

Troppe sfumature, con il dialetto lombardo Google Translate non ce la fa

Pubblicato: Mercoledì 24 Luglio 2024



Alla fine di giugno Google aveva annunciato l'obiettivo di aggiungere 110 nuove lingue al servizio di traduzione Google Translate, **tra le quali anche alcuni tra i principali dialetti italiani, compreso il lombardo.**

I risultati, per ora, non sembrano incoraggianti, secondo un report presentato oggi dai ricercatori del progetto **AlpiLinK**. Il traduttore non riuscirebbe infatti a cogliere le tantissime sfumature dei dialetti, che risentono molto delle variazioni territoriali.

Restando al dialetto lombardo, ad esempio, **Google Translate non riesce a tradurre “ragazza”** e taglia corto utilizzando la stessa parola dell'italiano, mentre dimentica completamente *tosa*, utilizzato nel milanese, o il bresciano *s'céta*.

L'aggiornamento di Google translate – con l'inserimento di lombardo, veneto, friulano, ligure, siciliano – è dei giorni scorsi, e il team di **AlpiLinK** – **macroprogetto per la mappatura digitale dei dialetti del Nord Italia** che vede come capofila l'Università di Verona – ha subito testato la traduzione basata sull'intelligenza artificiale e sul modello linguistico conversazionale PaLM2 lanciato da Google.

«Rispetto all'italiano il lombardo dimostra una variazione linguistica molto forte da territorio a territorio: certe parole e costruzioni milanesi non vengono riconosciute a Bergamo o Brescia e viceversa

– spiega **Stefan Rabanus**, docente di linguistica tedesca all’ateneo di Verona e coordinatore scientifico di AlpiLinK – I casi che abbiamo raccolto mostrano come la recente introduzione di alcuni dialetti italiani nell’aggiornamento di Google Translate , pur rappresentando di per sé una buona notizia perché **riconosce il valore dei dialetti e delle lingue minoritarie**, sia soggetta a **molti limiti**. A differenza delle lingue nazionali, **denominazioni come “lombardo” non fanno riferimento ad una lingua unitaria ma ad una molteplicità di dialetti diversi** presenti nello stesso territorio e questo rende ancor più vulnerabile il traduttore, che necessiterebbe di un corpus molto più ampio e diversificato e la possibilità di specificare il tipo di “lombardo” per non incorrere in falle ed errori».

Se i test con singole parole evidenziano dei problemi **non va meglio – con qualche eccezione – nella traduzione di frasi**: tu dormi più di lui restituita da Google risulta come *tu durmi püsee de lü*. In realtà la forma verbale durmi in molti dialetti lombardi non esiste perché la desinenza della seconda persona singolare finisce in -t – dormet in questo caso. È se il pronome soggetto tu è espresso in italiano come in tu dormi più di lui, i parlanti dei dialetti lombardi tendono a “sdoppiarlo” e la frase viene resa con due pronomi soggetto, un forte e uno clitico (debole), ad esempio *ti te dormet püsee de lü* a Milano o *te ta dormet (de) più de lü* a Brescia.

Un altro limite è rappresentato dal fatto che, mentre per l’italiano Translate consente anche il riconoscimento vocale, **nel caso del dialetto non è possibile ascoltare audio**.

Proprio sulla raccolta di contributi audio è basato invece il progetto AlpiLinK. **Tutte le persone che parlano un dialetto possono partecipare direttamente alla ricerca attraverso il sito di AlpiLinK – alpilink.it** – compilando in poco tempo l’audio-sondaggio dedicato, in cui viene proposto all’utente di utilizzare il proprio dialetto o la propria lingua locale per descrivere cosa accade in una scena o per tradurre le parole o frasi indicate, tra cui il sopra citato tu dormi più di lui. Un modello di ricerca partecipativa, che punta sul coinvolgimento “dal basso” e che sta raccogliendo un notevole interesse.

«Un corpus significativo – spiega Rabanus – perché fra il progetto AlpiLinK e il progetto precedente Vinko **sono già quasi 2500 le persone che hanno partecipato al progetto** per un totale di circa **225 mila file audio**. Un database open che abbiamo utilizzato anche per questa analisi dell’attendibilità dei risultati di Google Translate e che permette di cogliere appunto le diverse varietà di una stessa area dialettale, andando a verificare in modo puntuale, per ogni Comune, i termini utilizzati con la possibilità di ascoltare direttamente gli audio. **Un grande lavoro di mappatura che forse, in futuro, si potrà rivelare utile anche per allenare modelli di intelligenza artificiale** e contribuire a superare i limiti attuali».

L’invito a contribuire al progetto attraverso il sito alpilink.it è aperto.

Google translate ora traduce anche il dialetto lombardo

Redazione VareseNews
redazione@varesenews.it