

100 MILIONI DI PAROLE TRADOTTE CON IL TRADUTTORE DELLA PRESIDENZA DEL CONSIGLIO UE

19.11.2020 | Lingua e comprensione del testo |
Tecnologie linguistiche e multilinguismo | Saarbrücken
Intervista con il Prof. Dr. Josef van Genabith



EU Council Presidency Translator

„Der EU Council Presidency Translator ist eine sehr europäische Lösung, die zeigt, dass Europa gemeinsam im Bereich Sprachtechnologie und KI international auf höchstem Level mehr als konkurrenzfähig ist“

www.presidencymt.eu

Powered by DeepL DFK Connecting Europe eTranslation TILDE

Nota: versione italiana tradotta da un sistema Tilde e post-editata dall'intervista originariamente tedesca, generata dall'EUCPT.

- Prof. van Genabith, Lei è Direttore scientifico presso il DFKI e dirige dal 2014 l'unità di ricerca sulle Tecnologie Linguistiche Multilingue (MLT) a Saarbrücken. Qual'è stata la Sua carriera scientifica prima di trasferirsi a Saarbrücken?

Il successo del Traduttore della Presidenza del Consiglio dell'UE è una grande occasione per il nostro [team MLT](#) e per i nostri partner [DeepL](#), [Tilde](#) e [eTranslation](#)! Sono molto orgoglioso delle squadre e del lavoro che hanno svolto in stretta collaborazione con il Ministero degli Esteri tedesco! Io stesso mi interesso da molto tempo di linguaggio e tecnologia, ho studiato ingegneria elettronica e lingua e letteratura inglese alla RWTH di Aachen ed in seguito ho avuto molta fortuna: grazie ad una borsa di studio del British Council e successivamente del British Foreign & Commonwealth Office, ho potuto prima fare un master all'Università dell'Essex e poi conseguire il dottorato di ricerca sotto la guida di Louisa Sadler. All'inizio degli anni Novanta poi ho lavorato come PostDoc con Hans Kamp presso l'Istituto per l'elaborazione del linguaggio naturale (IMS) di Stoccarda. È stato un periodo bellissimo! Dopo sono stato 17 anni in Irlanda presso la School of Computing della Dublin City University, dove sono avanzato da docente a docente senior e professore associato. A Dublino avevo molta libertà ed ottimi colleghi e colleghe sia alla DCU che in altre università di Dublino e nelle tante aziende di high-tech con sede in Irlanda (IBM, Microsoft, Symantec), e siamo riusciti a sfruttare questa libertà: ho ricostituito il Denas National Center for Language Technology (NCLT) e sono stato il direttore fondatore del CNGL (Center for Next Generation Localisation, ora ADAPT e gestito da Vinny Wade). Grazie a questo lavoro, soprattutto alla CNGL, siamo stati coinvolti sempre più spesso in progetti internazionali, ad esempio dell'UE, nei quali era molto attivo il precedente direttore della nostra unità di Saarbrücken, Hans Uszkoreit. Tramite Hans Uszkoreit, che nel frattempo aveva costituito

***Light post-editing** implies minimal intervention by the post-editor to make the text understandable. Grammar, punctuation and spelling are correct, the translation is complete and accurate in content, but not necessarily idiomatic and fluent.

un'unità gemella a Berlino (ora SLT, guidato da Sebastian Möller), sono arrivato a Saarbrücken e al DFKI nel 2014, dopo 17 anni in Irlanda.

- Oltre al Suo lavoro al DFKI, ha anche una cattedra all'Università di Saarland. Come si completano il lavoro accademico e quello orientato all'applicazione?

