall'inglese in una prospettiva di internazionalizzazione europea, che preveda ad esempio, accanto ad insegnamenti impartiti in inglese (lingua franca), la formazione obbligatoria in almeno un'altra lingua europea, contribuendo così *de facto* alla realizzazione di un'università europea multilingue e innovativa.

Bibliografia

Bowker Lynne (2021). "Promoting linguistic diversity and inclusion: Incorporating machine translation literacy into information literacy instruction for undergraduate students". *The International Journal of Information, Diversity and Inclusion*, Volume 5, Issue 3, 127-151.

Bowker Lynne (2020). "Machine translation literacy instruction for international business students and business English instructors". *Journal of Business and Finance Librarianship*, Volume 25, Issue 1-2, 25-43.

Bowker Lynne, Buitrago Ciro Jairo (dir.) (2019). *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. Bingley: Emerald. DOI: 10.1108/9781787567214

Cennamo Ilaria, De Faria Pires Loic (2022). "Intelligence artificielle et traduction. Les défis pour la formation et la profession". In: Maggi Ludovica, Bordes Sarah (a cura di). *Intelligences pour la traduction. IA et interculturel: actions et interactions, Revue internationale d'interprétation et de traduction / International Journal of Interpretation and Translation FORUM [20:2] Special Issue*. Amsterdam: John Benjamins Publishing Company, 333-356. <https://doi.org/10.1075/forum.20.2>

Cennamo Ilaria, Mattioda Maria Margherita (2022). "La traduzione automatica neurale: uno strumento di sensibilizzazione per la formazione universitaria in lingua e traduzione francese". In: Rachele Raus (cur.) et al.. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l'aune de l'intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell'era dell'intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 307-331. <https://www.collane.unito.it/oa/items/show/132>

Commissione Europea (2021). *Piano coordinato sulla revisione dell'intelligenza COM(2021), 205 final*. Bruxelles, Commissione Europea. <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>

Commissione Europea (2020). *Un'agenda per le competenze per l'Europa per la competitività sostenibile, l'equità sociale e la resilienza COM(2020), 274 final*, Bruxelles, Commissione Europea. <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:52020DC0274&from=EN>

Commissione Europea (2020). *Libro bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia COM(2020), 65 final*. Bruxelles, Commissione Europea. https://commission.europa.eu/document/d2ec4039-c5be-423a-81ef-b9e44e79825b_it

Commissione Europea (2020). *The future of language education in Europe: case studies of innovative practices*, Lussemburgo: Ufficio Pubblicazioni dell'Unione Europea. https://nesetweb.eu/wp-content/uploads/2020/05/NESET_AR_2020_Future-of-language-education_Full-report.pdf

Commissione Europea (2018). *Piano coordinato sull'intelligenza artificiale COM(2018), 795 final*. Bruxelles, Commissione Europea. <https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52018DC0795&from=DE>

Commissione Europea (2018). *L'intelligenza artificiale per l'Europa COM(2018), 237 final*. Bruxelles, Commissione Europea. <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=COM%3A2018%3A237%3AFIN>

Curry Edward, Metzger Andreas, Zillner Sonja, Pazzaglia Jean-Christophe, García Robles Ana (2021). "Data economy 2.0: From big data value to AI value and a European data space". In: *The Elements of Big Data Value: Foundations of the Research and Innovation Ecosystem*. Cham: Springer International Publishing, 379-399. DOI: <https://doi.org/10.1007/978-3-030-68176-0>

ELIS (2022). *European Language Industry Survey 2022 Trends, expectations and concerns of the European language industry*. https://elis-survey.org/wp-content/uploads/2022/03/ELIS-2022-report.pdf?utm_source=elis-repository&utm_medium=website&utm_campaign=elis-report22&utm_id=elis-report-22

European Master's in Translation (EMT) (2022). *Competence Framework 2022*. <https://termcoord.eu/2022/11/updated-version-of-the-emt-competence-framework-now-available/>

Lecomte Philippe, Vigier Mary, Gaibrois Claudine, Beeler Betty (2023). *Understanding the Dynamics of Language and Multilingualism in Professional Contexts*. Cheltenham: Edward Elgar Publishing.

Loock Rudy, Léchauguette Sophie (2021). "Machine translation literacy and undergraduate students in applied languages: report on an exploratory study". *Tradumàtica*, 19, 204-225.

Miličević Petrović Maja, Bernardini Silvia, Ferraresi Adriano, Aragrande Gaia, Barrón-Cedeño Alberto (2021). *Language data and project specialist: A new modular profile for graduates in language-related disciplines*. UPSKILLS Intellectual output 1.6. Zenodo. <https://dx.doi.org/10.5281/zenodo.5030929>

Monti Johanna (2019). *Dalla Zairja alla traduzione automatica. Riflessioni sulla traduzione nell'era digitale*. Napoli: Paolo Loffredo Editore.

Raus Rachele, Silletti Alida, Zollo Silvia Domenica, Humbley John (2022). « Multilinguisme et variétés linguistiques en Europe à l'aune de l'intelligence artificielle ». In: Rachele Raus (cur.) et al.. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l'aune de l'intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell'era dell'intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing. <https://www.collane.unito.it/oa/items/show/132>.

Wilmot Natalie Victoria (2022). *Language Management: From Bricolage to Strategy in British Companies*. Bristol: Multilingual Matters.

Yvon François (2022). "Evaluer, diagnostiquer et analyser la traduction automatique neuronale". In: Maggi Ludovica, Bordes Sarah (a cura di). *Intelligences pour la traduction. IA et interculturel : actions et interactions, Revue internationale d'interprétation et de traduction / International Journal of Interpretation and Translation FORUM [20:2] Special Issue*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 315-332. DOI: <https://doi.org/10.1075/forum.20.2>

Zollo Silvia Domenica (2022). « De l'usage raisonné des ressources documentaires numériques pour le développement de la compétence instrumentale dans la didactique de la traduction spécialisée : retours d'expérience et perspectives en contexte LANSAD ». *Revue Traduction et Langues*, 21(1), 157-172.

Zollo Silvia Domenica, Calvi Silvia (2022). "Fraseologia, traduzione e digital literacy nel contesto universitario: riflessioni e proposte per un percorso didattico sperimentale". In: R. Raus et al. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l'aune de l'intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell'era dell'intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 263-284. <https://www.collane.unito.it/oa/items/show/132>

Sitografia

Commissione Europea. *European Universities Initiative*. <https://education.ec.europa.eu/education-levels/higher-education/european-universities-initiative>

Commissione Europea. *Piano per la cooperazione settoriale sulle competenze*, <https://ec.europa.eu/social/main.jsp?langId=it&catId=1415>

La Traduzione Automatica (TA) neurale può essere molto utile per favorire la comunicazione tra parlanti di lingue diverse. Infatti, è sempre più integrata in strumenti d'uso comune, dai sistemi di messaggistica istantanea multilingue o traduzione vocale a quelli di reperimento dell'informazione sulle piattaforme commerciali online. La TA neurale serve anche come supporto al multilinguismo delle istituzioni e dei cittadini dell'Unione Europea. Sebbene il contributo di tale tecnologia sia innegabile, occorre osservarne gli sviluppi con attenzione, investendo in ricerca e indicando politiche che sappiano orientarne il progresso e l'uso al rispetto di principi condivisi. La TA neurale deve diventare una risorsa e non un elemento che sottrae terreno ai diritti e ai valori dell'uguaglianza e dell'inclusione. Il motto europeo "Unita nella diversità" deve restare un caposaldo e l'Europa deve continuare a promuovere iniziative a favore della diversità linguistica e dell'apprendimento di almeno due lingue straniere, in linea con l'"Obiettivo di Barcellona" (2002). È inoltre fondamentale sensibilizzare l'opinione pubblica sui limiti della TA e sui benefici dell'apprendimento di una lingua straniera.

In quest'ottica le politiche a favore dell'apprendimento linguistico non possono ignorare il ruolo dirompente della TA neurale e dell'intelligenza artificiale (IA, specialmente i modelli generativi), che devono essere via via integrate nei percorsi di formazione in modo da sfruttarne le potenzialità ma anche evidenziarne i limiti e educare le nuove generazioni a un utilizzo critico e consapevole. La ricerca nel campo della cosiddetta 'alfabetizzazione alla TA' sta emergendo. Un progetto degno di nota, e l'unico finora finanziato dal programma Erasmus+ (Azione chiave 2: Cooperazione tra organizzazioni e istituzioni), è *MultiTrainMT - Machine Translation training for multilingual citizens*¹ coordinato dalla Universitat Autònoma de Barcelona. I risultati del progetto sono:

- 1) la creazione di un *syllabus* per un corso sulla TA per cittadini multilingui che include anche un modulo sull'apprendimento delle lingue straniere;
- 2) l'apertura di una piattaforma online, MutNMT, per conoscere la TA neurale dall'interno e allenare una piattaforma creata ad hoc con corpora caricati dagli utenti;
- 3) la pubblicazione di un volume divulgativo sulla TA, con relative attività di verifica della comprensione e acquisizione dei contenuti online.

¹ 2019-1-ES01-KA203-064245, 1° settembre 2019-31 agosto 2022, <https://erasmus-plus.ec.europa.eu/projects/search/details/2019-1-ES01-KA203-064245?etrans=it>

Intelligenza artificiale, traduzione automatica e apprendimento linguistico

Alessandra Molino,
Lucia Cinato
Università di Torino

Sono necessari altri progetti di questo tipo che si rivolgano a diverse tipologie di utenti e che sappiano differenziare gli scenari d'uso dell'IA e della TA, al fine di sviluppare politiche pertinenti, realistiche e mirate.

La pressoché mancanza di dati e informazioni su come la TA possa influenzare i processi e i risultati dell'apprendimento linguistico è fonte di preoccupazione tra insegnanti, glottodidatti e linguisti. Gli studi esistenti (si vedano, ad esempio, Somers *et al.* 2006; Thue Vold 2018; Fredholm 2019; Carré 2022) suggeriscono che il livello di competenza della lingua straniera e il possesso di conoscenze sul funzionamento della TA sono fattori importanti per il riconoscimento degli errori e delle inadeguatezze dei risultati prodotti dai sistemi. Tale consapevolezza dovrebbe limitare il rischio di un uso scorretto della lingua e di interiorizzare forme lessicali e sintattiche standardizzate. Tuttavia, sono ancora molti gli interrogativi. Come usano gli strumenti di TA gli apprendenti stranieri a diversi livelli di competenza, grado scolastico, fascia d'età? Al di là di un uso sporadico individuale, come si possono efficacemente integrare la TA e l'IA generativa nella didattica delle lingue? L'impiego delle piattaforme online di TA ha un effetto positivo sulla produzione scritta e orale degli apprendenti? Tale effetto è temporaneo o permanente? Il progetto *MTrill: Machine Translation Impact on Language Learning*² e coordinato dalla Dublin City University ha fornito indicazioni iniziali sugli effetti della TA. Attraverso un esperimento di *priming* sintattico per analizzare se, nella loro produzione spontanea in inglese come seconda lingua, i partecipanti adottavano le strutture sintattiche da loro precedentemente osservate in un lavoro svolto con l'ausilio della TA, il progetto MTrill ha evidenziato che in effetti la TA può lasciare delle tracce nei processi di apprendimento: «i partecipanti si affidavano all'*output* di GT [Google Translate] tanto da modificare il loro comportamento linguistico per rispecchiare le scelte del sistema»³ (Resende, Way 2012: 82).

La dimensione etica è un ulteriore aspetto che va affrontato in merito alla didattica delle lingue e alla formazione di cittadini consapevoli. L'insegnamento delle lingue straniere implica non solo l'apprendimento del codice linguistico, ma anche della cultura legata alle lingue apprese. Gli apprendenti devono essere sensibilizzati alle differenze culturali e alle varietà sociolinguistiche.

In conclusione, la formazione linguistica, la TA e l'IA, la consapevolezza, la didattica, l'etica, e gli strumenti digitali in genere —

² Finanziato da EXCELLENT SCIENCE — Marie Skłodowska-Curie Actions, 25 aprile 2019-16 luglio 2021, <https://cordis.europa.eu/project/id/8434550>

³ «[P]articipants trusted the GT output enough to change their linguistic behaviour in order to mirror the system's choices». La traduzione è la nostra.

ad esempio, corpora e dizionari elettronici — sono tutti elementi interconnessi e indispensabili nella formazione linguistica moderna. Tuttavia, come sottolineato, la loro integrazione necessita di poggiare su basi di dati e informazioni più solide e complete, con maggiori sperimentazioni e collaborazioni interdisciplinari e internazionali.

Bibliografia

Carré Alice, Kenny Dorothy, Rossi Caroline, Sánchez-Gijón Pilar and Torres-Hostench Olga (2022). “Machine translation for language learners”. In: Dorothy Kenny (a cura di) *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Berlino: Language Science Press, 187-207. DOI:10.5281/zenodo.6760024

Fredholm Kent (2019). “Effects of Google translate on lexical diversity: Vocabulary development among learners of Spanish as a foreign language”. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 13(26). DOI:10.26378/rnlael1326300

Resende Natalie, Way Andy (2021). “Can Google Translate rewire your L2 English processing?”. *Digital*, 1, 66-85. DOI: <https://doi.org/10.3390/digital1010006>

Somers Harold, Gaspari Federico, Niño Ana (2006). “Detecting inappropriate use of free online machine translation by language students — a special case of plagiarism detection”. In: *Proceedings of the 11th Annual Conference of the European Association for Machine Translation*. Oslo: European Association for Machine Translation, 41-48.

Thue Vold Eva (2018). “Using machine-translated texts to generate L3 learners’ metalinguistic talk”. In: Åsta Haukås, Camilla Bjørke, Magne Dypedahl (a cura di). *Metacognition in language learning and teaching*. New York: Routledge, 67-97. DOI: <https://doi.org/10.4324/9781351049146>

L'intelligenza artificiale e lo sviluppo delle nuove tecnologie hanno rivoluzionato inevitabilmente il ruolo e i compiti delle figure professionali operanti nel settore delle lingue (Commission de l'éthique en science et en technologie 2019), tra cui quelle del traduttore e del revisore specialistico, chiamati non solo a conoscere le lingue di lavoro, ma anche a padroneggiare l'utilizzo di strumenti tecnologici: dalla gestione del processo traduttivo a quella delle diverse testualità e terminologie. Questa realtà si traduce in un nuovo bisogno di formazione che spazia da percorsi universitari per studenti aspiranti traduttori e revisori a corsi di aggiornamento rivolti a figure professionali già inserite nel settore (Frérot, Karagouch 2016; Gambier 2009): una formazione in cui conoscenze teoriche incontrano competenze pratiche che permettono di simulare il lavoro del traduttore e revisore di oggi, con l'obiettivo di formare professionisti competenti e competitivi sul mercato del lavoro.

In risposta a queste esigenze, un gruppo di esperti dell'EMT (*European Master's in Translation*) ha definito e aggiornato le competenze richieste per la formazione in questo settore (*EMT Competence Framework 2022*), indentificandone cinque principali: competenze linguistiche e culturali, competenze traduttive (dal punto di vista strategico, metodologico e della conoscenza del dominio di specialità), competenze tecnologiche, competenze personali e interpersonali e, infine, competenze di gestione del servizio traduttivo. Un'analisi più approfondita di questo documento ha messo in evidenza l'urgenza di una formazione grazie alla quale i traduttori e revisori possano trarre vantaggio dalle opportunità che l'intelligenza artificiale e l'informatica offrono, coscienti tuttavia dei limiti rispettivi (Flöter-Durr 2022) che richiedono l'intervento umano consapevole e ben formato.

All'interno del progetto di ricerca "Diritti e variazioni linguistiche in Europa nell'era dell'I.A.", l'Osservatorio di Terminologie e Politiche Linguistiche (OTPL) dell'Università Cattolica del Sacro Cuore di Milano ha svolto attività laboratoriali per studenti magistrali e dottorandi (aprile 2021 – maggio 2022), al fine di sperimentare nuovi percorsi di formazione che sensibilizzino gli apprendenti a un utilizzo consapevole degli strumenti informatici, con un'attenzione particolare alla fase di revisione di testi tradotti in modo automatico (Calvi, Dankova 2022). È stato così osservato l'utilizzo della traduzione automatica nella comunicazione multilingue nelle riviste di specialità, quali *Nature* e *National Geographic*, valutandone la qualità e il conseguente grado di revisione umana (Guasco 2013), con i seguenti risultati: il traduttore automatico è meno performante per quanto riguarda le scelte

Intelligenza artificiale e formazione in traduzione: nuove competenze per traduttori e revisori specialistici

Maria Teresa Zanola

Osservatorio di Terminologie e Politiche Linguistiche (OTPL)

Università Cattolica del Sacro Cuore, Milano

REALITER – Rete panlatina di terminologia

**Anna Serpente,
Martina Ali**

Università Cattolica del Sacro Cuore, Milano

lessicali, terminologiche e l'utilizzo di preposizioni. Un'indagine *ad hoc* è stata condotta nell'ambito del dominio della moda sostenibile: in un ambito di grande attualità, si è rilevato che la traduzione automatica ha più difficoltà relativamente alla dimensione terminologica (Zanola 2018), osservazione confermata anche in occasione di un'ulteriore attività laboratoriale, incentrata sul rapporto tra traduzione automatica e terminologia nel dominio del cambiamento climatico e dell'ambiente.

L'esperienza condotta ha confermato la necessità di una formazione innovativa e attenta alla competenza tecnologica o informatica, senza trascurare le competenze linguistiche e culturale che ancora oggi continuano a costituire un elemento centrale dei percorsi universitari. Inoltre, nell'ambito della traduzione specialistica, l'analisi delle traduzioni automatiche ha messo in evidenza come un buon processo di revisione necessiti di una conoscenza teorica approfondita in ambito terminologico. Lo studio del dominio in diverse fonti e tipologie testuali permette di sviluppare una sensibilità maggiore in fase di revisione, soprattutto per quanto riguarda la variazione terminologica nelle sue diverse sfaccettature, diatopica, diacronica e diafasica, elementi spesso non presi in considerazione dalla traduzione automatica. In questa direzione, una formazione teorica in terminologia, unita a un saper fare pratico, costituirà un sicuro punto di forza in termini di professionalità e competitività. L'OTPL sperimenta altresì queste pratiche nella formazione condotta da più di dodici anni nell'ambito del Corso di Perfezionamento in "Terminologie specialistiche e servizi di traduzione".

Bibliografia

Calvi Silvia, Dankova Klara (2022). « Industrie de la langue et formation des traducteurs spécialisés ». *Revue Traduction et Langues*, 21(1), 190-204. <https://hdl.handle.net/10807/227695>

Commission de l'éthique en science et en technologie (2019). *Les effets de l'intelligence artificielle sur le monde du travail. Document de réflexion*. Gouvernement du Québec.

European Master's in Translation (EMT) (2022). *Competence Framework 2022*. <https://termcoord.eu/2022/11/updated-version-of-the-emt-competence-framework-now-available/>

Flöter-Durr Margarete. (2022). "Epistemological limits of current digital techniques of artificial intelligence in translation". *Lebende Sprachen*, 67(1), 4-44. DOI: <https://doi.org/10.1515/les-2022-0004>

Frérot Cécile, Karagouch Lionel (2016). *Outils d'aide à la traduction et formation de traducteurs : vers une adéquation des contenus pédagogiques avec la réalité technologique des traducteurs*. ILCEA , 27 | 2016, mis en ligne le 08 novembre 2016, consulté le 10 juillet 2023. URL : <http://journals.openedition.org/ilcea/3849> ; DOI : <https://doi.org/10.4000/ilcea.3849>

Groupe d'experts EMT (2009). *Compétences pour les traducteurs professionnels, experts en communication multilingue et multimedia*. Bruxelles: Commission européenne, Direction générale de la traduction.

Guasco Patrizia (2013). *La révision bilingue : principes et pratiques*. Milano: Educatt.

Zanola Maria Teresa (2018). *Che cos'è la terminologia*. Roma: Carocci.

