

*Tradurre col computer*

Hellmut Riediger

# Tradurre col computer

dalla traduzione automatica al CAT e ritorno

## Sommario

I.	Breve storia delle tecnologie per la traduzione .....	3
1.	Il computer come assistente .....	5
2.	Il ritorno della TA.....	7
II.	Traduzione automatica, sì o no?.....	8
III.	L'uso della traduzione automatica nella pratica professionale .....	11
IV.	Altri usi della TA .....	14
1.	La TA come dizionario o come banca dati terminologica .....	14
2.	La TA per la ricerca di testi paralleli.....	15
3.	La TA per tradurre sottotitoli.....	16
V.	Sistemi di traduzione automatica.....	16
1.	Sistemi basati su regole/ architetture «rule-based» .....	16
2.	Sistemi statistici/Approcci statistici.....	18
3.	Sistemi ibridi.....	20
4.	Sistemi neurali.....	20
VI.	Metodi di valutazione della traduzione automatica .....	21
VII.	Scrivere per la TA – Lingue naturali, sotto-linguaggi, linguaggi controllati .....	23
1.	Sotto-linguaggi.....	23
1.1	Cos'è un sotto-linguaggio?.....	24

1.2	Sotto-linguaggi e TA?.....	24
2.	Linguaggi controllati.....	26
1.3	Vantaggi.....	28
3.	Pre-editing .....	29
	Appendice: Come scrivere per la traduzione automatica (di Lara Bernardi) .....	29
	Bibliografia .....	37

## **I. Breve storia delle tecnologie per la traduzione**

I traduttori, come tutti coloro che scrivono testi, usano il computer, usano programmi di videoscrittura, sbrigano la loro corrispondenza e gestiscono i loro incarichi o progetti via e-mail o attraverso sistemi cloud, navigano e ricercano in rete.

Tuttavia quando si parla di tecnologie informatiche per la traduzione, ci si riferisce in particolare a quei sistemi progettati appositamente per l'attività traduttiva, da quelli che pretendono di sostituire il traduttore (sistemi di traduzione automatica), a quelli che forniscono al traduttore strumenti per lo svolgimento di attività tradizionalmente svolte manualmente, quali vari tipi di analisi sul testo di partenza e/o di arrivo, la creazione, gestione e consultazione di glossari, le ricerche in corpora linguistici e la produzione del testo di arrivo.

Sono inoltre di ausilio alla traduzione, in senso lato, tutti gli altri strumenti informatici e di trattamento automatico della lingua che in qualche modo elaborano il linguaggio umano, come, dizionari elettronici, banche dati terminologiche, strumenti per la ricerca e l'estrazione terminologica, motori e programmi di ricerca ed interfacce di ricerca speciali, programmi di elaborazione di testi, archivi e corpora testuali, strumenti per la ricerca e la sostituzione di testo, strumenti per il riconoscimento e la sintesi vocale, sistemi di riconoscimento ottico dei caratteri (OCR) e molti altri.

HT (Human Translation), CAT (Computer aided Translation), MT (Machine Translation), MAHT (Machine aided human translation) e HAMT (Human aided machine translation) sono denominazioni comunemente impiegate nella traduzione professionale per distinguere, rispettivamente, la Traduzione umana, la Traduzione assistita da computer, la Traduzione

automatica, la Traduzione umana assistita da computer, e la Traduzione automatica assistita. Questi riflettono diversi gradi di intervento da parte dell'essere umano o delle macchine lungo uno spettro continuo di tecnologie per la traduzione che va dalla traduzione interamente umana alla traduzione interamente automatica.

Già nel XVII secolo Cartesio e Leibniz immaginarono che per superare le barriere linguistiche potesse essere possibile realizzare un dizionario meccanico.

Ma per i primi tentativi concreti si dovette attendere il XX secolo.