La parte più importante del nostro lavoro sono i nostri collaboratori: è grazie a loro se il nostro lavoro ha così successo! I miei collaboratori all'università e al DFKI lavorano insieme in una varietà di squadre. Nei nostri meeting settimanali non importa chi è al DFKI o all'università. Facciamo parte del [dipartimento 1102](#) (Information Density and Linguistic Encoding) dell'università; all'università abbiamo un progetto per il Consiglio tedesco di ricerca sul post-editing multimodale, nel quale collaboriamo con grande successo con il team del Prof. Antonio Krüger al DFKI; sono a capo del programma europeo di Master in Tecnologie del Linguaggio e della Comunicazione ([LCT](#), Erasmus+), che è gestito egregiamente da uno dei miei dirigenti presso l'unità [MLT](#) (al DFKI) tramite un contratto part-time all'università. I team leader dei quattro gruppi dell'unità MLT – Machine Translation, Question Answering and Information Extraction, Talking Robots e Data and Resources – insegnano, tengono seminari e seguono gli studenti di PhD, MSc e BSc. Allo stesso modo, molti collaboratori dei gruppi MLT sono attivi presso l'università. Il tutto, formalmente e finanziariamente, è ovviamente separato rigorosamente in progetti. Ma il legame con l'università è molto forte. Il dipartimento "[Scienze e tecnologie del linguaggio](#)" dell'Università di Saarland è uno dei migliori in Europa. Noi dell'unità MLT del DFKI siamo molto forti nel campo della ricerca: nel 2020, ad esempio, abbiamo pubblicato più di 10 articoli in occasione delle più importanti conferenze internazionali del nostro settore (ACL, ICML, EMNLP, COLING, IJCAI) nel campo della tecnologia linguistica, dell'intelligenza artificiale e dell'apprendimento automatico. Questo è un grande successo e dimostra la qualità delle squadre. Dall'altro lato, la ricerca del DFKI orientata alle applicazioni è molto attraente per gli studenti ed i ricercatori dell'università: dove altro il proprio lavoro viene utilizzato in modo così visibile come per es. nel Traduttore della Presidenza del Consiglio dell'UE, dove nel giro di 4,5 mesi (ad oggi) sono state tradotte 100 milioni di parole? Direi che è fantastico.

- Il traduttore della Presidenza del Consiglio dell'UE ha promosso ulteriormente la visibilità dei servizi di traduzione automatica in Germania. Si tratta del lavoro congiunto di diversi attori, ma Lei ha guidato questo progetto. Quando ha iniziato a lavorarci? Come ha messo insieme il consorzio? E quanti scienziati erano coinvolti?

Il [Traduttore della Presidenza del Consiglio dell'UE](#) è una soluzione molto europea che dimostra che l'Europa unita è estremamente competitiva a livello internazionale nel campo della tecnologia linguistica e dell'IA: si basa su una combinazione di eccezionali competenze nel campo high-tech e IA in Germania (DeepL, DFKI), Lettonia (Tilde) e nella CE (eTranslation). Una partnership tra industria (DeepL, Tilde), settore pubblico (CE, eTranslation) ed un istituto di ricerca (DFKI). Il DFKI è a capo del progetto, il finanziamento è fornito dal Ministero degli Esteri tedesco, che gestisce la Presidenza tedesca del Consiglio dell'UE. Le competenze dei membri del consorzio si completano a vicenda in modo ideale: Tilde ha programmato per anni, con finanziamenti europei, la struttura di base del Traduttore della Presidenza, nella quale sono integrati i motori di traduzione di diversi fornitori, e contribuisce con i propri motori di traduzione. DeepL offre motori di traduzione di qualità eccezionale per 8 lingue. eTranslation (della CE) fornisce un servizio di traduzione automatica di base per tutte le 24 lingue ufficiali dell'UE. In stretta collaborazione con gli addetti alla traduzione dei ministeri, il DFKI ha sviluppato sistemi di traduzione automatica per il tedesco, il francese e lo spagnolo, adattati appositamente ai dati e alle esigenze dei ministeri. Tilde fa lo stesso per l'inglese, l'italiano e il polacco. Al DFKI, Stephan Busemann fornisce il supporto amministrativo al Traduttore della Presidenza. Io gestisco gli aspetti scientifici e tecnici. Cristina España Bonet, capo della squadra di Machine Translation dell'unità MLT e la sua collaboratrice Jingyi Zhang sviluppano i sistemi. Sono coadiuvate da due studentesse, Damyana Gateva e Anastasija Amman, del programma di laurea

magistrale "Scienze e tecnologie linguistiche" dell'università. Il DFKI gestisce anche il lavoro mediatico e di sensibilizzazione del Traduttore della Presidenza. Il tutto è supervisionato da Eileen Schnur e dalla sua collega Marlies Thönnissen nel team MLT ed è attivamente supportato dal dipartimento di Corporate Communications del DFKI.

- [Utilizzate reti neurali artificiali per la traduzione. Può descriverci il funzionamento della sua macchina di traduzione?](#)

Negli ultimi anni, i modelli neurali hanno permesso di fare salti quantici nella qualità di molte tecnologie linguistiche e di altre applicazioni nell'IA. I nostri sistemi utilizzano reti neurali profonde basate su modelli di trasformatori. Questi modelli utilizzano diversi tipi di attenzione e sono in gran parte altamente parallelizzabili.