RACCOMANDAZIONE 2

Occorre investire su nuovi profili professionali nell'industria linguistica.

In una società sempre più interconnessa, che sempre di più fonda la comunicazione quotidiana su sistemi di intelligenza artificiale (Vetere 2022), appare imprescindibile una riflessione sulle tematiche della diversità culturale e linguistica, sul rispetto dei parametri di inclusione e di uguaglianza nella rappresentatività delle diverse istanze di genere o gruppo sociale, linguistiche e culturali. Tale rispetto è ribadito dalle Conclusioni del Consiglio dell'Unione europea del 2022 (Consiglio dell'Unione europea 2022): «La diversità culturale e linguistica è componente intrinseca dell'Unione europea e dei suoi valori fondamentali. [...] Una politica ambiziosa in materia di diversità culturale e linguistica dovrebbe integrare pienamente le questioni di sostenibilità e tenere conto dell'innovazione tecnologica, anche nel settore digitale». Ora, l'apparente crescita esponenziale del multilinguismo nell'ambito delle applicazioni che l'intelligenza artificiale può supportare — la Risoluzione del Parlamento europeo del 3 maggio 2022 sull'intelligenza artificiale in un'era digitale (2020/2266(INI)) auspica l'attuazione e lo sviluppo della tecnologia dell'IA anche nel settore delle lingue minoritarie in modo tale da incoraggiarne la conoscenza e l'uso — non si accompagna necessariamente a una maggiore o più equa presa in carico delle diversità (Larsonneur 2021). Questo divario tra inclusività esposta e omogeneizzazione reale in relazione a modelli tarati sostanzialmente su realtà linguistiche dominanti può arrivare a produrre discriminazione più che una reale condivisione di contenuti, riproducendo dinamiche di dominazione nei confronti di gruppi sociali o culturali (Markl 2022) e precludendo a queste comunità una reale presenza e un vero *empowerment* all'interno della comunicazione globale.

Su queste basi, nuove competenze sono richieste al personale specializzato nell'ambito della comunicazione: è urgente la formazione di persone esperte che possano comprendere e gestire la complessità e la diversità dell'espressione linguistica salvaguardandone la varietà. Appare soprattutto urgente che le nuove figure professionali (nell'ambito del *prompt engineering*, del *post-editing*, della programmazione) padroneggino appieno le dimensioni diatopiche, diamesiche, diastratiche e diafasiche nell'ambito dei diversi sistemi linguistici e siano in grado di allenare le applicazioni basate su intelligenza artificiale al riconoscimento, alla definizione e all'identificazioni di tali dimensioni all'interno dei grandi corpora testuali. Da qui la necessità correlata di saper costruire corpora di grandi dimensioni, tarati *ad hoc* per l'allenamento delle macchine, in grado di salvaguardare *a monte* la variazione linguistica. Tali competenze permetteranno

Investire su nuovi profili di mestiere

Danio Maldussi

Università di Bergamo

Micaela Rossi

Università di Genova

la difesa della pluralità e della diversità, nel rispetto di una comunicazione multilingue efficace.

Si pensi, a titolo di esempio, all'universo delle piccole e media aziende dove si intensifica la domanda di sistemi di traduzione automatica "su misura" che sappiano rappresentare bisogni linguistico-terminologici specifici a livello settoriale e delle singole realtà aziendali. Tale necessità si palesa non solo a livello di aziende di produzione ma anche di servizi, come nel caso, ad esempio, delle imprese di servizi linguistici. A livello di singola azienda esistono bisogni ancora più specifici, bene rappresentati dal *jargon d'entreprise* (gergo aziendale) o *variation occupationnelle* (Bertaccini, Matteucci 2005) a garanzia di integrazione e senso di appartenenza. Il rispetto della pluralità e della diversità nella comunicazione diventa pertanto un asset strategico per tutte le realtà aziendali che operano in scenari competitivi.

Tuttavia, la gestione della complessità non implica solo l'individuazione di nuove figure professione ma deve altresì sfociare nella costruzione di prodotti che sappiano rispondere alle esigenze di ottimizzazione dei tradizionali "prodotti" linguistico-terminologici quali glossari terminologici e manuali d'uso. Ai professionisti sarà richiesto di sostituirli con banche dati terminologiche "scalabili" nelle lingue di lavoro e nei settori di appartenenza, nonché di progettare sistemi di traduzione automatica adattivi.

Bibliografia

Bertaccini Franco, Matteucci Alessandra. (2005). « L'approche variationniste à la pratique terminologique d'entreprise ». *Meta: Translators' Journal*, 50(4). <https://doi.org/10.7202/019910ar>

Consiglio dell'Unione europea (2022). *Conclusioni del Consiglio su rafforzare gli scambi culturali attraverso la mobilità degli artisti e dei professionisti creativi e della cultura nonché attraverso il multilinguismo nell'era digitale (2022/C 160/07)*. [https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:52022XG0413\(02\)](https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:52022XG0413(02))

Larsonneur Claire (2021). « Intelligence artificielle ET/OU diversité linguistique : les paradoxes du traitement automatique des langues ». *Hybrid* [en ligne], 7. DOI: <https://doi.org/10.4000/hybrid.650>

Markl Nina (2022). "Mind the data gap(s): Investigating power in speech and language datasets". In: Bharathi Raja Chakravarthi, B Bharathi, John P McCrae, Manel Zarrouk, Kalika Bali, Paul Buitelaar (Eds.). *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Dublin: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.ltedi-1>, 11-12.

Parlamento europeo (2022). *Risoluzione del Parlamento europeo del 3 maggio 2022 sull'intelligenza artificiale in un'era digitale* (2020/2266(INI)). Bruxelles: Parlamento Europeo. https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_IT.html

Vetere Guido (2023). "Elaborazione automatica dei linguaggi diversi dall'inglese: introduzione, stato dell'arte e prospettive". In: Rachele Raus et al. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l'aune de l'intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell'era dell'intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 69-87. <https://www.collane.unito.it/oa/items/show/132>

I cambiamenti nel mercato del lavoro legati allo sviluppo senza precedenti delle tecnologie, e in particolare l'intelligenza artificiale, non sono una novità. Nell'ambito delle professioni legate alle lingue, in particolare, la traduzione automatica è solo uno dei fenomeni che hanno modificato non solo le modalità di lavoro, ma anche i bisogni formativi, rendendo la presenza nei corsi universitari di tecnologie e di modelli di traduzione automatica una necessità. Con lo sviluppo sempre più rapido dell'intelligenza artificiale sta diventando necessario non solo conoscere risorse quali sistemi di traduzione automatica o modelli linguistici come ChatGPT, ma anche avere un'idea di come funzionano, su che cosa si basano e che cosa possono e non possono fare. Questo porta alla creazione di nuove professioni a cavallo tra lingue e ingegneria/programmazione, ma richiede anche un'appropriazione da parte di stakeholder quali le università. Per collegare meglio i bisogni del mercato del lavoro e quelli dell'istruzione superiore, il progetto Erasmus+ *UPgrading the SKILLS of Linguistics and Language Students - UPSKILLS*¹ ha individuato un nuovo, composito profilo professionale denominato *specialista in dati e progetti linguistici (language data and project specialist; Miličević Petrović et al. 2021)*.

L'identificazione di un nuovo profilo professionale nei campi legati alla lingua e alla linguistica nasce da una dettagliata analisi dei bisogni, che ha chiaramente evidenziato la necessità di un nuovo gruppo di competenze, e soprattutto di una nuova mentalità, per poter soddisfare le esigenze e le sfide professionali del settore e del corrispondente mercato del lavoro. La prima componente è stata un'indagine dei curricula dei corsi di laurea dedicati alle lingue (linguistica, lingue moderne, mediazione linguistica) in diverse università europee. L'indagine ha evidenziato la necessità di fornire contenuti formativi aggiuntivi in linea con le esigenze del mercato di lavoro, che consentissero a studenti e docenti di collegare e valorizzare competenze spesso già presenti nei curricula. Dalla seconda componente, l'analisi della letteratura accademica e dei report di settore, sono emersi come rilevanti sei gruppi/blocchi di competenze, con un'enfasi particolare sulle capacità trasversali legate alla ricerca. La terza componente, empirica, si è basata sull'analisi di un corpus di annunci di lavoro, il quale è stato esaminato per rilevare e classificare parole e frasi ricorrenti, portando all'identificazione delle abilità e delle competenze tipicamente richieste nell'industria delle lingue, nonché delle denominazioni utilizzate per queste nuove professioni ibride fra lingue e tecnologie. Sulla base di questo prelimi-

Nuovi profili professionali a cavallo tra lingue e tecnologie. Risultati dell'analisi dei bisogni condotta nel progetto Erasmus+ UPSKILLS

Silvia Bernardini,
Adriano Ferraresi,
Maja Miličević Petrović
Università di Bologna

¹ <https://upskillsproject.eu>

nare lavoro di ricerca è stato poi elaborato un questionario sottoposto alle aziende che assumono linguiste/i e professioniste/i delle lingue. Le abilità e le competenze menzionate nelle risposte al questionario hanno infine portato allo sviluppo di interviste semi-strutturate che sono state poi condotte con i rappresentanti del settore. Da tutte le componenti è emerso come i programmi di studio di ambito linguistico devono porre un maggiore accento sullo sviluppo di competenze tecnologiche, gestionali e trasversali, in particolare quelle legate alla ricerca.

Il profilo dello/la *specialista in dati e progetti linguistici* che è emerso dall'analisi dei bisogni così condotta non è inteso come legato a una singola posizione lavorativa, né come il titolo di un nuovo corso di studio. Si tratta di un profilo appositamente generico, modulare e adattabile alle esigenze di diversi corsi di studio e diverse posizioni nel mercato del lavoro. Allo scopo di mantenere questa fondamentale flessibilità, le implicazioni formative sono state strutturate intorno a due dimensioni: una dimensione verticale focalizzata sui sette ambiti principali individuati nell'analisi dei bisogni del progetto UPSKILLS (ambito disciplinare, (inter)culturale, tecnologico, legato ai dati, legato alla ricerca, organizzativo e trasversale) e una orizzontale basata sugli elementi standard dei risultati dell'apprendimento (conoscenze, abilità e competenze). L'idea centrale del profilo è di costituire una guida nella scelta delle conoscenze, abilità e competenze da conseguire/insegnare, tenendo presente che la maggior parte delle nuove posizioni lavorative nel settore richiedono un contributo maggiore o minore da ciascuno dei sette ambiti. I principali compiti e responsabilità da un lato, e le principali conoscenze, abilità e competenze dall'altro, possono essere riassunte come nelle figure 1 - 2.



Figura 1. Compiti e competenze dello/la specialista in dati e progetti linguistici (fonte: https://upskillsproject.eu/wp-content/uploads/2021/09/sess1.pres7_.Profile.pdf)²

² La traduzione dall'inglese all'italiano del testo presente nelle tre figure a cura dell'autore.



Figura 2. Compiti e competenze dello/a specialista in dati e progetti linguistici
(fonte: https://upskillsproject.eu/wp-content/uploads/2021/09/sess1.pres7_.Profile.pdf)

Sono inoltre stati definiti quattro sottoprofili più specifici, a seconda che il focus sia più sulla ricerca (*analista di dati linguistici* e *scienziato/a dei dati linguistici*) oppure sulla gestione di progetti (*manager di dati linguistici* e *manager di progetti linguistici*), e a seconda del livello di responsabilità e dell'esperienza lavorativa (minori per l'analista e il/la manager di dati linguistici, maggiori per lo/la scienziato/a dei dati linguistici e il/la manager di progetti linguistici). La figura 3 presenta a titolo esemplificativo alcuni compiti e responsabilità legati ai sottoprofili. Le relative competenze sono sottoinsiemi a geometria variabile di quelle proprie del profilo più generale. Maggiori informazioni su profili e sottoprofili, così come sulla metodologia e i risultati dell'analisi dei bisogni sono reperibili in Miličević Petrović *et al.* (2021).



Figura 3. Responsabilità e compiti legati ai quattro sottoprofili
(fonte: https://upskillsproject.eu/wp-content/uploads/2021/09/sess1.pres7_.Profile.pdf)

Oltre a riassumere il composito insieme di dati e informazioni accademici e professionali emersi dall'analisi dei bisogni condotta nel progetto UPSKILLS, riteniamo che il profilo e i sottoprofili possano contribuire a mettere ordine in un settore complesso e disomogeneo caratterizzato dalla multidisciplinarietà, guidando i/le docenti nella creazione di nuovi corsi e nuovi materiali didattici e aiutando le studentesse e gli studenti a valoriz-

zare i propri interessi e le conoscenze già acquisite in relazione a prospettive professionali in rapida e costante crescita.

Bibliografia

Miličević Petrović Maja, Bernardini Silvia, Ferraresi Adriano, Aragrande Gaia, Barrón-Cedeño Alberto (2021). *Language data and project specialist: A new modular profile for graduates in language-related disciplines*. UPSKILLS Intellectual output 1.6. Zenodo. <https://dx.doi.org/10.5281/zenodo.5030929>

La digitalizzazione che oggi coinvolge la società su scala globale amplifica un bisogno sempre più crescente, quello di poter disporre di dati linguistici affidabili per la specializzazione di tecnologie linguistiche rispetto al dominio applicativo. Gli algoritmi neurali necessitano dei neuroni umani molto di più di quanto questo non sia emerso finora nella ricerca. L'importanza di produrre e di disporre di dati affidabili è una preoccupazione, già da lunga data, dell'ingegneria e della linguistica computazionale che si occupa di costruzione e modellazione di tecnologie linguistiche (citiamo, ad esempio, Carletta 1996; Artstein, Poesio 2008; Bayerl, Paul 2011) con una propensione spiccata alla lingua inglese. Secondo i dati EUROSTAT (Vetere 2022: 79), in generale, l'uso dei media digitali per imprese e persone private è maggiormente diffuso in lingua inglese; si evince, pertanto, che migliori sono le tecnologie, maggiore sarà la diffusione e la capacità attrattiva di una lingua per la quale si dispiegano altrettanti investimenti tecnologici.

Al fine di preservare la diversità linguistica, il multilinguismo e il plurilinguismo¹, l'addestramento delle reti neurali in altre lingue si rivela capitale. Le proposte di regolamento del Parlamento europeo del 21 aprile 2021 sull'intelligenza artificiale (IA) e sui servizi orientati ai dati, mostrano una crescente consapevolezza delle istituzioni europee rispetto a tematiche dello scacchiere geopolitico che il progresso tecnologico ha introdotto. Le tecnologie del linguaggio naturale sono al centro di una delle tematiche più attuali nonché uno dei principali settori dell'IA.

L'IA è la capacità di un computer di imitare funzioni cognitive degli esseri umani quali l'apprendimento e la risoluzione dei problemi. Un computer utilizza strumenti matematici e logici per simulare le modalità con cui gli esseri umani ragionano al fine di apprendere nuove informazioni e prendere decisioni. Il *Machine Learning* (ML) è un'applicazione dell'IA, il cui processo consiste nell'utilizzare modelli matematici per aiutare un computer ad apprendere da solo sulla base dell'esperienza. Tale esperienza è resa possibile in larga parte grazie all'annotazione dei dati, ovvero un processo col quale dei dati vengono etichettati tramite l'aggiunta di metadati. La finalità perseguita è quella di mostrare alla macchina il risultato ideale, un risultato che il modello dovrà poi essere in grado di replicare su dati non analizzati in precedenza. L'annotazione si combina quasi esclusivamente con i corpora, e serve per facilitare l'estrazione di informazioni di natura

¹ Ricordiamo che per "multilinguismo" si intende la presenza di più lingue all'interno di uno stesso territorio o, nel nostro caso, in un'organizzazione con il relativo lavoro di traduzione e "plurilinguismo", ovvero la conoscenza di più lingue da parte degli individui (cfr. Gaboriaux, Raus, Robert, Vicari 2022: 9).

La figura professionale dell'annotatore per garantire una maggiore affidabilità dei dati provenienti dall'Intelligenza Artificiale

Michela Tonti
Università di Bergamo

linguistica o discorsiva. Pertanto, un corpus annotato contiene gli elementi che si desidera che il sistema di ML impari a riconoscere e ad acquisire, diventando così fruibile anche per progetti futuri. Il lavoro di annotazione su cui ci focalizziamo ha luogo prima che i corpora vengano resi disponibili al pubblico. Di conseguenza, a occuparsene, sono in primo luogo gli specialisti di linguistica computazionale di cui forniamo alcuni riferimenti bibliografici. I linguisti forniscono elenchi provvisori di categorie linguistiche rilevanti, come le componenti morfosintattiche e le condizioni contestuali per disambiguarle. Il resto del lavoro, in cui rientra una complessa opera di annotazione, è frutto dell'intervento della figura dell'annotatore.

Questa figura, relegata finora a intervenire solo nei momenti di preparazione dei corpora monolingui o paralleli, ha compiti che di riflesso sono definiti volta per volta a seconda delle categorie da annotare: fonetiche, morfologiche, sintattiche, semantiche, pragmatiche, stilistiche, discorsive. Sulla scorta della selezione di pubblicazioni operata dai linguisti computazionali con un focus sull'annotazione², si evince che l'affidabilità del risultato finale dipende in larga parte dalla correttezza del processo di annotazione e di *pre-editing*, ruoli che necessitano di essere sempre più valorizzati. Se l'obiettivo che ci prefiggiamo è quello di far luce sulla figura dell'annotatore, è imprescindibile non menzionare quella del *pre-editor*, visto che ci situiamo a monte dell'*output* concretamente ottenuto dall'utente finale. La figura del *pre-editor* è organicamente integrata nella produzione/traduzione di testi, volta per volta, e il testo così modificato o tradotto è destinato alla pubblicazione e quindi alla ricezione umana; nell'insieme ha competenze linguistiche diverse da quelle dell'annotatore. Il *modus operandi* di quest'ultima figura: quella dell'annotatore, è del tutto nuovo rispetto al ruolo tradizionale dell'analista del discorso che analizza a posteriori dei discorsi realmente prodotti, dato che si tratta di fare di volta in volta uno sforzo di predizione degli effetti che potrebbero intervenire nel caso in cui una data parola sia utilizzata in discorsi diversi tra loro, le ambiguità semantiche sono in effetti particolarmente insidiose e dispendiose in caso di *post-editing*. La mancanza di coerenza e di coesione dell'*output* o altre problematiche sul piano morfosintattico che necessitano pertanto di un intervento di *post-editing*, rappresentano un problema in termini di tempo e di costi che l'apprendimento supervisionato grazie al lavoro di annotazione permette di superare. Non è trascurabile il fatto che diverse parti del processo di annotazione possano essere automatizzate: ad esempio, il *soft-*

² Cfr. nella bibliografia.

ware Sketch Engine prevede una componente standard utilizzabile ogni volta che si carica un nuovo *corpus* la quale annota automaticamente, con una piccola percentuale di errore, tutte le categorie morfologiche e un certo numero di quelle sintattiche. Il lavoro deve invece essere realizzato manualmente quando si vogliono contrassegnare le unità rilevanti dal punto di vista sintattico, stilistico e pragmatico.

Infatti, i nuovi bisogni di annotazione manuale si riferiscono principalmente a elementi pragmatico-stilistici, o alle unità tematiche e informative importanti per un determinato compito o per un determinato dominio disciplinare. Ne deriva una notevole varietà di categorie non tradizionali da definire e da individuare, e conseguentemente la competenza richiesta agli annotatori cambia radicalmente da un progetto all'altro. Questo aspetto è di tutto valore, in quanto si evince la spiccata propensione alla duttilità, alla ricerca e alla formazione continua di cui deve dar prova chi fa annotazione.

In Figura 1, riportiamo un esempio di annotazione e di sviluppo di un modello. La figura dell'annotatore è rappresentata dai *task* elencati e rappresentati in blu mentre quelli in verde sono ricoperti da un *annotation lead* con il compito di coordinare il lavoro di annotazione precedentemente realizzato da annotatori che, in virtù dell'elevato numero di testi necessari per addestrare il sistema, saranno altrettanto numerosi. È quindi essenziale mirare ad armonizzare le annotazioni per assicurare l'affidabilità dei dati dell'*output* di cui parlavamo all'inizio.