Nel 1933 vennero registrati i primi brevetti di traduttori automatici, uno del francese Georges Artsrouni e uno del russo Petr Trojanskij. Quello di Artsrouni riguardava un dizionario a schede perforate chiamato «cerveau mécanique», in grado di tradurre in automatico parole di quattro lingue diverse in francese. Anche il brevetto di Trojanskij prevedeva un dizionario elettronico, ma con la funzione aggiuntiva d'interpretazione delle regole grammaticali, attraverso dei simboli universali inseriti in uno strumento di traduzione multilingue.

Negli anni '30 e '40 si gettano le basi della teoria dell'informazione e della crittografia che fece poi registrare notevoli successi durante la seconda guerra mondiale. Le ricerche e i modelli proposti, tra cui spiccano quelli di Alan Turing e Claude Shannon, diffusero la fiducia che fosse possibile sviluppare macchine in grado di eseguire qualunque calcolo risolvibile con un algoritmo. I primi a pensare all'uso del computer per tradurre furono il cristallografo britannico Andrew Booth del Birbeck College di Londra e il matematico americano Warren Weaver, autore con Shannon di una celebre volume sulla teoria matematica della comunicazione. Nel 1948 Booth progettò un dizionario elettronico bilingue. Nel 1949 Warren Weaver, stimolato anche dalle conversazioni e corrispondenze con Booth scrisse un Memorandum, in cui delineò le prospettive della traduzione automatica suscitando l'interesse della ricerca in questo campo negli Stati Uniti.

Nel giugno 1952, Yehoshua Bar-Hillel presiedette la prima conferenza sui traduttori automatici presso il MIT, da cui emerse la necessità dell'intervento umano prima o dopo i processi di traduzione (pre- e post-editing).

Nel 1954, la collaborazione tra la Georgetown University e IBM diede vita alla prima dimostrazione pubblica di un sistema di traduzione automatica. Con un vocabolario di sole 250 parole e sole 6 regole grammaticali fu tradotto in inglese un campione selezionato di frasi russe. Nonostante il suo carattere rudimentale, il sistema suscitò l'interesse della ricerca statunitense e ispirò altri progetti in tutto il mondo. Si diffusero grandi aspettative e molti pensarono che in tempi rapidi i traduttori automatici avrebbero potuto sostituire i traduttori umani.

Nel 1964, negli Stati Uniti fu istituito l'ALPAC (Automatic Language Processing Advisory Committee). Tuttavia già nel 1966, visti gli scarsi risultati ottenuti, la commissione tagliò i fondi alla ricerca. La sua celebre relazione considerò che i sistemi di traduzione automatica erano inutili, costosi e imprecisi e contribuì a determinare un rallentamento della ricerca nel settore.

È probabile che i primi ricercatori della TA avessero sottovalutato la complessità della questione, considerando il problema come puramente matematico e trascurando i già numerosi studi di linguistica riguardanti l'ambiguità lessico-strutturale, la dipendenza contestuale e la risoluzione delle anafore. Questo tipo di mentalità si riassume in una citazione dello stesso Weaver: «When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode'» (Weaver, 1955, p. 18). Inoltre c'è da dire che la linguistica degli anni Cinquanta non forniva delle regole esplicite che si potessero formalizzare in algoritmi. Infatti la ricerca di regole divenne uno degli obiettivi principali delle linguistiche teoriche soltanto alla fine degli anni '50 con l'avvento della linguistica generativa chomskyana. Solo successivamente i linguisti cercarono di classificare le differenze strutturali di costrutti paralleli applicando il concetto di interlingua, che fu coniato per la prima volta nel 1972 dallo psicolinguista americano Larry Selinker (Selinker, 1972) nel definire i processi mentali di chi sta apprendendo una seconda lingua.