- [Le reti neurali artificiali vengono costruite e testate con grandi quantità di dati linguistici. Da dove provengono questi dati e, per dare solo una stima, di quante parole si tratta?](#)

Per molte coppie linguistiche, i nostri dati di training sono costituiti da decine di milioni di coppie di frasi, in cui ogni coppia di frasi contiene una frase di partenza in una lingua e la sua traduzione nell'altra lingua. Da questo, i motori imparano a tradurre da soli. Questi dati si basano su traduzioni già effettuate da umani. Il motore impara quindi dagli umani. I dati provengono da raccolte di dati dell'UE, dall'ELRC (European Language Resource Coordination, anche questo gestito al MLT del DFKI) e da altre fonti. Inoltre, lavoriamo a stretto contatto con i team di traduzione dei ministeri per creare motori speciali che utilizzano i dati dei ministeri, appositamente progettati per soddisfare le esigenze dei ministeri. Questi vengono costantemente valutati dai traduttori dei ministeri, in modo da poterli migliorare continuamente nel corso del progetto.

- [Negli ultimi 150 giorni, il Traduttore della Presidenza è stato impiegato molto spesso dagli utenti. Sono state tradotte oltre 100 milioni di parole. Quali erano le coppie linguistiche più popolari? E ci sono state forse anche frasi tradotte con particolare frequenza?](#)

A differenza di altre offerte, il Traduttore della Presidenza è sicuro e protetto, tutti i server si trovano nell'UE, le trasmissioni sono criptate e tutti i dati vengono cancellati immediatamente dopo il completamento di una traduzione. Quindi abbiamo solo informazioni superficiali riguardo l'uso. Le cifre mostrano che la traduzione one-click dal sito web in lingua tedesca della Presidenza è stata accolta molto bene: da lì proviene circa il 47% dei 100 milioni di parole tradotte finora. Le lingue di destinazione preferite per la traduzione automatica del sito della Presidenza sono lo spagnolo, l'italiano e il portoghese (le versioni in francese e inglese sono state create manualmente). La parte leggermente maggiore risulta dalla traduzione di testi (22%), documenti (30%) e siti web (2%) sulla pagina del Traduttore, e qui la traduzione tra tedesco e inglese è quella più richiesta.

- [Cosa dicono i traduttori dei nuovi livelli della traduzione automatica? I traduttori vedono i motori di traduzione come concorrenti o come strumenti di supporto al loro lavoro? E come sta cambiando l'immagine professionale del traduttore?](#)

Nel progetto "EU Council Presidency Translator", lavoriamo a stretto contatto con i nostri colleghi dello staff di traduzione dei ministeri: sono loro a gestire la raccolta e la preparazione di dati all'interno dei ministeri al fine di adattare i motori a loro dedicati alle esigenze dei ministeri. Inoltre, testano e valutano i motori e con i loro risultati contribuiscono in modo decisivo al miglioramento dei sistemi. Nel flusso di lavoro di traduzione poi, sono un aiuto: se la qualità della traduzione è buona, il motore può contribuire ad aumentare la produttività di un traduttore umano. In questo senso, il profilo professionale del traduttore si dirige verso il controllo di qualità, la garanzia di qualità tramite il post-editing (la correzione) delle traduzioni prodotte automaticamente e la certificazione delle traduzioni e della loro qualità. La formazione dei traduttori oggi giorno tiene conto di questi cambiamenti: il corso accademico di traduzione "Scienza e tecnologia della traduzione"

dell'Università di Saarland ha elevato contenuto tecnologico, nel quale i futuri traduttori acquisiscono familiarità con le tecnologie linguistiche sviluppate dai loro colleghi nei corsi di Linguistica computazionale (Scienze e tecnologie del linguaggio) e di Informatica.

- [La presidenza tedesca del Consiglio dell'UE termina il 31.12.2020. Come verrà utilizzato il Traduttore della Presidenza? E indipendentemente da questo, quali sono i Suoi progetti futuri?](#)

Il Traduttore della Presidenza è stato accolto straordinariamente bene e ha superato tutti i record stabiliti dai precedenti Traduttori della Presidenza. Sono molto orgoglioso di quello che il team MLT del DFKI ha realizzato insieme ai colleghi di DeepL, Tilde e eTranslation! C'è un grande interesse ad utilizzare il Traduttore per le future presidenze UE, le trattative sono già in corso. Inoltre c'è un grande interesse da parte dell'industria per la tecnologia linguistica tedesca ed europea: tecnologia linguistica e IA "made in Europe". La traduzione automatica è solo una delle competenze della nostra unità MLT: altre sono quelle del gruppo "Question-Answering and Information Extraction" (soprattutto in campo biomedico), quelle del gruppo Talking Robots (incentrato sui sistemi di dialogo e sulla robotica di soccorso) e quelle del gruppo "Data and Resources" (che per molti anni ha guidato grandi progetti UE come ELRC). A questo si aggiunge la nostra unità di ricerca gemella SLT (Speech and Language Technology) a Berlino. I due laboratori (MLT a Saarbrücken e SLT a Berlino) lavorano a stretto contatto e si completano a vicenda nelle loro competenze.