L'apprendimento supervisionato della macchina e la figura dell'annotatore sono quindi, a nostro avviso, imprescindibili per il mercato dell'IA.



Figura 1. Fasi del processo di annotazione e sviluppo: un esempio a nostra cura.

Elenco di alcuni esempi di contributi realizzati con l'aiuto di annotazioni manuali

ELMo: Deep contextualized word representations di Matthew E. Peters *et al.* Questo lavoro presenta ELMo, un LLM (*Large Language Models*) pionieristico che genera rappresentazioni di parole contestualizzate sfruttando modelli linguistici profondi bidirezionali.

Probing Neural Network Comprehension of Natural Language Arguments di Timothy Niven e Hung-Yu Kao. Questo studio ha utilizzato gli LLM per verificare la loro comprensione degli argomenti in linguaggio naturale, valutandone capacità e limiti.

CoQA: A Conversational Question Answering Challenge di Siva Reddy *et al.* Questo progetto ha presentato il *dataset* CoQA, incentrato sulla risposta a domande poste in contesto conversazionale, progetto nel quale si richiede ai LLM di comprendere e generare risposte coerenti.

SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles di Giovanni Da San Martino *et al.* Questo contributo ha lanciato la sfida di rilevare le tecniche di propaganda negli articoli di cronaca, incoraggiando l'uso degli LLM e dell'annotazione manuale per migliorare l'accuratezza del rilevamento.

Bibliografia

Artstein Ron, Poesio Massimo (2008). "Inter-coder agreement for computational linguistics". *Computational Linguistics*, 34(4), 555–596.

Artstein Ron, Poesio Massimo (2005). "Bias decreases in proportion to the number of annotators". *Proceedings of FG-MoL 2005*, 141–150.

Bayerl Petra Saskia, Karsten Paul Ingmar (2011). "What Determines Inter-Coder Agreement in Manual Annotations? A Meta-Analytic Investigation". *Computational Linguistics*, 37 (4), 699–725, DOI: https://doi.org/10.1162/COLI_a_00074

Carletta Jean (1996). "Assessing agreement on classification tasks: The kappa statistic". *Computational Linguistics*, 22(2), 249–254.

Commissione europea (2021). *Proposta di Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (Legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione*. Bruxelles: Commissione europea, <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=celex%3A52021PC0206>

Commissione europea (2020). *Legge sui servizi digitali*. Bruxelles: Commissione europea, <https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52020PC0825>

Da San Martino Giovanni, Barrón-Cedeño Alberto, Wachsmuth Henning, Petrov Rotislav, Nakov Preslaw. (2020). “SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles”. In: Herbelot A. et al. *Proceedings of the Fourteenth Workshop on Semantic Evaluation*. Barcellona: SemEval, 1377-1414.

Gaboriaux Chloé, Raus Rachele, Robert Cécile, Vicari Stefano (a cura di) (2022). « Le multilinguisme dans les organisations internationales ». *Mots. Les langages du politique*, n°128. DOI: <https://doi.org/10.4000/mots.29135>

Niven Timothy, Kao Hung-Yu (2019). “Probing Neural Network Comprehension of Natural Language Arguments”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Firenze: Association for Computational Linguistics (ACL), 4658-4664. DOI: [10.18653/v1/P19-1459](https://doi.org/10.18653/v1/P19-1459)

Peters Matthew E., Neumann Mark, Iyyer Mohit, Gardner Matt, Clark Christopher, Lee Kenton, Zettlemoyer Luke (2018). “Deep Contextualized Word Representations”. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. New Orleans: Association for Computational Linguistics, 2227-2237. DOI: [10.18653/v1/N18-1202](https://doi.org/10.18653/v1/N18-1202)

Reddy Siva, Chen Danqi, Manning Christopher D. (2019). “CoQA: A Conversational Question Answering Challenge”. In: *Transactions of the Association for Computational Linguistics*, 7. Cambridge: MIT Press, 249-266.

Seretan Violeta, Roturier Johann., Silva David, Bouillon Pierre (2014). “The ACCEPT Portal: An Online Framework for the Pre-editing and Post-editing of User-Generated Content”. In: *Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation*. Gothenburg: Association for Computational Linguistics, 66-71.

Vetere Guido (2023). “Elaborazione automatica dei linguaggi diversi dall’inglese: introduzione, stato dell’arte e prospettive”. In: Raus R. et al. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l’aune de l’intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell’era dell’intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 69-87. <https://www.collane.unito.it/oa/items/show/132>.

Sitografia

Sketch Engine, <http://www.sketchengine.eu>

Bias di genere e intelligenza artificiale: una mancanza di competenze?

Mara Floris
Università Vita-Salute

Nel novembre del 2017, lo scrittore americano Alex Shams ha creato un certo scalpore su diversi *social media* che hanno ripreso un *tweet* nel quale scriveva: «Il turco è una lingua neutra dal punto di vista del genere. Non ci sono ‘lui’ o ‘lei’: tutto è semplicemente ‘o’. Ma guardate cosa succede quando Google traduce in inglese»¹. Il testo era seguito da uno *screenshot*, nel quale alcune brevi frasi contenenti il pronome turco neutro “o” sono tradotte dal turco all’inglese, ottenendo il risultato in figura 1:

I pronomi personali vengono tradotti al maschile o al femminile a seconda dello stereotipo di genere relativo alla professione indicata: “ingegneri” e “dotto-ri” sono uomini, “infermiere” e “segretarie” donne. Lo stesso succede quando ai pronomi personali vengono accostati sentimenti e atteggiamenti: le traduzioni rinforzano lo stereotipo di donna fragile ed emotiva (Prates *et al.* 2020).

In letteratura, la presenza di un *bias* di genere nelle traduzioni automatiche e in altri sistemi di *Natural Language Processing* (NLP) è ampiamente attestata (Chen *et al.* 2021; Costa-jussà 2019; Sun *et al.* 2019) Il problema è duplice: manca una formazione specifica sui *bias* di genere e mancano anche competenze linguistiche specifiche sulle lingue diverse dall’inglese.

Dopo le numerose segnalazioni, Google ha provveduto a correggere il tiro. Attualmente, se provate a tradurre dal turco “o bir doktor”, otterrete “lei è un dottore”. Molto meglio, anche se il termine corretto in italiano è “dottoressa”, un errore che avrebbe potuto essere evitato facilmente da una persona con le competenze adeguate dal punto di vista linguistico e con un’appropriata formazione sull’esistenza di un *bias* di genere.

Turkish - detected	English
o bir aşçı	she is a cook
o bir mühendis	he is an engineer
o bir doktor	he is a doctor
o bir hemşire	she is a nurse
o bir temizlikçi	he is a cleaner
o bir polis	He-she is a police
o bir asker	he is a soldier
o bir öğretmen	She's a teacher
o bir sekreter	he is a secretary
o bir arkadaş	he is a friend
o bir sevgili	she is a lover
onu sevmiyor	she does not like her
onu seviyor	she loves him
onu görüyor	she sees it
onu göremiyor	he can not see him
o onu kucaklıyor	she is embracing her
o onu kucaklamıyor	he does not embrace it
o evli	she is married
o bekar	he is single
o mutlu	he's happy
o mutsuz	she is unhappy
o çalışkan	he is hard working
o tembel	she is lazy

Figura 1. Screenshot del tweet di Alex Shams del 28 novembre 2017 (fonte: https://twitter.com/alexshams_/status/935291317252493312)

¹ «Turkish is a gender-neutral language. There is no ‘he’ or ‘she’ — everything is just ‘o’. But look what happens when Google translates to English» (traduzione mia).

Un'area in cui i *bias* di genere sono particolarmente diffusi è quella dell'elaborazione del linguaggio naturale (NLP) e delle traduzioni (cfr. Luccili *et al.* 2020). Per *bias* di genere si intende il favoritismo, o la discriminazione sistematica, nei confronti di un particolare genere, con conseguente disparità di trattamento e di opportunità. Sebbene nelle dichiarazioni degli intenti dei produttori l'intelligenza artificiale (IA) miri a essere obiettiva e imparziale, spesso rispecchia i *bias* presenti nei dati su cui viene addestrata. I dati utilizzati per addestrare i modelli di IA possono contenere preconcetti, ereditati dalla società in cui è stata progettata e che portano a risultati distorti.

Oltre al sopra menzionato caso della traduzione, un altro caso in cui è presente un *bias* di genere sono i sistemi di *sentiment analysis*. L'analisi dei sentimenti di un testo effettuata da un sistema di NLP potrebbe associare erroneamente parole o espressioni tipicamente utilizzate da donne o uomini con sentimenti positivi o negativi, creando un'interpretazione distorta e rinforzano stereotipi di genere.

Altri esempi della presenza di tale problema nel NLP sono i sistemi di *screening* dei *curricula*. Se un sistema di selezione automatizzata dei *curricula* addestrato su dati storici mostra una preferenza per determinati termini o esperienze tipicamente associate a un genere specifico, tale preferenza potrebbe scoraggiare le candidate femminili o maschili a seconda del *bias* incorporato.

Per affrontare in modo completo tutte le possibili conseguenze di tali discriminazioni nel mondo delle tecnologie di NLP, è necessario compiere diversi passi importanti, che includono lo sviluppo di nuove professioni, l'acquisizione di competenze specialistiche da parte dei linguisti e uno sforzo congiunto per interrompere il processo attraverso il quale i *bias* di genere si riflettono nelle tecnologie che produciamo. In tal senso, un aspetto cruciale riguarda la creazione di nuovi profili professionali con una formazione linguistica, in particolare sulle lingue cui è presente il genere grammaticale, e con un'adeguata preparazione sui *bias* di genere e sulle manifestazioni linguistiche di questo fenomeno. Tradizionalmente, lo sviluppo dell'NLP ha fatto affidamento principalmente sulle competenze di informatici e ingegneri, che potrebbero non possedere una comprensione approfondita delle sfumature e della mancanza d'inclusione linguistica. Incorporando specialisti del linguaggio con una formazione adeguata in lingue diverse dall'inglese, possiamo garantire una prospettiva più completa e culturalmente diversificata nel processo di sviluppo. Tali figure possono contribuire con la loro esperienza all'analisi delle strutture linguistiche, con un occhio di riguardo per il problema del genere, all'identificazione di discriminazioni

potenziali e alla proposta di strategie per attenuarle o eliminarle.

Inoltre, i linguisti con una formazione specifica in sociolinguistica, analisi del discorso e studi di genere, permetterebbero una comprensione maggiore di come i *bias* di genere si manifestano nel linguaggio. Sfruttando la propria esperienza, possono contribuire attivamente allo sviluppo di algoritmi e modelli più sensibili alle diverse espressioni e identità di genere.

Bloccare il processo attraverso il quale i *bias* di genere si riflettono nelle tecnologie NLP è un obiettivo fondamentale. Ciò comporta la loro identificazione e correzione in varie fasi del processo di sviluppo. Stabilendo metodologie di valutazione rigorose, possiamo valutare sistematicamente i modelli di NLP alla ricerca delle discriminazioni di genere e perfezionarli iterativamente per garantire equità e inclusione. Inoltre, è essenziale creare insiemi di dati di formazione diversificati e rappresentativi. Ciò può essere ottenuto incorporando le prospettive delle comunità emarginate e consultando individui con diverse identità di genere durante i processi di raccolta e annotazione dei dati. Lavorando attivamente per ridurre i pregiudizi di genere nei sistemi NLP, possiamo promuovere tecnologie più eque e giuste che contribuiscono a una società più inclusiva.

Bibliografia

Chen Yan, Mahoney Christopher, Grasso Isabella, Wali Esmā, Matthews Abigail, Middleton Thomas, Nije Mariana, Matthews Jeanna (2021). "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 (AIES '21)*. New York: Association for Computing Machinery, 24–34. DOI: <https://doi.org/10.1145/3461702.3462530>

Costa-jussà Marta R. (2019). "An analysis of gender bias studies in natural language processing". *Nature Machine Intelligence*, 1(11), 495-496.

Luccioli Alessandra, Dolei Ester, Xausa Chiara (2020). "Investigating Gender Bias in Machine Translation. A Case Study between English and Italian". In: Adriano Ferraresi, Roberta Pederzoli, Sofia Cavalcanti, Randy Scansani (a cura di), *MediAzioni 29*: B29-B49. <http://www.mediazioni.sitlec.unibo.it>

Prates Marcelo O., Avelar Pedro H., Lamb Luís C. (2020). "Assessing gender bias in machine translation: a case study with google translate". *Neural Computing and Applications*, 32, 6363-6381.

Sun Tony, Gaut Andrew, Tang Shirlyn, Yuxin Huang, ElSherief Mai, Zhao Jieyu, Mirza Diba, Belding Elizabeth, Chang Kai-Wei, Yang Wang William (2019). "Mitigating gender bias in natural language processing: Literature review". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Firenze: Association for Computational Linguistics (ACL), 1630-1640. <https://aclanthology.org/P19-1159/>

RACCOMANDAZIONE 3

Occorre investire sulla creazione di corpora multilingui elaborati a partire da materiale autentico nazionale rispettoso della diatopia linguistica.

La diversità linguistica, una risorsa minacciata

La diversità linguistica, componente singolare della biodiversità (Le Coadic 2010: 53-56), è attualmente minacciata dal paradigma economico-tecnologico, oltre che da un'ideologia "neobabelica"¹. Secondo l'UNESCO, le lingue vive sono circa 6700, distribuite in modo estremamente sbilanciato:

- 1) in termini quantitativi assoluti (numerosità): la maggioranza delle lingue è mediamente praticata da un numero relativamente debole di individui, mentre poche lingue (cinese mandarino, inglese, spagnolo...) sono praticate dalla maggioranza della popolazione mondiale;
- 2) in termini quanti-qualitativi (status): alcune lingue di grande diffusione, generalmente (co)ufficiali o seconde, come l'arabo standard o il kiswahili, sono le lingue materne di un numero modesto di persone;
- 3) in termini di diffusione areale (geografia): zone di altissima concentrazione di diversità linguistica (subcontinente indiano e Himalaya, Sud-Est asiatico, America centrale e meridionale, Africa subsahariana... senza contare le grandi aree urbane) si alternano con aree relativamente più omogenee (Europa occidentale, America nordorientale, Nordafrica, Asia nordorientale...).

Non è facile definire cosa sia una lingua "minoritaria", e soprattutto è complicato stilare un elenco, in quanto il carattere "minoritario" dipende molto dal contesto². Più semplice è definire cosa sia una lingua "minacciata": l'*Atlas des langues en danger dans le monde*³ recensisce circa 2500 lingue a rischio di estinzione nei prossimi anni, ossia quasi il 40% delle lingue attualmente praticate nel mondo.

Con ogni evidenza, questa "minaccia" di estinzione non riguarda solo il piano strettamente linguistico, ma presenta gravi ric-

Gestione della diversità linguistica e intelligenza artificiale: rischi e opportunità

Giovanni Agresti

*Centre National de la
Recherche Scientifique — CNRS
Université Bordeaux Montaigne*

¹ La molteplicità delle lingue è tradizionalmente interpretata (e in buona misura lo è ancora) come una barriera alla comunicazione pratico-economica e come un ostacolo alla comunicazione *tout court* e, di conseguenza, alla pacifica convivenza tra popoli. Tuttavia, le ultime due guerre che hanno insanguinato l'Europa — il conflitto nei Balcani negli anni 1990 e l'attuale guerra in Ucraina — mostrano tragicamente come l'identità di lingua non sia affatto garanzia di pace, stabilità, dialogo tra paesi diversi. L'orizzonte di una lingua unica come strumento della costruzione democratica europea (De Mauro 2014), appare oggi più un fantasma di procedenza biblica o un'opzione al servizio dell'egemonia finanziaria globalizzante, che una ragionevole prospettiva di equità (Gazzola 2016) e di sviluppo sostenibile (in senso ambientale e sociale).

² L'italiano è minoritario nel Parlamento europeo ma maggioritario in Italia; il catalano è maggioritario in alcune città della Catalogna, minoritario in Spagna, ultra-minoritario in Sardegna; il greco di Calabria è sempre ultra minoritario... e così via.

³ <https://unesdoc.unesco.org/ark:/48223/pf0000189451>

dute in termini sociali, economici, ambientali – come, tra gli altri, ha evidenziato anche il Papa nella sua enciclica *Laudato si'* (Francesco 2015: 111-114). In effetti, le tre prospettive menzionate (numerosità, status, geografia) permettono di leggere e, talvolta, anticipare conflitti potenziali, in quanto ogni lingua è collegata a fattori molto sensibili quali l'identità (individuale e collettiva), la memoria storica, il potere politico ed economico, e i diritti-doveri che intervengono in tali ambiti (Poggeschi 2015). La gestione della diversità linguistica, intesa nel senso di una complessità multidimensionale che non riguarda mai esclusivamente il piano linguistico-culturale e mobilita porzioni ristrette o molto estese della società, è in definitiva un dossier molto delicato, che coinvolge non solo le politiche pubbliche ma anche gli interventi “dal basso” (attivismo della società civile) e “a mezz'altezza” (Djordjevic 2018). Tale gestione merita, pertanto, un'estrema attenzione.

Gestione della diversità linguistica e intelligenza artificiale: rischi e opportunità

Gestire la diversità linguistica significa, da sempre, governare reti di relazioni umane, sociali, di straordinaria complessità e profondità – e quindi gestire un potere: il mito della Torre di Babele, quale ne sia l'interpretazione, ne è una prova universalmente nota. Ora, le tecnologie digitali amplificano esponenzialmente questa reticolarità, moltiplicando i piani d'interazione: non solo in presenza (oralità, sincrona) e in assenza (scrittura, asincrona), ma anche in presenza-assenza (comunicazione sincrona a distanza), per non considerare poi la questione dell'accessibilità in linea dei testi (biblioteche nazionali, motori di ricerca, banche dati, corpora ecc.). Queste innovazioni tecniche ridefiniscono il paradigma della comunicazione umana, e segnatamente le nozioni di “distanza” tra interattanti, di “spazio” e “tempo” di comunicazione e di “comunità linguistica”. Estendendo vertiginosamente funzionalità e raggio d'azione della comunicazione di tutti i giorni, il digitale accentua o, meglio, esaspera, il valore economico delle lingue naturali, incoraggiando sempre più – *de jure* e *de facto* – un processo di centralizzazione linguistica. Quest'ultima necessita, per funzionare ad alto rendimento, di una semplificazione che agisce su almeno quattro piani:

- 1) riduzione del numero di lingue parlate in un dato contesto (regione, Stato, continente, mondo) e/o adozione di una lingua comune per ridurre le “barriere” all'intercomprensione;
- 2) semplificazione del vocabolario di ciascuna lingua;

- 3) banalizzazione/standardizzazione del discorso, secondo schemi ricorrenti, prevedibili, e quindi, di fatto, sempre potenzialmente stereotipi e di rapida produzione;
- 4) accaparramento del potere di gestione della diversità linguistica da parte della tecnologia stessa e dei suoi depositari.

In questa cornice, culturalmente preoccupante, occorre assumere una postura fortemente critica e vigile. In materia di gestione della diversità linguistica è necessario porsi due domande:

- 1) come reagiscono, di fronte a tali derive centraliste, le lingue-culture / le comunità linguistiche minoritarie, segnatamente quelle numericamente e politicamente più fragili, geograficamente e culturalmente più periferiche, solitamente legate a una comunicazione “personale”, di prossimità?
- 2) Quale funzione può svolgere la cosiddetta “Intelligenza artificiale” (IA), punta attualmente più avanzata della tecnologia digitale e oggetto di numerose riflessioni anche a carattere etico⁴? Accentuare sempre più il centralismo linguistico-culturale-economico (quindi anche politico), oppure equilibrarlo⁵? Una volta di più, è nell’uso delle tecnologie, di per sé neutre, che si determina il carattere virtuoso o nefasto delle stesse⁶.