Ciononostante in piena Guerra Fredda e sulla scia degli esperimenti della Georgetown, nel 1968 grazie a Peter Toma, nacque Systran, un sistema di traduzione diretta dal russo all'inglese che a partire dal 1970 fu impiegato dall'Aeronautica degli Stati Uniti e successivamente anche dalla NATO, da General Motors e Xerox. Tuttavia, nel complesso la ricerca rallentò e la TA rimase confinata nell'ambito di grandi istituzioni. Nel 1975 la Commissione delle Comunità Europee approvò un piano d'azione per fronteggiare i problemi del multilinguismo della Comunità. Il piano prevedeva tra l'altro l'uso di Systran, lo sviluppo di traduttori automatici solo per determinate lingue (come TITUS) e la progettazione a lungo termine del sistema multilingue di traduzione automatica su larga scala EUROTRA. Nel 1976, in Canada la Montreal University sviluppò un altro sistema operativo chiamato Météo.

### **1. CAT: Il computer come assistente**

Gli sforzi maggiori si concentrarono ora sulla ricerca di altri modi di mettere la tecnologia informatica al servizio del traduttore umano per farle svolgere mansioni che richiedessero minore intelligenza umana. Non più traduzione automatica, quindi, bensì assistenza al traduttore umano nelle attività particolarmente dispendiose e ripetitive, come la ricerca di parole o di frasi già tradotte o la gestione della terminologia.

Le prime soluzioni, negli anni '70, consistettero essenzialmente in banche dati terminologiche, il cui uso, a causa dei costi elevati, era limitato a poche grandi aziende. Tra la fine degli anni '70 e l'inizio degli anni '80 si cominciò a chiedersi se non fosse possibile realizzare programmi in grado di ricordarsi se una parte di un nuovo testo da tradurre era già stata tradotta in precedenza e di visualizzare questa parte in modo automatico insieme al testo già tradotto. Nasceva così il concetto di memoria di traduzione.

Negli anni '80 sistemi del genere vennero effettivamente realizzati per le attività di traduzione interna di alcune grandi aziende. I primi furono TSS sviluppato da ALPS e Text Tools della società

olandese INK, poi commercializzato in Germania dalla società Trados. Diverse altre aziende realizzarono strumenti simili a uso interno.

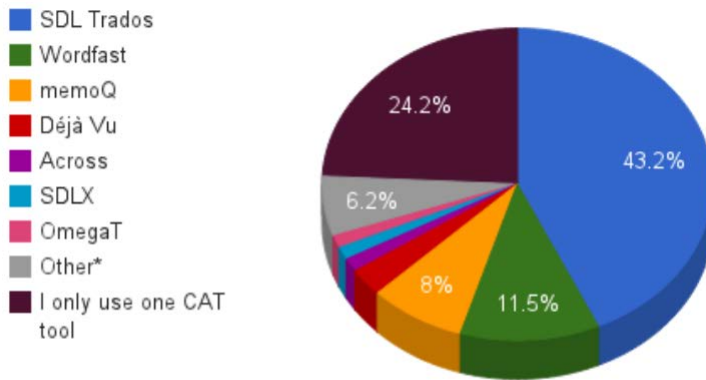
Nella prima metà degli anni '90 tali pacchetti iniziarono a essere venduti anche al grande pubblico. Nel 1992 esordirono sul mercato Translation Manager/2 (TM/2) di IBM e la Translator's Workbench di Trados. L'anno dopo uscì Déjà Vu di Atril e nel 1994 Star mise in commercio il programma Transit. Con la diffusione dei PC si sviluppò anche la necessità della localizzazione del software, per la quale furono sviluppati nuovi specifici programmi. Tuttavia i prezzi alti (nell'ordine delle migliaia di attuali euro) e gli elevati requisiti hardware ostacolarono a lungo la diffusione di questi sistemi fra i traduttori autonomi, finché alla fine degli '90 e il primo decennio del 2000, videro la luce diversi software economici oppure gratuiti come Wordfast (1999), Omega T (2002), Memo Q (2009), Translator's Toolkit di Google (2008), Memsource (2011) e MateCat (2014).