Sul versante dei rischi, si evidenzia una marcata dissimmetria: alla disponibilità di grandissimi corpora linguistici nelle lingue di larga diffusione internazionale fa riscontro una notevole scarsità e limitatezza di banche dati nelle lingue minoritarie: l’IA lavora meglio con le “grandi lingue” e, fatalmente, meno con le “piccole lingue”. In quest’ottica, si potrebbe dire che “piove sul bagnato” e l’IA, senza un deliberato intervento di pianificazione linguistica, non può che contribuire ad accentuare il centralismo linguistico-culturale su scala globale. Occorre allora documentare sempre più e sempre meglio un numero crescente di lingue meno diffuse

⁴ Sul tema dell’IA, ricordiamo appena come le posizioni ideologiche siano le più varie: da un lato, «L’intelligenza artificiale non esiste» (Julia 2020), ma dall’altro è stato affermato che «Non c’è disciplina scientifica che abbia trasformato il mondo quanto l’intelligenza artificiale» (Ganascia 2021:150). Da parte sua, l’UNESCO considera l’intelligenza artificiale come un insieme di tecnologie avanzate che permettono «alle macchine di imitare alcune delle funzionalità dell’intelligenza umana, comprese caratteristiche come la percezione, l’apprendimento, il ragionamento, la risoluzione di problemi, l’interazione linguistica e persino la produzione di lavori creativi» (UNESCO 2019:10). Il corsivo è nostro.

⁵ Per un’analisi più approfondita del rapporto tra minoranze linguistiche e IA, cf. Agresti 2023.

⁶ Circa l’uso e gli impatti sociali dell’IA, cf. OCDE 2019, Kiyindou 2019. Sul versante dei progetti di ricerca in corso sul tema, si segnaleranno in particolare i lavori dell’asse «Numérique» (“digitale”) della rete di ricerca neo-aquitana sulle francofonie FrancophoNéA (<https://httpfrancophonea.fr>), coordinati dallo stesso Alain Kiyindou.

(pianificazione del corpus), il che, a termine, potrebbe consentire agli utenti di queste lingue di disporre di risorse digitali sempre più estese ed efficaci, su tutte la traduzione automatica da e per tali lingue. La traduzione automatica rappresenta in effetti forse la più grande opportunità per realizzare una gestione virtuosa della diversità linguistica, in quanto permetterebbe di allentare la concorrenza tra lingue in termini di mercato e accrescimento del capitale umano. Ma per raggiungere questo obiettivo occorre investire robustamente in progetti di ricerca: raccolta sul campo di forme orali e scritte della lingua X, standardizzazione ortografica (un problema tradizionalmente molto sentito e sensibile in ambito minoritario), meglio se di tipo “polinomico” (Marcellesi *et al.* 2003), ossia rispettoso della variazione diatopica⁷; digitalizzazione e costituzione di corpora linguistici, meglio se multimediali⁸, in libero accesso; sviluppo di correttori ortografici e di dispositivi per l'aiuto alla redazione di testi in lingua minoritaria sui supporti di comunicazione contemporanei (smartphone, tablet ecc.), anche perché questi ultimi contribuiscono anche a “svecchiare” l'immagine stessa della lingua minoritaria, migliorandone quindi lo status a livello di rappresentazioni sociali e, conseguentemente, a livello di uso e trasmissione (Strubell 1999).

Bibliografia

Agresti Giovanni (2023). « Intelligence artificielle et langues minoritaires : du bon ménage? Quelques pistes de réflexion ». In: Rachele Raus (cur.) *et al.*. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l'aune de l'intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell'era dell'intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 47-68. <https://www.collane.unito.it/oa/items/show/132>.

De Mauro Tullio (2014). *In Europa son già 103. Troppe lingue per una democrazia?* Bari: Laterza.

⁷ L'approccio polinomico è spesso fondamentale. Nei contesti linguistici minoritari, segnatamente ultraminoritari, non esiste in genere una capitale o un'accademia che possa imporre una norma unica, uno standard, come a livello di uno Stato in rapporto alla lingua nazionale. Imporre una norma che possa risultare non familiare, non autentica, e quindi artificiale, ai locutori di tale lingua minoritaria è un rischio troppo elevato che occorre evitare di correre.

⁸ Un esempio di raccolta multimediale della memoria orale delle montagne europee è il progetto Rete Tramontana (<https://www.re-tramontana.org/fr/>), che intende contribuire in modo significativo alla costituzione di corpora linguistici ed etnografici “periferici” ed altamente conservativi.

Djordjević-Léonard Ksenija (2018). « Linguistes, activistes et locuteurs : trois terrains croisés (vepse, tabarquin, croate molisain) ». *Études finno-ougriennes*, 49-50. DOI: <https://doi.org/10.4000/efo.9951>

Francesco (2015). *Laudato si'. Lettera Enciclica sulla cura della casa comune*. Roma: Libreria Editrice Vaticana.

Ganascia Jean-Gabrielle (2021). « Intelligence Artificielle : Des Big-Data au Cerveau ». In *Les signatures neurobiologiques de la conscience. Neurobiologie fonctionnelle, phénomènes de conscience, cognition, automates « intelligents », éthique*. Les Ulis: EDP Sciences, 147-162.

Gazzola Michele (2016). « Multilinguisme et équité: l'impact d'un changement de régime linguistique européen en Espagne, France et Italie ». In: Giovanni Agresti, Joseph-G. Turi (a cura di). *Représentations sociales des langues et politiques linguistiques. Déterminismes, implications, regards croisés. Actes du Premier Congrès mondial des droits linguistiques, Vol. I*. Roma: Aracne, 269-286.

Julia Luc (2020). *L'intelligence artificielle n'existe pas*. Parigi: J'ai lu.

Kiyindou Alain (2019). *Intelligence artificielle. Pratique et enjeux pour le développement*. Parigi: L'Harmattan.

Le Coadic Ronan (2010). « Diversité, liberté, vitalité ». In: Giovanni Agresti, Mariapia D'Angelo (a cura di). *Rovesciare Babele. Economia ed ecologia delle lingue regionali o minoritarie*. Roma: Aracne, 51-72.

Marcellesi Jean-Baptiste, Bulot Thierry, Blanchet Philippe (a cura di) (2003). *Sociolinguistique. Epistémologie, Langues régionales, Polynomie*. Parigi: L'Harmattan.

OCDE (2019). *Considérations de politique publique. L'intelligence artificielle dans la société*. Parigi: Éditions OCDE. <https://doi.org/10.1787/b7f8cd16-fr>.

Poggeschi Giovanni (2015). «La mediazione linguistica e culturale come strumento esemplare per la vigenza dei diritti linguistici di prima specie». *Lingue e Linguaggi*, 16, 435-443. DOI: [10.1285/i22390359v16p435](https://doi.org/10.1285/i22390359v16p435)

Strubell Miquel (1999). «From Language Planning to Language Policies and Language Politics». In: Pater J. Weber (a cura di). *Contact + Confl(c)t: language planning and minorities*. Bonn: Dummler, 237-248.

UNESCO (2019). *Steering AI and Advanced ICTs for Knowledge Societies*. Parigi: Unesco.

Questo contributo discute il ruolo dei corpora come risorse per realizzare l'uguaglianza tra tutte le lingue ufficiali (e, di conseguenza, tra tutti i cittadini) dell'Unione Europea (UE). Il multilinguismo è uno degli aspetti costitutivi dell'identità culturale, sociale e politica dell'UE fin dalla fondazione, e la sua rilevanza anche in proiezione futura è stata ribadita con forza nella risoluzione relativa all'uguaglianza linguistica nell'era digitale in Europa, adottata dal Parlamento Europeo a larghissima maggioranza nel settembre 2018 (Parlamento Europeo, 2018), anche in risposta alle crescenti preoccupazioni relative all'indebolimento e ai rischi di morte ed estinzione di alcune lingue europee (si vedano ad esempio Moseley 2010; Rehm, Uszkoreit 2012; Kornai 2013; Ceberio Berger *et al.* 2018).

Entro questa cornice, in cui convivono motivi di allarme e slanci propositivi volti alla tutela e alla promozione del multilinguismo in Europa, questo contributo si concentra sulle ragioni per cui i corpora elettronici costituiscono risorse essenziali per assicurare la necessaria vitalità digitale delle lingue europee in un'ottica paritaria, non sulla base di un astratto principio ugualitario di esclusivo interesse degli studiosi e dei linguisti, ma per assicurare il benessere e la rilevanza di tutte le comunità di parlanti, anche a livello educativo, culturale, sociale, politico ed economico. Le tecnologie linguistiche odierne basate sull'intelligenza artificiale (IA) necessitano di dati digitali mono- e/o multilingui per essere realizzate: dai correttori ortografici, agli strumenti di traduzione automatica fino ai sistemi di sintesi e riconoscimento vocale, i progressi e i vantaggi dell'IA possono essere sfruttati solamente grazie alla disponibilità di dati linguistici in formato elettronico, in grandi quantità, di elevata qualità e corrispondenti agli ambiti di utilizzo e ai domini di applicazione delle tecnologie (Vetere 2023).

Due progetti internazionali finanziati dalla Commissione Europea, *European Language Grid*¹ (ELG) (Rehm 2023) e *European Language Equality*² (ELE) (Rehm *et al.* 2022; Rehm, Way 2023) hanno effettuato ampie raccolte di dati e strumenti linguistici e svolto consultazioni capillari con esperti e rappresentanti di tutte le comunità linguistiche d'Europa, rivolgendosi non solamente a linguisti e studiosi, ma anche a esponenti dell'industria e del mercato, utenti, consumatori, attivisti, politici e decisori a livello degli organismi dell'UE, degli Stati Membri e delle amministrazioni regionali e delle autorità locali, per porre le basi di un ambizioso programma per raggiungere l'uguaglianza digitale tra tutte le lingue europee. Sulla base di questo imponente lavoro, la

I corpora come risorse per l'uguaglianza digitale tra le lingue ufficiali dell'Unione Europea

Federico Gaspari

Università di Napoli "Federico II"

¹ <https://live.european-language-grid.eu>

² <https://european-language-equality.eu>

disponibilità di corpora elettronici è risultata un fattore decisivo per livellare al rialzo le prospettive di vitalità e sostenibilità di tutte le lingue ufficiali dell'UE nell'era digitale.

Una constatazione che emerge dall'analisi delle risorse linguistiche attualmente disponibili riguarda la situazione di evidente disuguaglianza e grave squilibrio, in cui risultano privilegiate pochissime lingue dominanti, a discapito dell'ampia maggioranza. ELG e ELE hanno creato e messo a disposizione online un osservatorio interattivo³ che, tramite visualizzazioni dinamiche guidate direttamente dall'utente e fondate su informazioni costantemente aggiornate, consente di misurare l'attuale stato di supporto digitale delle varie lingue europee e di confrontare il rispettivo livello di prontezza digitale, aggregando o scorporando le varie tipologie di strumenti, applicazioni o risorse linguistiche di interesse presenti nel Catalogo ELG. La Figura 1 presenta la disponibilità di corpora elettronici per le 24 lingue ufficiali dell'UE⁴, riportandone rispettivamente il conteggio complessivo in modalità aggregata (pannello 1) e disgiunta, con il dettaglio specifico dei corpora monolingui (pannello 2), bilingui (pannello 3) e multilingui (pannello 4).

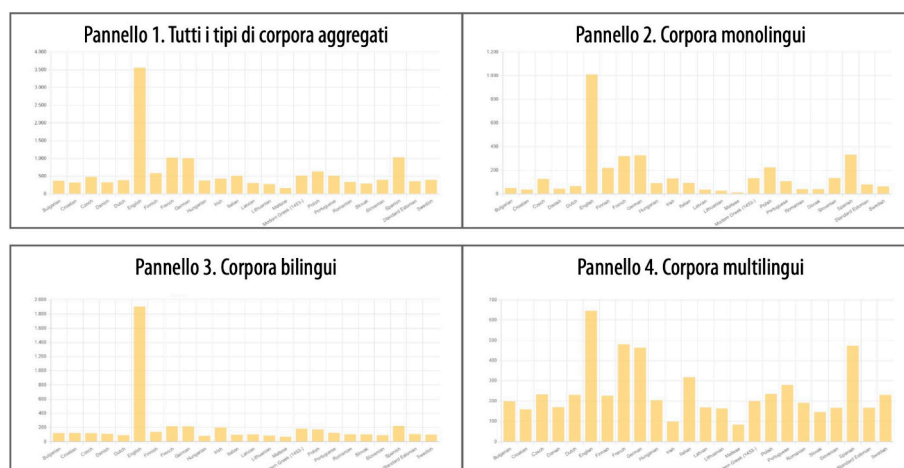


Figura 1. Disponibilità delle varie tipologie di corpora elettronici per le lingue ufficiali dell'UE nel Catalogo ELG (fonte: <https://live.european-language-grid.eu/catalogue/dashboard>)

Se la posizione privilegiata di alcune lingue è prevedibile, anche sulla base di considerazioni legate alle dimensioni delle relative comunità di parlanti e alla loro diffusione internazionale,

³ L'osservatorio interattivo online è disponibile all'indirizzo <https://live.european-language-grid.eu/catalogue/dashboard>.

⁴ I grafici riportati nella Figura 1 sono tratti dall'osservatorio già citato e sono basati su dati aggiornati al 31 maggio 2023. Si limitano alle lingue ufficiali dell'UE, ma il database interattivo raccoglie e permette di visualizzare e confrontare con le medesime modalità anche dati aggiornati relativi a una settantina di altre lingue regionali, minoritarie e co-ufficiali dell'Europa, appositamente tutelate (Consiglio d'Europa, 1992), alcune delle quali sono utilizzate da comunità di poche migliaia di persone.

non solamente in Europa ma anche al di fuori di essa (con l'inglese, lo spagnolo, il francese e il tedesco che spiccano complessivamente per la disponibilità di corpora), emergono anche situazioni sorprendenti e preoccupanti: in primo luogo, l'enorme divario in termini di disponibilità di corpora di tutti i tipi tra l'inglese e le altre lingue ufficiali dell'UE, comprese quelle relativamente più ricche di risorse linguistiche. In secondo luogo, lingue ufficiali utilizzate in alcuni dei Paesi più grandi d'Europa (ad esempio l'Italia e la Polonia) hanno una scarsa disponibilità di corpora in relazione alle rispettive popolazioni. Inoltre, nei casi di corpora con dati in due o più lingue, questi abbracciano perlopiù le lingue privilegiate, a partire dall'inglese, e includono soltanto in maniera molto limitata le altre lingue meno supportate. Infine, volendosi limitare per brevità alle disuguaglianze principali, è palese la penuria di corpora digitali di tutte le tipologie analizzate per la maggior parte delle lingue ufficiali europee, molte delle quali si trovano in una situazione deficitaria non dissimile, per la verità, da quella di moltissime delle altre circa 70 lingue regionali, minoritarie o co-ufficiali d'Europa censite nel Catalogo ELG e visualizzabili nel dettaglio tramite l'osservatorio interattivo online.

Va chiarito che questa rapida disamina complessiva si basa esclusivamente sul conteggio dei corpora disponibili nel Catalogo ELG per le varie lingue ufficiali dell'UE, senza considerarne le dimensioni, l'effettiva qualità e la diversificazione in termini di natura dei dati (esclusivamente testuali e/o anche orali e/o video) nonché in relazione ai domini settoriali e alle categorie testuali coperte (per ulteriori considerazioni relative all'importanza di tipologie variegata e innovative di corpora, si veda Gaspari 2022: 50 e segg.). Non sorprenderà che un'analisi più approfondita di questi fattori rende ancora più manifesto e pronunciato lo squilibrio della disponibilità di corpora a favore dell'inglese, seguito (sempre a considerevole distanza) dal terzetto di lingue già ricordate sopra, con tutte le altre lingue ufficiali europee notevolmente svantaggiate (al riguardo, cfr. anche Vetere 2023).

Dalle considerazioni svolte, basate su dati empirici validati dalle comunità di esperti consultati dai progetti internazionali ELG e ELE, appare quindi palese la disuguaglianza tra le lingue ufficiali europee a livello di disponibilità di corpora digitali, che sono cruciali per avvalersi degli ultimi sviluppi dell'IA nello sviluppo di tecnologie linguistiche. Lungi dal riguardare astrattamente sistemi linguistici in modo avulso dalla vita reale, e dunque dall'essere di esclusivo interesse degli studiosi di linguistica o di tecnologie, questi squilibri investono al contrario direttamente i cittadini europei, in quanto membri delle rispettive comunità di parlanti. In-

fatti, le disuguaglianze riscontrate riflettono, e allo stesso tempo aggravano, le differenze in termini di sostegno attuale e di prospettive future di prosperità che le varie comunità linguistiche d'Europa hanno nell'era digitale, anche per quanto riguarda il progresso educativo, culturale, sociale, politico ed economico delle rispettive popolazioni. Queste asimmetrie hanno rilievo sia nei rapporti interni tra i cittadini europei sia, in un'ottica più ampia, rispetto a quelli di Paesi anche al di fuori dell'UE, in un mondo sempre più globalizzato e interconnesso, in cui le pressioni verso l'omogeneizzazione e l'adeguamento a modelli linguistico-culturali dominanti imposti dall'alto, esercitate anche attraverso la supremazia tecnologica, risultano sempre più intense.

In conclusione, appare quanto mai urgente una presa di coscienza collettiva da parte di tutti i cittadini europei sui temi dell'uguaglianza linguistica digitale e delle prospettive di sviluppo connesse. In questo senso, a livello istituzionale, vanno promosse con coraggio e lungimiranza politiche di ricerca e sviluppo sostenute da adeguati finanziamenti da parte dei politici e dei decisori interessati, a partire da coloro che operano nelle Istituzioni Europee, passando per gli Stati Membri, fino alle amministrazioni regionali e alle autorità locali che hanno a cuore la tutela dei diritti linguistici e delle identità socio-culturali delle proprie comunità.

Bibliografia

Ceberio Berger Klara, Gurrutxaga Hernaiz Antton, Baroni Paola, Hicks Davyth, Kruse Eleonore, Quochi Valeria, Russo Irene, Salonen Tuomo, Sarhimaa Anneli, Soria Claudia (2018). *Digital Language Survival Kit: The DLDP Recommendations to Improve Digital Vitality*. Bruxelles: European Commission; Erasmus+ Programme. www.dldp.eu/sites/default/files/documents/DLDP_Digital-Language-Survival-Kit.pdf

Consiglio d'Europa (1992). *Carta europea delle lingue regionali o minoritarie*. Strasburgo: Consiglio d'Europa. <https://rm.coe.int/168007c095>.

Gaspari Federico (2022). "Expanding the Reach of Corpus-Based Translation Studies: The Opportunities that Lie Ahead". In: Sylviane Granger e Marie-Aude Lefer (a cura di). *Extending the Scope of Corpus-Based Translation Studies*. London: Bloomsbury, 42-63.

Kornai András. (2013). "Digital Language Death". *PLoS ONE* 8(10): e77056. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0077056>

Moseley Christopher, Nicolas Alexandre (a cura di) (2010). *Atlas of the World's Languages in Danger*. 3ª edizione. Parigi: UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000187026> (accessed 23/07/2023).

Parlamento Europeo (2018). *Risoluzione del Parlamento Europeo dell'11 settembre 2018 sull'uguaglianza linguistica nell'era digitale (2018/2028(INI))*. Bruxelles: Parlamento europeo. http://www.europarl.europa.eu/doceo/document/TA-8-2018-0332_IT.html

Rehm Georg (a cura di) (2023). *European Language Grid: A Language Technology Platform for Multilingual Europe*. Berlino: Springer.

Rehm Georg, Gaspari Federico, Rigau German, Giagkou Maria, Piperidis Stelios, Grützner-Zahn Annika, Resende Natalia, Hajič Jan, Way Andy (2022). "The European Language Equality Project: Enabling digital language equality for all European languages by 2030". In: Željko Jozić e Sabine Kirchmeier (a cura di). *The Role of National Language Institutions in the Digital Age*. Budapest: Hungarian Research Centre for Linguistics, 17-47.

Rehm Georg, Uszkoreit Hans (a cura di) (2012). "About META-NET". In: Georg Rehm, Hans Uszkoreit (eds). *The Danish Language in the Digital Age. White Paper Series*. Berlin: Springer. DOI: https://doi.org/10.1007/978-3-642-30627-3_10

Rehm Goerg, Way Andy (a cura di) (2023). *European Language Equality: A Strategic Agenda for Digital Language Equality*. Berlino: Springer.

Vetere Guido (2023). "Elaborazione automatica dei linguaggi diversi dall'inglese: introduzione, stato dell'arte e prospettive". In: Rachele Raus et al. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l'aune de l'intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell'era dell'intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 69-87. <https://www.collane.unito.it/oa/items/show/132>

La lingua umana, in virtù della sua natura dinamica e complessa, è strettamente influenzata da fattori e variabili sociali, culturali e geografici relativi ad un determinato contesto spazio-temporale antropico. L'analisi delle varietà presenti all'interno di una comunità linguistica rivela una sorprendente diversità di forme e manifestazioni che confermano l'assunto universale per cui ogni lingua possiede una certa variazione interna (Berruto, Cerruti 2019).

Ogni parlante, realizzando forme e costrutti linguistici lungo le quattro dimensioni di variazione¹, si discosta dallo standard² modificando e adattando il proprio sistema linguistico alle esigenze pratiche del quotidiano, generando dunque un progressivo allargamento del concetto di norma³.

In tale contesto e con il rapido sviluppo dell'intelligenza artificiale (IA), si delineano criticità che potrebbero mettere a rischio il diversificato sistema linguistico. Le varietà linguistiche, considerate dall'industria globale come problematiche e disturbanti, rischiano di essere cancellate dal rapido sviluppo dell'IA e dalla creazione di quel "mondo ideale" in cui si parla una sola lingua⁴. Anche considerando possibile l'idea di sviluppare delle IA in grado comunicare in tutte le lingue del mondo, verrebbe comunque a mancare la gestione delle numerose varietà interne. Un esempio sono gli assistenti digitali: programmati per utilizzare un linguaggio standardizzato, limitano la creatività e l'espressività nelle interazioni.

Guardando al panorama linguistico italiano, la diffusione dell'IA ha suscitato interrogativi sulla conservazione delle diversità e unicità linguistiche. Per tale ragione, risulta indispensabile l'adozione di un approccio globale e integrato che coinvolga le comunità locali e diversi *stakeholder* nell'ideazione e validazione di modelli in grado di comprendere e utilizzare le peculiarità linguistiche e culturali presenti su tutto il territorio nazionale⁵.

Sebbene i tratti linguistici costituenti dello standard italiano siano descritti e rappresentati all'interno di vari corpora linguistici, lo stesso non si può dire per tutte le altre varietà, ampiamente

La lingua italiana: salvaguardia delle varietà linguistiche dall'impatto con l'intelligenza artificiale

Chiara Russo

Università degli Studi di Catania

¹ Ci si riferisce alla variazione diatopica, diastratica, diafasica e diamesica.

² In linguistica, il termine italiano standard indica una specifica varietà assunta dai parlanti come modello e quindi non connotata socialmente o regionalmente (Marzullo M. 2005).

³ Nel processo evolutivo della lingua italiana, sia la varietà standard che il parlato hanno rivestito un ruolo fondamentale: lo standard si è adeguato alle necessità del quotidiano, mentre il parlato ha esercitato pressione sulle strutture della lingua scritta, portando ad una serie di modifiche strutturali del sistema linguistico.

⁴ Nel mondo dell'industria è forte la tentazione di utilizzare la lingua inglese come "lingua ponte" affidandosi esclusivamente alla traduzione automatica.

⁵ Il discorso è replicabile per tutte le lingue europee, anch'esse ricche di varietà interna.

diffuse tra le comunità di parlanti ma ad oggi poco rappresentate nello scenario tecnologico. Le banche dati e i corpora che rappresentano le varietà della lingua italiana, rispettivamente scritte e parlate, presentano problemi di rappresentatività, flessibilità e scalabilità⁶.

La tutela delle diverse varietà linguistiche nell'ambito di un utilizzo massivo dell'intelligenza artificiale rappresenta una sfida di notevole complessità ma si delineano possibili misure e strategie da adottare per affrontarla.

Riveste un ruolo fondamentale la raccolta di dati rappresentativi e di alta qualità che permettano all'IA di essere addestrata su un insieme diversificato di dati. Tale procedura concorrerebbe alla prevenzione di modelli linguistici basati su una o poche varietà, che potrebbero compromettere il riflesso della reale complessità della lingua stessa e delle corrispondenti culture locali. Considerando la difficoltà di tale compito, sarebbe auspicabile l'attuazione di normative, su base nazionale o regionale, che prevedano la definizione di modelli specifici per garantire un'elevata accuratezza.

Inoltre, di fondamentale importanza sono il monitoraggio e il continuo aggiornamento attraverso il coinvolgimento di personale esperto in linguistica in grado di gestire accuratamente i dati raccolti. In sintesi, la diversità linguistica italiana rappresenta un patrimonio culturale prezioso che testimonia la ricchezza culturale del Paese e l'IA, se correttamente utilizzata, può diventare un ottimo strumento per la valorizzazione, conservazione e tutela delle varietà anche attraverso la creazione di risorse linguistiche digitali quali applicazioni educative e di apprendimento⁷.

Bibliografia

Berruto Gaetano (2012). *Sociolinguistica dell'italiano contemporaneo*. Roma: Carocci.

Berruto Gaetano, Cerruti Massimo (2019). *Manuale di Sociolinguistica*. Novara: De Agostini Scuola SpA.

⁶ L'elenco delle banche dati, corpora e archivi testuali attualmente disponibili è consultabile al seguente link: <https://accademiadellacrusca.it/it/contenuti/banche-dati-corpora-e-archivi-testuali/6228>

⁷ Interessante è il progetto *Lahjoita puhetta*, sostenuto dalla Yle, dalla società di sviluppo statale finlandese Vake e con il contributo di esperti dell'Università di Helsinki. Vincitore del *festival Prix Europa 2021*, il progetto prevede la raccolta di brevi discorsi, aperti all'intera popolazione e su base volontaria, per lo sviluppo di un'intelligenza artificiale capace di comprendere e modellare le caratteristiche delle varietà della lingua finlandese. L'obiettivo è quello di sviluppare applicazioni e servizi a comando vocale che funzionino senza problemi con tutti i parlanti. Cfr. <https://www.yle.fi/lahjoitapuhetta>

De Mauro Tullio (2021). *Storia linguistica dell'Italia repubblicana dal 1946 ai nostri giorni*. Bari-Roma: Laterza.

De Mauro Tullio. (2020). *Storia linguistica dell'Italia unita*. Bari-Roma: Laterza.

Marzullo Mara (2005). *Etimologia e origine della parola standard*. Firenze: Accademia della Crusca. <https://accademiadellacrusca.it/it/consulenza/etimologia-e-origine-della-parola-standard/154>

Sabatini Francesco (1985). "L'italiano dell'uso medio: una realtà tra le varie linguistiche italiane". In: Gunther Holtus, Edgar Radtke (Hrsg.). *Gesprochenes Italienisch in Geschichte und Gegenwart*. Tübingen: Narr.

Numerose ricerche sono state condotte negli ultimi sessant'anni sulla necessità di riconoscere la legittimità linguistica del francese in uso nei paesi della francofonia e, in particolare, del francese parlato nel Canada francofono¹. Questa battaglia per il riconoscimento della variazione diatopica della lingua francese viene continuata attualmente da altri linguisti, tra cui menzioniamo Nadine Vincent et Wim Remysen (2016), autori del primo dizionario, basato interamente su corpora testuali, concepito in Québec per tutti i francofoni e francofili interessati a una descrizione “aperta” del francese².

Se è dunque possibile affermare oggi che la legittimità linguistica del francese del Québec, a lungo discriminato come un francese “deviato” rispetto alla norma di riferimento di Francia, sia stata raggiunta in campo lessicografico (Zotti 2012), e questo sia a livello transnazionale che nazionale (si pensi all'integrazione di numerosi “francofonismi” nelle ultime edizioni dei dizionari francesi Robert et Larousse, cf. Cormier *et al.* 2013), nell'ambito della traduzione automatica siamo ancora lontani da questo traguardo. Un'indagine (Zotti 2021) condotta su un campione di testi della letteratura quebecchese³ ha fatto emergere la quasi totale assenza di varianti diatopiche del Québec della lingua francese nei corpora su cui si basano i principali strumenti di traduzione automatica disponibili gratuitamente on line: *Google Translate* e *DeepL*. I testi bilingui del Parlamento canadese costituiscono un'importante banca dati per il binomio francese/inglese, ma non per altre coppie linguistiche, come, nel caso specifico francese/italiano, e, in ogni caso la copertura resta limitata al dominio giuridico-amministrativo. Il tasso di parole del francese quebecchese non tradotte o tradotte in maniera errata risulta dunque elevato sia per le varianti lessematiche (forme in uso solo in francese del Québec) che per le varianti semantiche (forme esistenti nel francese di riferimento con un significato diverso). Queste mancanze o errori comportano dei rischi relativi alla diffusione di traduzioni sbagliate nei vari campi del sapere, pericolosa se estesa anche al mondo della formazione.

Una riflessione sui dati che nutrono gli strumenti di traduzione automatica si impone. Se la letteratura scientifica risulta ormai abbastanza ampia rispetto al “*gender bias*” (Temmerman 2021)

¹ Si pensi agli studi di Claude Poirier, direttore per molti anni del laboratorio di ricerca *Trésor de la Langue Française au Québec* all'Université Laval à Québec, e dei suoi allievi, tra cui menzioniamo Louis Mercier e Hélène Cajolet-Laganière 2004.

² USITO, cfr. <https://usito.usherbrooke.ca/>

³ Poesia: Gaston Miron; prosa: novelle e romanzi del “terroir”, fortemente connotati socio-culturalmente e caratterizzati dall'abbondanza di amerindianismi e *realia*.

e ai cosiddetti “*bias* algoritmici” (Conseil de l’Europe 2019), sono molto meno numerosi gli studiosi che si sono dedicati a ricerche e a test sulla performance degli strumenti di TA per quanto riguarda la traduzione di testi in lingue “poco dotate” (Le 2019), ovvero di lingue e “varianti regionali” che soffrono la mancanza di risorse linguistiche, come i corpora paralleli, fondamentali per lo sviluppo di sistemi performanti di trattamento automatico delle lingue naturali. Questa riflessione si collega a un’altra preoccupazione relativa alla creazione, ormai generalizzata, di risorse plurilingui di traduzione (ad es. *BabelNet*) basate sul *web crawling*, ovvero sulla estrazione automatica di dati “non verificati” dal *web* che non fanno altro che generare traduzioni poco precise in ambiti specialistici, nonché stereotipi culturali e linguistici.

L’incapacità dei sistemi di traduzione automatica statistica di generare *output* diversificati e la loro tendenza a riprodurre i “modelli” più frequenti ignorando quelli meno frequenti è un’evidenza. L’insufficienza nel generare *output* diversificati è stata osservata anche nei modelli neurali per attività che comportano la generazione di linguaggio (es. ChatGPT). Infatti, mentre per entrambi i modelli di apprendimento la generazione di traduzioni accurate è stato l’obiettivo principale, il mantenimento della ricchezza lessicale e la creazione di *output* diversificati non sono stati invece una priorità. Un esempio molto semplice ma particolarmente significativo nell’ambito della variazione diatopica della lingua francese è il seguente: se si chiede a uno strumento di TA di tradurre in francese un sintagma italiano come “il sindaco di < città >”, è necessario che il sistema di traduzione automatica generi, in lingua francese, sia “le *maire* de Lyon/Paris/Bordeaux” sia “le *bourgmestre* de Bruxelles/Liège/Anvers”, essendo il termine “*bourgmestre*” in uso in Belgio. Allo stesso modo, “la *maire* de ...”, forma femminile di “*maire*” in uso in Québec, in contrapposizione alla forma “*une maire*” usata in Francia per il genere femminile, dovrebbe essere tradotta con il corrispettivo italiano per il genere femminile “la sindaca di...”, e non ricondotta alla forma più diffusa di genere maschile “il sindaco” (cfr. *Google Translate* e *DeepL*, ultima consultazione 14/04/2023).

Consideriamo che la tendenza degli attuali sistemi di traduzione automatica a generalizzare possa avere seri inconvenienti conducendo alla generalizzazione di una visione monolitica della complessità linguistica delle società umane. L’eccessiva generalizzazione su un *input* visto e l’esacerbazione delle forme dominanti potrebbero non solo portare a una perdita di scelta lessicale, ma potrebbero anche essere la causa sottostante dell’esacerbazione di pregiudizi sociali rispetto alle minoranze linguistiche.

Come osservato dallo stesso Consiglio d'Europa (2019): «La mancanza di diversità e inclusione che caratterizza la progettazione dei sistemi di intelligenza artificiale è una delle principali preoccupazioni: invece di rendere le nostre decisioni più obiettive, questi sistemi rischiano di rafforzare la discriminazione e il pregiudizio dando loro l'apparenza di oggettività». L'IA è una tecnologia strategica che offre molti vantaggi ai cittadini, alle imprese e alla società nel suo complesso, a condizione che sia etica, sostenibile, incentrata sull'uomo e rispetti i diritti e i valori fondamentali. E a condizione che tenga conto del multilinguismo e del plurilinguismo (Temmerman 2021).

Si evince che il lavoro di specialisti di linguistica dei corpora nel migliorare i modelli di traduzione automatica neurale fornendo dati monolingui di "qualità" (Sennrich *et al.* 2016) rispettosi del multilinguismo interno ed esterno sia essenziale e che la preservazione della diversità, sebbene finora non considerata una priorità in questo ambito, sia importante anche nella traduzione automatica.

Bibliografia

Bentivogli Luisa, Bisazza Arianna, Cettolo Mauro, Marcello Federico (2016). "Neural versus Phrase-Based Machine Translation Quality: a Case Study". In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin: EMNLP, 257-267.

Conseil de l'Europe (2019). *Comblent l'écart. Comment garantir les droits de l'homme pour tous. Compilation du Carnet des droits de l'homme*. Articles publiés en 2018 et 2019 par Dunja Mijatović, Commissaire aux droits de l'homme du Conseil de l'Europe. Bruxelles: Conseil de l'Europe.

Cormier Monique C., Francoeur Aline, Boulanger Jean-Claude (a cura di) (2003). *Les dictionnaires Le Robert : Genèse et évolution*. Montréal: Presses de l'Université de Montréal. <http://books.openedition.org/pum/13849>

Le Ngoc Tan (2019). *Traduction automatique pour une paire de langues peu dotée*. Thèse de Doctorat en informatique cognitive. Montréal: Université du Québec.

Mercier Louis, Cajolet-Laganière Hélène (a cura di) (2004). « Français du Canada - Français de France VI » *Actes du Sixième colloque international d'Orford, du 26 au 29 septembre 2000 (Canadiana Romanica, 18)*. Tübingen: Niemeyer, 365.

Sennrich Rico, Haddow Barry, Birch Alexandra (2016). "Improving Neural Machine Translation Models with Monolingual Data". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*. Berlin: Association for Computational Linguistics, 86-96. Doi: 10.18653/v1/P16-1009

Temmerman Rita (2021). « La créativité plurilingue, un obstacle pour l'IA? », intervento al convegno *Linguistic Rights and Language Varieties in Europe in the Age of AI*, Università di Torino, 21 aprile 2021. <https://www.jmcoe.unito.it/content/linguistic-rights-and-language-varieties-europe-age-ai>

Wim Remysen, Nadine Vincent (a cura di) (2016). *La langue française au Québec et ailleurs: Patrimoine linguistique, socioculture et modèles de référence*. Frankfurt: Peter Lang.

Zotti Valeria (2021). « Intelligence Artificielle et diversité linguistique : quel regard sur la langue française au Québec ? ». Intervento al Convegno internazionale *'Regards croisés sur le Québec et la France'*. Centro Interuniversitario di Studi Quebecchesi (Trento, 20-22 maggio 2021). USITO: <https://usito.usherbrooke.ca>

Zotti Valeria (2019). « Ressources numériques pour la traduction des mots désignant des *realia* ». *Etudes de Linguistique Appliquée*, 194, 227-246.

Zotti Valeria (2012). « La légitimité linguistique du français québécois passe-t-elle par la reconnaissance lexicographique d'une langue et d'une culture autres? La contribution d'un dictionnaire différentiel bilingue français québécois / italien ». In: *Lexiques Identités Cultures*. Verona : QuiEdit, 471-489.

L'attuale disponibilità di risorse computazionali e di tecnologie di Intelligenza Artificiale capaci di processare lingue e testi è fortemente sbilanciata a favore dell'inglese; la maggior parte dei dati e dei software *open-source*, spesso gli unici alla portata di piccole e medie imprese, è sviluppata solo sulla lingua inglese lasciando scoperte tutte le altre lingue (Vetere 2022). Di conseguenza, vi è una barriera all'accesso ai servizi più avanzati che la tecnologia può fornire, una barriera causata semplicemente dall'appartenere a una comunità linguistica diversa da quella anglofona. Questa situazione comporta una limitazione della specificità e, soprattutto, della qualità e delle possibilità di sviluppo dei mercati collegati alle lingue diverse dall'inglese: potremmo definirla una vera e propria situazione di minoranza economica e politica, col rischio che l'intero mercato europeo si ritrovi nell'impossibilità di creare innovazione in vari campi chiave per il suo sviluppo.

Uno degli ambiti linguistici oggi più toccati dall'elaborazione automatica dei dati è infatti quello relativo alle lingue speciali, strumenti indispensabili per la comunicazione efficace per i settori dell'economia e della politica (Morresi 1998), non senza ricadute sui fenomeni di contatto linguistico e sull'evoluzione stessa degli stili e dei mezzi espressivi delle lingue comuni (Cortelazzo 1994). Da sempre, l'innovazione tecnologica porta alla coniazione di termini specifici, tanto che si può dire che la situazione delle lingue speciali riflette sul piano internazionale la capacità di sviluppo e di investimento di uno Stato in un dato settore economico: come esempio, si pensi proprio alla situazione della lingua speciale dell'informatica, dominata dall'inglese; ma, per contro, la lingua speciale della produzione vinicola riflette invece la situazione di leadership dell'Italia e della Francia sul mercato globale.

La comparazione di corpora multilingui consente quindi di monitorare la situazione delle lingue speciali sul piano internazionale, individuandone le equivalenze traduttive e, al contempo, le differenze e le specificità di ogni mercato locale. L'apprendimento dell'Intelligenza Artificiale in questo campo è particolarmente sfidante perché dovrebbe, in una situazione ottimale, applicare un approccio cosiddetto onomasiologico (vale a dire che apprenda le equivalenze traduttive fra lingua e lingua a partire dai concetti) e non solo esclusivamente semasiologico (che invece traduce a partire dai significati); da questo punto di vista, la precisione richiesta dalle lingue speciali comporta ancora un'attenta fase di *post-editing* alla traduzione automatica.

Per citare un campo particolarmente sensibile, si pensi a quello giuridico, dove vige un continuo conflitto dinamico fra le richieste di un'armonizzazione internazionale e le necessità di garanti-

Corpora multilingui e lingue speciali: preservare la diatopia linguistica

Marta Muscariello
Università IULM di Milano

re le specificità di ogni singolo sistema nazionale. La creazione di corpora specifici, distinti per ogni Stato (quindi tenendo separati anche i corpora di Stati con la stessa lingua), consente la comparazione della situazione linguistica e concettuale della giurisprudenza e del diritto sia all'interno del mondo europeo sia rispetto alla più ampia situazione globale. La situazione di ogni Stato viene così a delinarsi nella propria specificità giuridica e, al contempo, nel suo rapporto dinamico con il diritto sovranazionale europeo attualmente in via di sviluppo (Rossini Favretti 1999; Felici, Mori 2019).

Bibliografia

Cortelazzo Michele (1994). *Le lingue speciali: la dimensione verticale*. Padova: Unipress.

Felici Annarita, Mori Laura (2019). "Corpora di italiano legislativo a confronto: dall'Unione Europea alla Cancelleria svizzera". In: Bruno Moretti, Aline Kunz, Silvia Natale e Etna Krakenberger (a cura di). *Le tendenze dell'italiano contemporaneo rivisitate. Atti del LII Congresso Internazionale di Studi della Società di Linguistica Italiana (Berna, 6-8 settembre 2018)*. Milano: Officinaventuno.

Morresi Ruggero (a cura di) (1998). *Le lingue speciali. Atti del convegno di studi, Università di Macerata, 17-19 ottobre 1994*. Roma: Il Calamo.

Rossini Favretti Rema (1999). "Equivalenze traduttive in corpora giuridici multilingue". *Quaderni di Libri e Riviste d'Italia*. Roma: Istituto Poligrafico e Zecca dello Stato, 47-66.

Vetere Guido (2023). "Elaborazione automatica dei linguaggi diversi dall'inglese: introduzione, stato dell'arte e prospettive". In: Rachele Raus et al. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l'aune de l'intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell'era dell'intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 69-87. <https://www.collane.unito.it/oa/items/show/132>



RACCOMANDAZIONE 4

Occorre investire nello sviluppo di tecnologie linguistiche e computazionali fattivamente made in EU



R4

Il rapido sviluppo dell'intelligenza artificiale (IA) ha portato a una dipendenza sempre maggiore dalle risorse IA sviluppate da grandi aziende principalmente al di fuori dell'Europa. Questa condizione, da un lato, mette in discussione l'indipendenza tecnologica dell'Europa e, dall'altro, solleva preoccupazioni sulla sicurezza dei dati degli utenti europei. In questo contesto, l'utilizzo passivo di modelli IA per il processamento del linguaggio naturale potrebbe comportare il rischio di omogeneizzazione linguistica e culturale. Lo sviluppo di risorse IA autonome e indipendenti in Europa potrebbe offrire una soluzione per preservare la diversità linguistica e culturale del continente, garantire la protezione dei dati degli utenti europei e promuovere l'indipendenza tecnologica.

L'industria dell'intelligenza artificiale rappresenta una risorsa strategica di grande importanza per molti paesi. In Europa, molte delle tecnologie basate sull'IA utilizzate per il processamento del linguaggio naturale sono fornite da grandi aziende tecnologiche, molte delle quali hanno sede al di fuori dei confini europei. Di conseguenza, l'Europa sta diventando sempre più dipendente dalle tecnologie importate da questi paesi. Questo può comportare una serie di rischi, in quanto le tecnologie importate potrebbero non essere conformi alle specifiche esigenze e alle direttive europee in materia di protezione dei dati personali. Inoltre, le aziende tecnologiche straniere potrebbero non essere disposte a condividere le risorse e le tecnologie con le autorità europee, rendendo difficile per l'Europa sviluppare la propria indipendenza tecnologica.

È importante notare che molti dei modelli di intelligenza artificiale disponibili oggi sono stati sviluppati principalmente per la lingua inglese o adattati dalla lingua inglese per altre lingue. La quantità di modelli di intelligenza artificiale disponibili per le lingue europee è generalmente molto inferiore rispetto a quella disponibile per la lingua inglese¹. Ciò può rappresentare una sfida per la ricerca e lo sviluppo di nuovi modelli di linguaggio personalizzati per le lingue europee, riducendo il potenziale di sviluppo tecnologico e limitando la capacità dell'Europa di competere a livello globale nel campo dell'IA.

Per questo motivi, diventa ancora più importante investire nella creazione di risorse di intelligenza artificiale interne per le lingue europee, al fine di fornire gli strumenti necessari per la ricerca e lo sviluppo di nuovi modelli di linguaggio e garantire la competitività dell'Europa nel campo dell'Intelligenza Artificiale. Ad esem-

¹ Ad aprile 2023, sulla piattaforma huggingface.co, sono presenti 14.819 modelli disponibili per la lingua Inglese, 618 (4%) per la lingua Italiana, 1.303 (9%) per la lingua francese.

Indipendenza tecnologica e diversità culturale nell'intelligenza artificiale europea

Moreno La Quatra
Università degli studi di Enna "Kore"

pio, il progetto europeo *European Language Grid* (ELG)² si propone di creare una piattaforma centralizzata per la collezione di dati linguistici provenienti da varie nazioni e lingue europee. Una strategia univoca per l'utilizzo concreto di questi dati consentirebbe all'Europa di sviluppare modelli di linguaggio personalizzati che soddisfino le esigenze specifiche della popolazione europea e siano conformi alle direttive europee in materia di protezione dei dati personali.

Oltre a risolvere i problemi legati alla dipendenza tecnologica, lo sviluppo di risorse d'IA autonome e indipendenti porterebbe anche a preservare la diversità culturale e linguistica europea. L'utilizzo passivo di modelli IA e il loro addestramento su corpora non curati potrebbe causare il rischio di omogeneizzazione linguistica e culturale, minacciando la varietà linguistica europea. L'accesso a risorse indipendenti e autonome potrebbe promuovere la salvaguardia della diversità culturale e la valorizzazione delle lingue minoritarie europee, che altrimenti potrebbero essere trascurate dalle tecnologie sviluppate da istituzioni esterne all'Europa.

L'importanza della diversità linguistica nell'intelligenza artificiale europea

La diversità linguistica e culturale dell'Europa rappresenta una sfida unica per lo sviluppo di soluzioni di intelligenza artificiale. In Europa, infatti, vi sono almeno 24 lingue ufficiali riconosciute dall'Unione Europea, oltre a molte altre lingue minoritarie e regionali. Ciò significa che le soluzioni IA utilizzate in Europa devono essere in grado di comprendere e utilizzare una vasta gamma di lingue europee, al fine di garantire un'equa rappresentazione della varietà linguistica.

È importante sottolineare che molte lingue parlate nell'Unione Europea hanno caratteristiche specifiche che non sono presenti o rilevanti in altre lingue, come ad esempio l'inclusione di genere nel francese e nell'italiano. Per affrontare questa sfida, è necessario sviluppare risorse di intelligenza artificiale specifiche per le lingue europee che tengano conto delle loro caratteristiche linguistiche (La Quatra, Cagliari, 2022; Sarti, Nissim 2022; Martin *et al.* 2020). In questo contesto, è importante evidenziare come l'inclusività nel linguaggio non si limiti solo alla gestione del genere grammaticale, ma riguarda anche la rappresentazione di gruppi e minoranze specifiche come le persone con disabilità, le

² <https://live.european-language-grid.eu/>

persone LGBTQ+ e le persone di colore. L'uso di soluzioni basate su modelli di lingua inclusivi e adattati alle specifiche linguistiche delle lingue europee può contribuire a promuovere la diversità e l'inclusione in tutti i contesti, dalla comunicazione aziendale all'istruzione e alla politica (Attanasio *et al.* 2021; Attanasio *et al.* 2022).

Lo sviluppo di modelli d'intelligenza artificiale europei rappresenta una soluzione a questa sfida multilingue, consentendo la creazione di soluzioni altamente personalizzate e adattate alle esigenze linguistiche specifiche della popolazione europea. I modelli di linguaggio sarebbero in grado di analizzare, riconoscere e generare testo che preservi le sfumature linguistiche e culturali specifiche di ogni regione e paese.

Corpora non curati: rischi e sfide nei modelli di linguaggio

Per creare modelli di linguaggio efficaci, è necessario che siano addestrati su una grande quantità di dati (Raffel *et al.* 2020). Dopo essere stati addestrati, i modelli di linguaggio naturale dimostrano una notevole capacità di generalizzazione e sono in grado di gestire efficacemente un'ampia varietà di compiti. In questo scenario, i dati utilizzati per l'addestramento sono spesso ottenuti da risorse *web* non curate, come ad esempio i social media, i *blog* e i siti *web*.

L'uso di corpora non curati comporta una serie di rischi. In primo luogo, i dati non curati possono contenere *bias* intrinseci, che possono propagarsi all'interno dei modelli stessi. Questi *bias* possono riguardare questioni di genere, razza, etnia o altro, e possono portare a risultati discriminatori e imprevedibili a priori (Dodge *et al.* 2021). I corpora non curati possono contenere linguaggio d'odio o altri tipi di contenuti inappropriati, che possono essere incorporati nei modelli di linguaggio. In secondo luogo, l'uso di corpora non curati non consente di controllare la varietà o la regione di provenienza del testo utilizzato durante l'addestramento. Questo può portare a modelli di linguaggio che non riflettono la varietà linguistica e culturale richiesta in un contesto europeo.

Per affrontare questi rischi si rende necessario l'utilizzo di risorse di riferimento curate e specifiche per ciascuna lingua e cultura. Ciò richiede un notevole sforzo di raccolta e cura dei dati che può portare a modelli di intelligenza artificiale più inclusivi e conformi all'utilizzo nei contesti europei. Inoltre, l'utilizzo di corpora curati può contribuire alla riduzione di *bias* intrinseci e portare alla prevenzione della diffusione di contenuti inappropriati (Meade *et al.* 2022).

L'intelligenza artificiale europea: sfida e opportunità per il futuro

In sintesi, l'investimento nella creazione di risorse di intelligenza artificiale europee rappresenta un passo cruciale per mantenere l'indipendenza tecnologica dell'Europa e garantire il progresso continuo in questo campo in rapida evoluzione. Tuttavia, per raggiungere tale obiettivo, è necessario impegnarsi a lungo termine nello sviluppo di competenze tecnologiche e nella definizione di politiche europee che garantiscano la massima protezione della *privacy* dei cittadini e la qualità dei dati utilizzati per addestrare i modelli.

In uno scenario europeo, la qualità dei dati utilizzati per addestrare i modelli di intelligenza artificiale è un altro aspetto cruciale. È importante che i dati siano rappresentativi e inclusivi, per preservare la varietà linguistica ed evitare il rischio di discriminazione. La collaborazione tra università e governi è fondamentale per lo sviluppo dell'intelligenza artificiale europea. Le università svolgono un ruolo cruciale nello sviluppo di competenze tecnologiche avanzate, mentre i governi possono promuovere politiche che incentivino la ricerca e lo sviluppo dell'IA, nonché la sua adozione in diversi settori industriali.

Bibliografia

Attanasio Giuseppe, Greco Salvatore, La Quatra Moreno, Cagliero Luca, Tonti Michela, Cerquitelli Tania, Raus Rachele (2022). "L'analyse du discours et l'intelligence artificielle pour réaliser une écriture inclusive: le projet E-MIMIC. *SHS Web of Conferences*, Vol. 138, 01007, 1-15. DOI: <https://doi.org/10.1051/shsconf/202213801007>

Attanasio Giuseppe, Greco Salvatore, La Quatra Moreno, Cagliero Luca, Tonti Michela, Cerquitelli Tania, Raus Rachele (2021). "E-MIMIC: Empowering Multilingual Inclusive Communication". In: *2021 IEEE International Conference on Big Data (Big Data)*. 4227-4234. DOI: [10.1109/BigData52589.2021.9671868](https://doi.org/10.1109/BigData52589.2021.9671868)

Raffel Colin, Shazeer Noam, Roberts Adam, Lee Katherine, Narang Sharan, Matena Micheal, Zhou Yanqj, Li Wei, Liu Peter J. (2020), "Exploring the limits of transfer learning with a unified text-to-text transformer". *Journal of Machine Learning Research*, Vol. 21,1-67.

Dodge Jesse, Maarten Sap, Marasović Ana, Agnew William, Ilharco Gabriel, Groeneveld Dirk, Mitchell Margaret, Gardner Matt (2021), "Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Online and Punta Cana: Association for Computational Linguistics, 1286-1305. DOI: [10.18653/v1/2021.emnlp-main.98](https://doi.org/10.18653/v1/2021.emnlp-main.98)

La Quatra Moreno, Cagliero Luca (2023). "BART-IT: An Efficient Sequence-to-Sequence Model for Italian Text Summarization". *Future Internet*, 15, 15. DOI: <https://dx.doi.org/10.3390/fi15010015>

Louis Martin, Muller Benjamin, Ortiz Suárez Pedro Javier, Dupont Yoann, Romary Laurent, de la Clergerie Éric, Seddah Djamé, Sagot Benoit (2020). "CamemBERT: a Tasty French Language Model". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, 7203–7219. DOI: 10.18653/v1/2020.acl-main.645

Meade Nicholas, Poole-Dayana Elinor, Reddy Siva (2022). "An Empirical Survey of the Effectiveness of Debiasing Techniques for Pre-trained Language Models". In: *Vol.1: Long Papers di Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Dublin: Association for Computational Linguistics, 1878–1898.

Sarti Gabriele, Nissim Malvina (2022). "IT5: Large-scale Text-to-text Pretraining for Italian Language Understanding and Generation". arXiv preprint [arXiv:2203.03759](https://arxiv.org/abs/2203.03759).



R4

Titolo azione proposta

La trascrizione automatica a servizio e a tutela della diatopia linguistica nell'UE

Azione proposta

Nella prospettiva didattico-formativa, formare docenti e discenti, a livello scolastico (istruzione secondaria) e universitario, che siano in grado di far revisionare (docenti) e revisionare (discenti) a livello madrelingua, ma anche di lingua straniera o di lingua seconda, per una lingua “minoritaria” o “ultramminoritaria” dell'UE, documenti audiovisivi trascritti automaticamente in tale lingua relativi alla comunicazione politico-istituzionale, in modo da permettere a questa utenza di sviluppare una maggiore consapevolezza della lingua interessata e affinché ne siano sviluppati o implementati corpora orali e trascritti.

Scopi

1. tutela della diatopia linguistica e del multilinguismo a livello europeo attraverso la creazione o, laddove già esistenti, l'implementazione di corpora delle lingue “minoritarie” e “ultramminoritarie” volti a preservarle nonché a essere utilizzati come strumenti di indagine linguistica, da utilizzare;
2. ai fini dello sviluppo di *software* di trascrizione automatica;
3. e come punti di partenza per discenti per riflettere sulle specificità di tali lingue che emergono da una trascrizione automatica di documenti audiovisivi politico-istituzionali e per risolvere, in *post-editing*, problemi legati all'aspetto morfosintattico e semantico.

Sviluppo dell'azione proposta

Sull'esempio di una lingua “minoritaria” dell'UE quale il basco (Sarasola *et al.* 2023) — adottiamo le denominazioni di “lingua minoritaria” e “lingua ultraminoritaria” proposte da Agresti (2023) —, le tecnologie linguistiche nell'ambito dell'intelligenza artificiale (IA) possono rivelarsi un volano per la rivitalizzazione di lingue il cui repertorio diminuisce a causa della contrazione del numero di soggetti che le parlano in modo attivo o passivo e in contesti più o meno istituzionali e formali. Tra i fattori che permettono di contrastare questo fenomeno, la formazione di corpo docente esperto nell'insegnamento delle lingue “minoritarie” o “ultramminoritarie” a livello madrelingua o di lingua seconda o di lingua straniera, in grado di insegnarle a livello scolastico e universitario, è una risorsa fondamentale (si pensi all'insegnamento a livello scolastico e universitario, nel contesto della Francia, ad

**Investire nello
sviluppo di
tecnologie
“made in UE”**

Alida Maria Silletti

Università di Bari

esempio, dell'occitano (Verny 2009), insieme allo sviluppo di corpora generalisti e specializzati che scaturiscono dall'uso della lingua in determinati contesti (es. l'uso di queste lingue nei media, nell'amministrazione o nella comunicazione politico-istituzionale). Qualsiasi azione in tal senso comporta investimenti economici importanti ed è noto, come rileva Vetere (2023), che una tendenza maggioritaria in ottica di investimenti in ambito linguistico e a livello europeo consiste nell'"anglicizzazione della vita linguistica europea" e, di conseguenza, in una riduzione della diversità, oltre che linguistica, anche culturale (Fischer e Pulaczewska 2009). Se Vetere (2023) identifica in maggiori investimenti nell'ambito della traduzione automatica uno strumento efficace per preservare il multilinguismo, la proposta che avanziamo per difendere la diatopia linguistica e il multilinguismo dell'UE, in particolare per le lingue "minoritarie" e "ultramioritarie", e per sistematizzarne l'uso in contesti politico-istituzionali sfrutta le potenzialità della trascrizione automatica. Questa declinazione dell'IA, che si basa su un processo organizzato in tre tappe — analisi acustica; associazione di frequenze sonore con parole; analisi della parola ottenuta attraverso un modello basato sul linguaggio, un modello basato sulla pronuncia e uno sulla fonetica —, oltre a richiedere grandi risorse economiche, necessita anche di un'enorme quantità di dati affinché vengano identificate le stringhe di parole più probabili pronunciate da un soggetto locutore all'interno di un segnale sonoro¹. Se, pertanto, il risultato è la restituzione, da parte del sistema, del discorso più probabile sulla base dei dati precedentemente immessi, è evidente non solo che maggiori sono i dati a disposizione e minore sarà, statisticamente, la probabilità di errore nella restituzione — e quindi migliore sarà la resa per la lingua inglese, che possiede i *software* più potenziati (Vetere 2023) — ma anche che investire su questo strumento di IA per le lingue "minoritarie" e "ultramioritarie" rappresenta un enorme potenziale. Infatti, sul modello di una sperimentazione condotta in Italia, in aula, su discenti universitari di livello intermedio della lingua francese come lingua straniera (Silletti 2022), emerge che l'esercizio di revisione di una trascrizione automatica a partire da un documento audiovisivo che ha come oggetto la comunicazione politico-istituzionale — ovvero un orale più "controllato" della conversazione spontanea — comporta lo sviluppo di competenze legate alla prosodia della lingua e alla trascrizione che richiedono un lavoro rilevante anche a livello grammaticale (morfosintassi della frase e semantica). L'auspicio è pertanto far sì che anche le lingue "minoritarie"

¹ <https://www.journaldunet.fr/web-tech/guide-de-l-intelligence-artificielle/1501849-reconnaissance-vocale/>

e “ultraminoritarie” dell’UE siano dotate di strumenti che permettano di realizzare *software* di trascrizione automatica o di potenziarli, laddove già esistenti. Questo consentirebbe di arricchire il repertorio linguistico disponibile per quella specifica lingua a partire da un uso autentico della stessa. Tutto ciò richiede investimenti importanti che la rete europea META (Vetere 2023) ma anche soggetti investitori a livello regionale o locale dovrebbero attuare l fine di dotare l’Unione europea di una rete di tecnologie “*made in UE*” che tuteli anche la diatopia linguistica dell’UE.

Bibliografia

Agresti Giovanni (2023). « Intelligence artificielle et langues minoritaires : du bon ménage? Quelques pistes de réflexion ». In: Rachele Raus (cur.) et al. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l’aune de l’intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell’era dell’intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 47-68. <https://www.collane.unito.it/oa/items/show/132>.

Crochet-Damais Antoine (2022). “Reconnaissance vocale : définition, algorithmes et fonctionnement”. *Journal du Net*, 31/05/2022. <https://www.journaldunet.fr/web-tech/guide-de-l-intelligence-artificielle/1501849-reconnaissance-vocale/>

Fischer Roswitha, Pulaczewska Hanna (a cura di) (2008). *Anglicisms in Europe: Linguistic Diversity in a Global Context*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Sarasola Kepa, Aldabe Itziar, Aranberri Nora (2023). “Enabling additional official languages in the EU for 2025 with language-centred AI”. In: Rachele Raus (cur.) et al. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l’aune de l’intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell’era dell’intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 91-105. <https://www.collane.unito.it/oa/items/show/132>.

Silletti Alida Maria (2022). « La macrosyntaxe à l’épreuve de la transcription générée automatiquement : le cas des ‘parenthèses’ » intervento al Convegno internazionale *Franc’parler. Français parlé: données. Représentatins, questionnements théoriques*, Università di Torino: 16-17 giugno 2022.

Verny Marie-Jeanne (2009). “Enseigner l’occitan au XXIe siècle. Défis et enjeux”. *Trema*, 31, 69-83. DOI: <https://doi.org/10.4000/trema.962>.

Vetere Guido (2023). "Elaborazione automatica dei linguaggi diversi dall'inglese: introduzione, stato dell'arte e prospettive". In: Rachele Raus et al. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l'aune de l'intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell'era dell'intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 69-87. <https://www.collane.unito.it/oa/items/show/132>

La corsa all'Intelligenza Artificiale e i suoi pochi partecipanti

Recenti studi sulle forme odierne di Intelligenza Artificiale (IA) generativa¹ mostrano indiscutibilmente che la qualità del risultato, sia esso un testo o un'immagine, è spesso legata agli investimenti a monte per la creazione dell'IA stessa (Bowman 2023: 1). La necessità di grandi investimenti dedicati e, al contempo, di competenti centri di ricerca e sviluppo, ha lasciato lo sviluppo di IA moderne a completo appannaggio di poche aziende, colossi nel mercato della tecnologia, nessuno dei quali ha sede legale nell'Unione Europea².

La “corsa all'Intelligenza Artificiale” a trazione privata — e, ormai, esplicitamente orientata alla costruzione di un prodotto di consumo — costituisce insieme un rischio e un grosso incentivo per la costruzione di nuove IA da enti accreditati nell'UE. Il resto di questo contributo analizza due macro aree che costituiscono un rischio per l'UE, ovvero la sempre minore trasparenza e i “*bias*” sociali che l'IA rischia di perpetuare, e conclude con una lista di possibilità e sfide incombenti.

I rischi di una IA poco trasparente

Guidati da — seppur legittimi — criteri a tutela di un possibile vantaggio economico, *tutti* i produttori delle IA generative più diffuse non rilasciano né informazioni chiave sullo sviluppo, come i dati usati per l'addestramento, né i modelli stessi, che sono, invece, accessibili dietro interfacce a pagamento. Questa pratica pone una serie di rischi per l'UE e i Paesi membri, per una serie di motivi.

Le linee guida europee per l'*Intelligenza Artificiale sicura*³ sottolineano l'importanza della trasparenza⁴ e richiedono che sia possibile misurare parametri quali la presenza di concetti stereotipati o la sostenibilità ambientale. La quasi totale assenza di dettagli sullo sviluppo e l'uso delle IA odierne rende impossibile

¹ Con IA “generativa” si intende, genericamente, una tecnologia che può generare dei nuovi contenuti, quasi sempre inediti, a partire da una forma di richiesta in ingresso. Ne sono un esempio GPT-3 per il linguaggio e DALL-E per le immagini, a cui è possibile chiedere “un sonetto su un tordo nello stile di Dante”, e “un dipinto della Torre Eiffel nello stile de La Notte Stellata” di Picasso, rispettivamente. Sia GPT-3 che DALL-E sono sviluppate dall'azienda statunitense OpenAI.

² I recenti ChatGPT, DALL-E, Bing, Bard, LLaMA, Claude sono tutti sviluppati da aziende private, tutte con sede legale negli Stati Uniti.

³ <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1.html>

⁴ Ad esempio, deve essere possibile “spiegare” ad un essere umano il processo che ha portato un algoritmo ad una certa decisione.

Verso un'intelligenza artificiale europea e trasparente

Giuseppe Attanasio
Università Bocconi

assolvere alle suddette linee guida. È impossibile, ad esempio, impiegare tecniche del campo dell'*Explainable AI* ("IA spiegabile"), branca di ricerca ormai consolidata, che richiede l'accesso al modello per attribuire un dato *output* ad un dato *input*. Ugualmente, è difficile stimare l'impatto ambientale in quanto non sono note né la dimensione del modello — correlata positivamente con la CO₂ emessa per l'addestramento (Strubell *et al.* 2019: 2) — né la quantità e il tipo di dati di addestramento. Infine, l'accesso limitato ostacola le nuove proposte di "*watermarking*", ovvero tecniche per verificare formalmente se un contenuto è stato prodotto da una IA⁵, che richiedono accesso parziale o totale al modello e alle sue funzionalità, e che sono cruciali per la verificabilità.

Più genericamente, i modelli "chiusi" non sono ispezionabili e l'intera catena di verifica resta in testa all'azienda proprietaria — che non sempre aderisce alle legislazioni europee vigenti⁶ — violando il principio di verificabilità indipendente delle linee guida europee. Verificabilità e validazione sono di estrema importanza tanto più questi modelli si dimostrano fragili e pericolosi: esistono già prove di come le IA testuali possano mostrarsi aggressive verso gli utenti⁷, fornire false informazioni⁸, o il cui comportamento sia facilmente modificabile con attacchi che fanno leva su specifiche formulazioni delle richieste⁹.

Le aggiunte complessità di questi modelli rendono impossibile assicurare che l'IA sia *sicura*: non esistono, ad oggi, metodi affidabili che permettano di "forzare" l'IA a seguire principi o linee guida (Bowman 2023: 1). Ne segue che lo sviluppo di IA europee diventa fondamentale per invertire questa tendenza. Nuovi investimenti faciliteranno nuove ricerche sulla validazione e la sicurezza delle IA, favorendo tanto la valutazione di modelli esistenti quanto la costruzione di nuovi modelli "*made in EU*", i cui tutti dettagli — a partire dai dati, per finire alle componenti del modello — siano liberamente accessibili ad enti di ricerca e partner industriali.

Ideologie e stereotipi che l'IA perpetua

I dati usati per l'addestramento costituiscono un tema altresì impellente. Il rischio, in questo caso, è duplice.

⁵ <https://aiguide.substack.com/p/on-detecting-whether-text-was-generated>

⁶ <https://www.independent.co.uk/tech/chatgpt-ban-italy-gdpr-data-protection-b2311738.html>

⁷ <https://time.com/6256529/bing-openai-chatgpt-danger-alignment/>

⁸ <https://cybernews.com/tech/chatgpts-bard-ai-answers-hallucination/>

⁹ <https://www.washingtonpost.com/technology/2023/02/14/chatgpt-dan-jailbreak/>

Da un lato, il tipo di dati e la loro provenienza possono essere problematici. È risaputo che la prima fonte di dati di addestramento sia internet. Il *web*, però, non rispecchia la totalità delle visioni del globo, ma solo quelle di chi ne può fruire l'accesso più o meno costante (Bender *et al.* 2021: 3). Perciò le visioni, le ideologie — e gli stereotipi — di una certa fetta di popolazione possono essere apprese e perpetuate da uno strumento, l'IA in questione, che raggiunge molte più persone. Ad esempio, studi recenti dimostrano che le IA che generano immagini associano la figura dell'Amministratore Delegato ad un uomo, o quella di un terrorista ai tratti mediorientali (Bianchi *et al.* 2023: 4).

D'altro canto, *quasi tutti* i corpora linguistici allo stato dell'arte — cioè sufficientemente estesi per addestrare una IA per il linguaggio — sono in lingua inglese.

Costruire IA anglocentriche, però, limita intrinsecamente le possibilità delle stesse su altre lingue in molteplici casi. Nozza *et al.* (2022: 5), ad esempio, mostrano come modelli italo-inglesi, cioè addestrati su una combinazione di dati in Italiano e Inglese, permettono di identificare il linguaggio d'odio più efficacemente rispetto alle loro controparti monolingua. L'inglese, inoltre, non condivide le peculiarità di molte delle lingue europee. Ne è un esempio l'inflessione del genere nella grammatica delle lingue romanze: è un tema, infatti, lo studio del rapporto tra genere usato e posizioni lavorative (Stanovsky *et al.* 2019: 6), o nuove attenzioni sull'uso di un linguaggio o una traduzione inclusiva (Attanasio *et al.*: 7, Piergentili *et al.* 2023: 8).

La costruzione di corpora sotto la supervisione di enti accreditati nell'UE ha un duplice obiettivo. Da un lato, è necessario fornire risorse in cui è possibile verificare la qualità e i valori che vi sono codificati, misurarne e mitigarne gli stereotipi prima che vengano utilizzati per l'addestramento. Dall'altro, e più genericamente, incrementare la presenza delle lingue degli stati membri, bilanciando il rapporto con l'inglese, per catturare peculiarità linguistiche altrimenti dimenticate.

Sfide incumbenti

Nuovi investimenti nel campo dell'intelligenza artificiale da parte dell'Unione europea sono necessari per invertire la tendenza che vede come principale risultato dei modelli non trasparenti, non sicuri, e che danno priorità alla sola lingua inglese.

Gli incentivi sono molteplici, e passano dalle certezze che studi recenti sul rapporto costi-benefici, dal supporto nelle intenzioni

di altri enti indipendenti¹⁰, e, soprattutto, sulla necessità di costruire IA sicure, secondo le lingue guida stilate, peraltro, dalla stessa Unione europea.

Bibliografia

Attanasio Giuseppe, Greco Salvatore, La Quatra Moreno, Cagliero Luca, Tonti Michela, Cerquitelli Tania, Raus Rachele (2021). "E-MIMIC: Empowering Multilingual Inclusive Communication". In: *2021 IEEE International Conference on Big Data (Big Data)*. 4227-4234. DOI: 10.1109/BigData52589.2021.9671868

Bender Emily M., Gebru Timmit, McMillan-Major Aangelina, Shmitchell Shmargaret (2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?". In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. New York: Association for Computing Machinery, 610-23. DOI: <https://doi.org/10.1145/3442188.3445922>

Bianchi Federico, Kalluri Pratyusha, Durmus Esin, Ladhak Faisal, Cheng Myra, Nozza Debora, Hashimoto Tatsunori, Jurafsky Dan, Zou James, Caliskan Aylin (2022). "Easily accessible text-to-image generation amplifies demographic stereotypes at large scale". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. New York: Association for Computing Machinery, 1493-1504. DOI: <https://doi.org/10.1145/3593013.3594095>

Bowman Samuel R. (2023), "Eight Things to Know about Large Language Models". *arXiv preprint arXiv:2304.00612*.

Nozza Debora, Bianchi Federico, Attanasio Giuseppe (2022). "HATE-ITA: hate speech detection in Italian social media text". In: *Proceedings of the Sixth Workshop on Online Abuse and Harms (WOAH)*. Seattle: Association for Computational Linguistics, 252-260. DOI: 10.18653/v1/2022.woah-1.24

Piergentili Andrea, Fucci Dennis, Savoldi Beatrice, Bentivogli Luisa, Negri Matteo (2023). "From Inclusive Language to Gender-Neutral Machine Translation". <https://arxiv.org/abs/2301.10075>

Stanovsky Gabriel, Smith Noah A., Zettlemoyer Luke (2021). "Evaluating gender bias in machine translation". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence: Association for Computational Linguistics, 1679-1684. DOI: 10.18653/v1/P19-1164

Strubell Emma, Ganesh Ananya, McCallum Andrew (2019). "Energy and Policy Considerations for Deep Learning in NLP". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence: Association for Computational Linguistics, 3645-3650. DOI: <https://doi.org/10.18653/v1/P19-1355>

¹⁰ <https://laion.ai/notes/letter-to-the-eu-parliament/>

Per fronteggiare crisi ed emergenze di carattere nazionale o internazionale, avere a disposizione strumenti tecnologici per diffondere rapidamente informazioni utili alla popolazione interessata è una parte fondamentale della strategia per mitigare il danno, prestare soccorsi in maniera adeguata, o semplicemente anticipare ed informare il pubblico per prevenire danni peggiori.

In una società fortemente interconnessa a livello internazionale, superare le barriere linguistiche e culturali è spesso una necessità in tempo di crisi, e quindi strumenti di traduzione automatica appositamente adattati possono essere di grande aiuto in situazioni del genere, ma occorre una adeguata preparazione a monte in termini di raccolta ed organizzazione di informazioni, creazione di uno o più corpora multilingui, e competenze scientifiche e tecniche per la creazione o adattamento di sistemi di traduzione automatica.

Nel caso del terremoto ad Haiti nel 2010, il governo locale era in grave difficoltà nel fornire aiuto alla popolazione, e le risorse internazionali a disposizione, governative e non-governative, avevano il problema di riuscire a comunicare nella lingua locale (Creolo Haitiano). Attraverso una collaborazione volontaria tra enti accademici e aziende private, venne realizzato dapprima un sistema di traduzione umana per lo smistamento e triage dei messaggi di richieste d'aiuto via SMS, e nel giro di un'altra settimana, un sistema di traduzione automatica gratuito, grazie all'utilizzo di un corpus Creolo haitiano-inglese creato da ricercatori delle varie organizzazioni e gli stessi traduttori volontari coinvolti nel sistema iniziale.

Il successo dell'operazione ha portato alla stesura di raccomandazioni per creare risorse simili in situazioni di crisi (Lewis, Munro, Vogel 2011: 501).

Ci troviamo sempre più spesso ad affrontare crisi o multi-crisi di livello internazionale, dove non c'è un singolo evento localizzato (terremoto o altro disastro naturale, attentato terroristico), ma una serie di eventi che partono da una regione o nazione specifica, e coinvolgono un numero sempre più ampio di paesi, raggiungendo potenzialmente ogni angolo del globo. È il caso della pandemia del virus COVID-19 nel 2020, o della multi-crisi umanitaria ed energetica causata dalla invasione russa in Ucraina nel 2022.

In queste situazioni, gli effetti della crisi non sono immediati a livello internazionale, e gli obiettivi principali sono la diffusione di informazioni certe e verificate per il coordinamento della ricerca, del commercio, della logistica, e del trattamento dell'informazione in generale.

Proposta di creazione di un gruppo di lavoro dell'Unione europea per l'elaborazione di corpora multilingui e multimodali in risposta a situazioni di multi-crisi

Federico Garcea
Università di Bologna

Nel caso del COVID-19, si trattava di rendere disponibili le informazioni in possesso dell'Organizzazione Mondiale della Sanità in lingue “minori”, ovvero lingue non incluse tra quelle in cui sono tradotti con rapidità i nuovi contenuti via via pubblicati dall'OMS, e necessarie per raggiungere centinaia di milioni di individui in Africa e Asia. Di questo sforzo si è occupato il team virtuale di TICO-19 (*Translation Initiative for COvid-19*), che ha creato un corpus in 35 lingue diverse con le informazioni conosciute sul virus al momento della stesura, includendo strategie di mitigazione e contenimento, terapie conosciute e la loro risposta osservata, ecc. (Anastasopoulos *et al.*, 2020).

Questa risorsa è uno strumento fondamentale sia per i traduttori umani, come memoria di traduzione, sia per sistemi automatici e i loro ricercatori, permettendo di adattare facilmente i sistemi esistenti con l'adeguata terminologia e informazioni coerenti e puntuali.

Si tratta di una risorsa utile anche a contrastare la diffusione di notizie false e strumentalizzazioni a fini politici o criminali, poiché rende possibile verificare l'attendibilità delle informazioni utilizzando la traduzione assistita o automatica.

Si sottolinea che questo tipo di risorse è forse ancora più importante oggi, al tempo della diffusione degli LLM (*Large Language Models*) e GPT (*Generative Pre-trained Transformer*), che sono in grado di generare informazioni in un linguaggio fluido e coerente, ma non necessariamente fattuale: le risorse possono essere consumate dai sistemi di GPT per aggiornarsi, e per enti governativi e non per creare delle applicazioni simil-GPT in base alle nuove risorse per informare direttamente il pubblico e rispondere ai loro quesiti in maniera automatica o semi-automatica.

Negli Stati Uniti, l'Università di Washington (UW) ha fondato insieme ad altri enti accademici il gruppo *Language Technologies for Crisis Preparedness and Response* (LT4CPR) per proseguire il lavoro in questo campo e coordinare al meglio gli sforzi nelle prossime crisi.

Sosteniamo che sarebbe utile la creazione e sviluppo di un gruppo di lavoro coordinato a livello europeo che sia in grado di gestire le stesse problematiche, curando la creazione e mantenimento di corpora multilingui sui soggetti più rilevanti per la popolazione europea in occasione di crisi umanitarie o economiche. Occorre creare e gestire risorse in tutte le lingue ufficiali della EU, ed inoltre includere le lingue delle minoranze linguistiche riconosciute nell'Unione, ed infine considerare le lingue più diffuse tra le comunità internazionali dei residenti della EU (es. turco, arabo, ucraino).

Da sottolineare anche l'importanza della creazione di corpora multimodali, ovvero che incorporino informazioni di tipo testuale ed audiovisivo. La sola informazione testuale non è sufficiente a gestire il volume e modalità di comunicazione spesso necessaria per raggiungere tutte le fasce della popolazione.

Ci sono situazioni in cui operatori sanitari, sociali, umanitari devono comunicare faccia a faccia (di persona o virtualmente), e la impossibilità di usufruire di sistemi di riconoscimento e sintesi vocale o comunque di traduzione automatica multimodale è spesso un fattore limitante.

La disponibilità di risorse multilingue e multimodali a livello europeo, rispettosi della diatopia linguistica, rappresenterebbe uno strumento efficace per migliorare la qualità e velocità dell'informazione sanitaria, legale, sociale, economica tra tutti i cittadini e residenti dell'Unione Europea, e potrebbe letteralmente consentire di salvare un numero maggiore di vite in scenari di crisi, o comunque di mitigare le conseguenze sulla popolazione, soprattutto delle sue fasce più fragili.

Bibliografia

Anastasopoulos Antonios, Cattelan Alessandro, Dou Zi_Yi, Federico Marcello, Federmann Christian, Genzel Dmitriy, Guzmán Francisco, Hu Junjie, Macduff Hughues, Koehn Philippe, Lazar Rosie, Lewis Will, Neubig Graham, Niu Mengmeng, Öktem Alp, Paquin Eric, Tang Grace, Tur Sylwia (2020). "TICO-19: the Translation Initiative for COvid-19". In: *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*. Association for Computational Linguistics. <https://aclanthology.org/2020.nlpcovid19-2.5/>

Lewis William, Munro Robert, Vogel Stephan (2011). "Crisis MT: Developing A Cookbook for MT in Crisis Situations". In: *Proceedings of the Sixth Workshop on Statistical Machine Translation*. Edinburgo: Association for Computational Linguistics, 501–511. <https://aclanthology.org/W11-2164/>



R4

La disponibilità di risorse terminologiche plurilingui, autentiche, attuali e affidabili costituisce un patrimonio di inestimabile valore, utile per diverse applicazioni che vanno dalla traduzione e dall'interpretariato all'analisi terminologica per la trasmissione delle conoscenze. In questo contesto, la Rete Panlatina di Terminologia REALITER¹ promuove attività di ricerca e applicate ai fini della diffusione e dello sviluppo delle terminologie specialistiche nell'area romanza (Gilardoni 2011, Zanola 2014). In più di trent'anni di attività, numerose sono state le tematiche trattate all'interno di lessici plurilingue, dall'informatica al commercio, dalle biotecnologie alla medicina, dalla moda allo sport. Oltre ai traduttori specialistici, i lessici si rivolgono sia a diverse figure professionali, quali esperti della disciplina, consulenti linguistici, giornalisti, revisori, redattori dei testi, ecc., sia a un pubblico più generale desideroso di informarsi sul corretto utilizzo di determinati termini in una prospettiva di comunicazione plurilingue (Calvi 2020). In particolare, i lavori dei lessici REALITER si basano su precisi principi metodologici, al fine di "favorire lo sviluppo armonico delle lingue neolatine, tenendo conto della loro origine comune"²: tutte le lingue della Rete, comprese le loro varianti, hanno la stessa importanza all'interno dei progetti. Da un punto di vista teorico/ metodologico sono adottati gli approcci variazionista e sistemico; la Rete incoraggia la collaborazione per la realizzazione dei lavori terminologici, non solo tra i terminologi di lingue e paesi diversi, ma anche tra gli esperti di settore di interesse. Infine, per garantire l'elaborazione di prodotti di qualità, si rispettano i criteri di accessibilità, attualità e affidabilità (Principi metodologici del lavoro terminologico 2000, Zanola 2012).

Per rendere più accessibili i lessici plurilingui di REALITER, fondamentale è la collaborazione con CLARIN-IT (Infrastruttura Comune Italiana per le Risorse e le Tecnologie Linguistiche), avviata nel 2021. In tale contesto, i lessici REALITER sono trasformati in formati che rispettano i principi FAIR del *web* semantico, secondo i quali i dati devono essere rintracciabili, accessibili, interoperabili e riutilizzabili (Cimiano 2020). I lessici e i loro metadati sono inseriti nella collezione "REALITER - OTPL"³, per poi essere caricati sulla piattaforma terminologica di *Social Sciences and Humanities for the European Open Science Cloud* – SSHOC⁴, strumento per la condivisione dei dati e di conseguenza della conoscenza.

Risorse terminologiche plurilingui nel rispetto dei principi FAIR del web semantico

Silvia Calvi
Klara Dankova
Lucrezia Marzo
Maria Teresa Zanola
*Università Cattolica
del Sacro Cuore, Milano*

¹ <https://www.realiter.net/>

² <https://www.realiter.net/presentazione/regolamento>

³ <https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/000-c0-111/565>

⁴ <https://sshopencloud.eu/>

Investire in lavori terminologici di questa natura significa raccogliere un patrimonio scientifico e culturale che potrà sicuramente migliorare l'efficacia e l'efficienza delle applicazioni dell'intelligenza artificiale. Nel dominio della traduzione automatica e assistita, queste risorse potranno essere integrate all'interno di sistemi automatici per assicurare un lavoro di qualità.

Bibliografia

Calvi Silvia (2020). "Il lessico dell'arrampicata sportiva: metodologia per la progettazione ed elaborazione di un lessico plurilingue". In: Manuel Célio Conceição, Maria Teresa Zanola (a cura di). *Terminologia e mediação linguística: métodos, práticas e atividades*. Universidade do Algarve, 287-301. <http://hdl.handle.net/10400.1/15043>

Cimiano Philipp et al. (2020). *Linguistic Linked Data. Representation, Generation and Applications*. Berlino: Springer. DOI: <https://doi.org/10.1007/978-3-030-30225-2>

Gilardoni Silvia (2011). "I lessici della Rete Panlatina di Terminologia". In: Maria Teresa Zanola, Francesca Bonadonna (a cura di). *Terminologie specialistiche e prodotti terminologici*. Milano: EDUCatt, 101-112.

Monachini Monica, Frontini Francesca (2016). "CLARIN, l'infrastruttura europea delle risorse linguistiche per le scienze umane e sociali e il suo network italiano CLARIN-IT". *Italian Journal of Computational Linguistics*, 2, 11-30. DOI: <https://doi.org/10.4000/ijcol.387>

REALITER. *Principi metodologici del lavoro terminologico*. <http://www.realiter.net/wp-content/uploads/2013/06/Principi-metodologici-del-lavoro-terminologico.pdf>

Zanola Maria Teresa (2014). "Le réseau Realiter, un acteur du plurilinguisme". *Plaisance*, 11, 149-165.

Zanola Maria Teresa (2012). *Costruire un glossario: la terminologia dei sistemi fotovoltaici*. Milano: Vita e Pensiero.

Zanola Maria Teresa, Villa Maria Luisa, Dankova Klara (2023). "Langages et savoirs : intelligence artificielle et traduction automatique dans la communication scientifique". In: Rachele Raus (cur.) et al. *De Europa Special Issue. Multilinguisme et variétés linguistiques en Europe à l'aune de l'intelligence artificielle, Multilinguismo e variazioni linguistiche in Europa nell'era dell'intelligenza artificiale, Multilingualism and Language Varieties in Europe in the Age of Artificial Intelligence*. Torino, Milano: Università degli Studi di Torino, Ledizioni LediPublishing, 107-127. <https://www.collane.unito.it/oa/items/show/132>.

Le reti neurali profonde, soprattutto l'architettura *Transformer* (Vaswani *et al.*, 2017), hanno permesso di far progredire notevolmente la traduzione automatica (Sutskever *et al.*, 2014; Bahdanau *et al.*, 2016). Molteplici servizi basati su tale tecnologia permettono di finalizzare traduzioni di buona qualità che tuttavia restano spesso letterali (Bhardwaj *et al.* 2020), possono contenere controsensi od omissioni e sono meno pertinenti in alcuni ambiti specifici, come ad esempio i settori finanziario e automobilistico.

Per migliorare la traduzione automatica a fini professionali, sarebbe fondamentale elaborare un quadro di riferimento europeo comune più adatto che annoveri metriche orientate alle pratiche, dati pertinenti e assolutamente multisetoriali. Ciò implica il coinvolgimento di vari attori della traduzione: ricercatori nelle TALN e in traduttologia, società che forniscono *software* e dispositivi di traduzione, nonché traduttori e servizi di traduzione. Realizzare ciò a livello europeo diventa essenziale se si vuole puntare a una strategia di successo sia per ridurre le disuguaglianze tecnologiche che differenziano le lingue europee¹ sia, soprattutto, per permettere all'Europa di fungere da capofila per l'integrazione delle tecnologie in ambito professionale.

In tal senso, è utile presentare l'attuale situazione della valutazione della traduzione automatica, come le campagne WMT², che si concentrano sul valutare le tecnologie della traduzione e la qualità delle traduzioni, o quelle IWSLT³ che riguardano la traduzione dell'oralità (interpretazione). Nell'ambito delle attività del WMT, i dati sono annotati da professionisti della traduzione che utilizzano la tassonomia MQM al fine d'identificare gli errori tipici della traduzione. Tra i loro compiti c'è anche quello di valutare il lavoro di *post-editing* e di predire se una traduzione contiene i cosiddetti errori "catastrofici".

Malgrado tali sforzi, resta difficile valutare i risultati delle attività svolte. È noto che alcuni sistemi vanno oltre il semplice rilevamento degli errori, tuttavia pochi o nessuno li analizzano. Non sorprende, perciò, se alcuni ricercatori (Marie *et al.* 2021) hanno osservato che, nell'analizzare le pubblicazioni principali sulla traduzione automatica tra il 2010 e il 2020, si continua ad utilizzare il punteggio BLEU (Papineni *et al.* 2002) per contare le sequenze di parole in comune tra la traduzione automatica e la traduzione di riferimento.

Valutare la qualità di una traduzione di un documento intero resta un'attività poco diffusa (Specia *et al.* 2020; Zerva *et al.*

Per un quadro di riferimento europeo comune di valutazione delle tecnologie della traduzione basate sull'IA

Philippe Langlais

RALI, DIRO, Università di Montréal

François Yvon

Sorbonne Université e CNRS

¹ Cfr. il progetto *European Language Equality* al link https://european-language-equality.eu/wp-content/uploads/2022/11/ELE___Deliverable_D3_4__SRIIA_and_Roadmap___final_version_-1.pdf

² <http://www2.statmt.org/wmt23/>

³ <https://iwslt.org>

2022), sebbene sia qualcosa di estremamente utile in ambito professionale.

Infine, ci sono troppi pochi studi sulla gestione delle Memorie di Traduzione (MdT) e sul loro uso nel processo traduttivo, che pure sono strumenti essenziali a livello professionale.

Quanto ai dati aggregati e resi disponibili per addestrare e testare i sistemi di apprendimento profondo, essi scaturiscono dai dibattiti parlamentari europei (Koehn 2005), da quelli dell'ONU (Ziemski *et al.* 2016) e dai corpora paralleli estratti dal *web* (Espla *et al.* 2019). Ci sono poi dei dati specifici che sono stati valutati in passato per dei settori particolari, per l'ambito mediatico, ecc.

Tuttavia, malgrado l'abbondanza di tali dati (almeno per alcune coppie linguistiche), non si presta attenzione, ad esempio, a disaggregare in modo pertinente i corpora per l'addestramento e quelli per i test. Questi ultimi sono spesso infarciti di frasi ricche di stereotipi ed estremamente ripetitive, peraltro già presenti nei primi. Ciò rischia di "contaminare" i test e ha ripercussioni sulla valutazione dei risultati che diventa ottimistica. Ci sono ancora troppi pochi studi sulla valutazione degli esempi su cui si addestrano i modelli e in tal senso la quantità va a discapito della qualità. Al contrario, selezionare i dati su criteri specifici permetterebbe di addestrare modelli più robusti e di assemblare dati più congrui per effettuare i test (cfr. ad esempio, il caso virtuoso di Varshney *et al.* 2022).

Va anche considerato che il documento resta un elemento secondario nell'organizzare i dati e che i corpora sono solitamente segmentati in frasi equivalenti per coppie linguistiche (i cosiddetti corpora allineati). Ciò rende difficile rendere coeso il testo tradotto, dato che è la frase l'unità di base.

Più in generale, sviluppare un sistema unico, capace di trattare di vari ambiti simultaneamente, è una questione che sta diventando sempre più fondamentale ma che ancora non è stata sufficientemente presa in considerazione e che, perciò, resta attualmente senza soluzione (Pham *et al.* 2021). Nella maggior parte dei casi, gli studi in tale ambito sono condotti in settori molto diversi tra loro (biomedico, finanziario, tecnologico) e in numero ridotto, non potendo quindi restituire la ricchezza degli ambiti che sono presi in considerazione dai servizi di traduzione. Ad esempio, si è constatato (Frenetti 2021) che un sistema di traduzione neurale generico aveva difficoltà a tradurre testi in alcuni settori presi in considerazione dal *Bureau de la traduction du Canada*, ovvero dai servizi di traduzione utilizzati dal Governo canadese, e che in tal senso a nulla sono serviti i tentativi tecnici di fornire al sistema ulteriori informazioni su quei settori.

Riassumendo, è possibile affermare che, nonostante gli indiscussi progressi compiuti nella traduzione automatica, gli attuali quadri di riferimento per valutare tale traduzione restano insufficienti e vanno ripensati affinché sia possibile sviluppare tecnologie più utili e funzionali ai fini professionali. Ciò richiede un maggior lavoro sulla preparazione dei dati e delle annotazioni, nonché elaborare metriche rappresentative e sviluppare nuove tecnologie (traduzione interattiva e/o con l'ausilio di MdT, dispositivi che permettano di gestire i flussi traduttivi, ecc.).

Elaborare un quadro comune europeo di valutazione è certamente un progetto ambizioso, che richiede peraltro uno sforzo sinergico dei diversi attori del mondo della traduzione, ma anche una sfida che l'Europa, forte del proprio multilinguismo, è certamente in grado di affrontare.

Bibliografia

Bahdanau Dzmitry, Cho KyungHyun, Bengi Yoshua (2014). "Neural Machine Translation by Jointly Learning to Align and Translate". *ICLR*. <https://arxiv.org/pdf/1409.0473.pdf>

Bhardwaj Shivendra, Hermelo David Alfonso, Langlais Phillippe, Bernier-Colborne Gabriel, Goutte Cyril, Simard Michel (2021). "Human or Neural Translation?". In: *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona: International Committee on Computational Linguistics, 6553–6564. DOI: 10.18653/v1/2020.coling-main.576

Devlin Jacob, Chang Ming-Wei, Lee Kenton, Toutanova Kristina (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *Volume 1 (Long and Short Papers) di Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Minneapolis: Association for Computational Linguistics, 4171–4186. DOI: 10.18653/v1/N19-1423

Espla Miquel et al. (2019). "ParaCrawl : Web-scale parallel corpora for the languages of the EU". In: *Proceedings of Machine Translation Summit XVII: Translator, Project and User Tracks*. Dublino: European Association for Machine Translation, 118–119. <https://aclanthology.org/W19-6721.pdf>

Frenette Xavier (2021). *Utilisation du plongement du domaine pour l'adaptation non supervisée en traduction automatique*. Tesi di laurea magistrale, Università di Montréal. DOI: <https://doi.org/1866/26528>

Koehn Philipp (2005). "Europarl: A parallel corpus for statistical machine translation". In: *Proceedings of Machine Translation Summit X: Papers*. <https://aclanthology.org/2005.mtsummit-papers.11.pdf>

Marie Benjamin, Fujita Atsushi, Rubino Raphael (2021). “Scientific Credibility of Machine Translation Research: A Meta-Evaluation of 769 Papers”. In: *Vol. 1: Long Papers. Proceedings of the joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language*. Association for Computational Linguistics, 7297–7306. <https://aclanthology.org/2021.acl-long.566.pdf>

Papineni Kishore, Roukos Salim, Ward Todd, Zhu Wei-Jing (2002). “Bleu: a Method for Automatic Evaluation of Machine Translation”. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311-318. <https://doi.org/10.3115/1073083.1073135>

Pham MinhQuang et al. (2021). “Revisiting Multi-Domain Machine Translation”. *Transactions of the Association for Computational Linguistics*, 9:17–35. DOI: 10.1162/tacL_a_00351

Specia Lucia et al. (2020). “Findings of the WMT 2020 Shared Task on Quality Estimation”. In: *Proceedings of the Fifth Conference on Machine Translation*. Association for Computational Linguistics, 743-764. <https://aclanthology.org/2020.wmt-1.79.pdf>

Sutskever Ilya, Vinyals Orion, Le Quoc V. (2014). “Sequence to sequence learning with neural networks”. In: *Advances in neural information processing systems*. <https://proceedings.neurips.cc/paper/2014/file/a14ac55a4f27472c5d894ec1c3c743d2-Paper.pdf>

Varshney Neeraj et al. (2022). “ILDAE : Instance-Level Difficulty Analysis of Evaluation Data”. In: *Volume 1: Long Papers : Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Dublin: Association for Computational Linguistics, 3412-3425. DOI: 10.18653/v1/2022.acl-long.240

Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, Gomez Aidan N., Kaiser Lukasz, Polosukhin Illia (2017). “Attention is All you Need”. In: *Advances in Neural Information Processing Systems*. DOI: <https://doi.org/10.48550/arXiv.1706.03762>

Zerva Chrysoula et al. (2022). “Findings of the WMT 2022 Shared Task on Quality Estimation”. In: *Proceedings of the Seventh Conference on Machine Translation (WMT)*. Abu Dhabi: Association for Computational Linguistics, 69-99. <https://aclanthology.org/2022.wmt-1.3.pdf>

Ziemski Michal et al. (2016). “The United Nations Parallel Corpus v1.0”. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portoroz: European Language Resources Association (ELRA), 3530-3534. <https://aclanthology.org/L16-1561.pdf>

Allegato

Centre National de la Recherche scientifique — CNRS

REALITER — Rete panlatina di terminologia

Kedge Business School

Sorbonne Université

Universidad del País Vasco

Universidade Estadual de Campinas

Università Bocconi

Università Cattolica del Sacro Cuore — Milano

Università degli Studi di Enna “Kore”

Università degli Studi di Napoli “Federico II”

Università di Bari

Università di Bergamo

Università di Bologna

Università di Catania

Università di Genova

Università di Modena e di Reggio Emilia

Università di Napoli “Parthenope”

Università di Roma 2

Università di Torino

Università L'Orientale di Napoli

Università Vita-Salute San Raffaele

Université Bordeaux Montaigne

Université Catholique de Lille

Université de Montréal

Université de Paris 12

Université de Toulouse

Université Lumière Lyon II,

Université Mohammed V — Rabat

Université Paris Cité

Université Paris 8

University of Geneva

University of South Australia

Vrije Universiteit Brussel

**Lista delle
Università e degli
enti di ricerca che
hanno collaborato
in forme diverse,
tramite il loro
personale, alle
ricerche condotte
dal *Panel* sui diritti
linguistici e l'IA del
Centro d'eccellenza
Jean Monnet
*Artificial
Intelligence for
European
Integration***



GLOSSARIO

Glossario

Algoritmo: in informatica, successione di calcoli che permettono di risolvere un problema.

Apprendimento profondo (*deep learning*): campo di ricerca dell'intelligenza artificiale che permette a dei neuroni artificiali di elaborare delle informazioni ai fini all'apprendimento da parte delle reti neurali.

Bias: forme di distorsione della realtà, errori che possono produrre veri e propri pregiudizi.

Corpus (*plur. Corpora*): selezione di dati opportunamente effettuata ai fini della ricerca sulla base di criteri specifici.

Diatopia linguistica: il variare delle lingue rispetto ai territori dove sono parlate in virtù di elementi sociolinguistici. Ne è un esempio la variazione regionale dell'italiano o le lingue francesi parlate in varie zone geografiche come la Francia, il Belgio, la Svizzera...

Industria delle lingue: l'insieme dei prodotti, tecniche, attività o servizi che richiedono il trattamento automatico della lingua naturale. Un esempio può essere la produzione di dispositivi di traduzione automatica, di interpretazione automatica, ecc.

Intelligenza artificiale (IA): scienza che si propone di sviluppare sistemi informatici intelligenti, riproducendo i processi mentali umani; più generalmente, per metonimia, rinvia anche ai modelli informatici e ai dispositivi materiali che si basano sull'apprendimento profondo (*deep learning*).

Large language model: modello di linguaggio che prevede l'apprendimento profondo automatizzato, ovvero non supervisionato da personale umano, o semiautomatizzato e basato su grandi quantità di dati. Si tratta quindi di un sistema avanzato d'intelligenza artificiale che sfrutta una grandissima quantità di dati per arrivare a riprodurre e generare il linguaggio umano, come nell'esempio di *ChatGPT*.

Lingua minoritaria: lingue utilizzate all'interno di un determinato territorio di uno Stato da gruppi di cittadini numericamente più piccolo rispetto al resto della popolazione. Rispetto a questa definizione, parafrasata dalla *Carta europea delle lingue regionali e minoritarie* adottata dal Consiglio d'Europa nel 1992, va precisato che nel presente report, come precisato da Agresti, tale definizione può indicare anche le lingue nazionali in ragione del loro minor utilizzo in determinati contesti rispetto ad altre lingue.

Modello di linguaggio o modello linguistico: uso di tecniche di calcolo statistico probabilistico effettuato da reti neurali appositamente addestrate in modo automatizzato o semiautomatizzato per predire l'utilizzo di una o più parole in una frase.

Modello informatico: insieme di meccanismi di astrazione che mirano a descrivere la struttura della conoscenza concreta. La realtà fattuale viene perciò rappresentata in modo astratto tramite dei linguaggi formali (informatici).

Modello informatico supervisionato: modello informatico che permette l'apprendimento dei dispositivi d'IA in modo da prevedere l'intervento umano.

Modello informatico non supervisionato: modello informatico che permette l'apprendimento dei dispositivi d'IA in modo automatico, senza l'intervento umano.

Modello matematico: insieme di meccanismi di astrazione che mirano a rappresentare in modo quantitativo i fenomeni naturali.

Neuroni artificiali (reti di): modelli matematici che imitano le funzioni del cervello umano. Le reti si compongono di unità di base (neuroni), che sono semplici unità di calcolo.



Il sostegno della Commissione europea alla produzione di questa pubblicazione non costituisce un'approvazione del contenuto, che riflette esclusivamente il punto di vista degli autori, e la Commissione non può essere ritenuta responsabile per l'uso che può essere fatto delle informazioni ivi contenute.

Notes

Note



Quest'opera è distribuita con
Licenza Creative Commons Attribuzione.
Condividi allo stesso modo 4.0 Internazionale.
Copyright © 2023

Finito di stampare nel mese di novembre 2023 da
Rotomail SpA - Vignate (MI)

Per un'intelligenza artificiale a favore del multilinguismo europeo

*Raccomandazioni
strategiche rivolte ai
decisori europei*

a cura di
Rachele Raus

€ 29,00 |



**Artificial Intelligence
for European Integration**
Jean Monnet Centre of Excellence



9 791256 000036 >