Oggi i sistemi di traduzione assistita sono diventati fondamentali nell'ambito della traduzione professionale, per tutte quelle tipologie testuali caratterizzate da qualche forma di ripetitività (testi tecnici, manuali, istruzioni d'uso, ma anche certi tipi di saggistica) poiché in cosiddette memorie di traduzione registrano (e ripropongono) le coppie di segmenti originali e tradotti da un traduttore umano. Nel corso della traduzione di un nuovo testo la memoria di traduzione viene automaticamente consultata per riutilizzare parti di testo già tradotte precedentemente. Vengono infatti confrontati i nuovi segmenti da tradurre, con quelli precedentemente tradotti e archiviati, cercando tra essi una possibile corrispondenza esatta o anche parziale (fuzzy match). L'applicazione segnala al traduttore con precisione, oltre al grado di somiglianza, le differenze riscontrate. Il traduttore deciderà di volta in volta, se confermare la traduzione trovata nella memoria, riadattarla o scriverne una nuova. In base al contenuto della memoria si può provvedere alla pretraduzione del testo per i segmenti con corrispondenza esatta. Man mano che si lavora con esse, le memorie crescono e diventano un patrimonio di lavoro, di cui avere grande cura. Esistono vari formati di memorizzazione proprietari legati a differenti programmi: per es., TMW è stato il formato tipico di SDL Trados, programma tra i più diffusi. Attualmente per garantire l'interoperabilità e la sopravvivenza nel tempo delle memorie, i programmi più aggiornati supportano il formato standard TMX, basato su XML. Inoltre tutti i sistemi CAT permettono di creare al momento o caricare database terminologici. Un altro grande vantaggio dei sistemi CAT è il fatto che restituiscono il testo con il medesimo layout del testo di partenza.

Tutti i sistemi CAT oggi più usati, come SDL Trados, Star Transit, Across, Atril Déjà Vu, WordFast, Omega T, memoQ e Memosurce comprendono programmi per la gestione di memorie di traduzione, la gestione della terminologia, l'allineamento dei file, la gestione del progetto e sempre più frequentemente sono integrati con sistemi di traduzione automatica.

Specifiche applicazioni come SDL Passolo vengono usate invece per la localizzazione di software.

**Which CAT tool do you use most?**



CAT software most used – grafico da Proz.com

## 2. Il ritorno della TA

Dopo che negli anni '80 furono avviati nuovi progetti, e in concorrenza con Systran nacquero sistemi commerciali come METAL, Globalink, Tovna, entrarono sul mercato la ricerca e i prodotti giapponesi, in particolare per le grandi industrie dell'elettronica di massa.

Negli anni '90, un gruppo di ricercatori dell'IBM pubblicò i risultati degli esperimenti di prova di un sistema basato puramente su dati statistici, e anche in Giappone fu sviluppato un metodo simile. Si trattava di sistemi innovativi rispetto ai precedenti sistemi basati su regole, non tanto per le regole sintattiche o semantiche, ma per il loro metodo di approccio lessicale, basato su corpora, cioè archivi di testi già tradotti.

Alla fine degli anni '90 si registrò una forte crescita delle vendite quando i sistemi, prima molto costosi, furono convertiti in prodotti per PC a prezzi contenuti accessibili al vasto pubblico. La diffusione del Web portò i traduttori automatici on-line. Alla domanda di sistemi per tradurre documenti da pubblicare si affiancò il bisogno di traduzioni immediate per le applicazioni di Internet, come la posta elettronica e i siti internet, per le quali è sufficiente una traduzione che renda il senso generale. Nel 1997, la collaborazione tra Systran e AltaVista (dal 2004 acquisita da Yahoo!) diede vita a BabelFish, il primo sistema di traduzione gratuito on-line con ben diciotto coppie di traduzione. Seguendo questo esempio molti portali cominciarono a offrire servizi di traduzione gratuita. Tra questi, Google Translate (<http://translate.google.com>) è forse il più conosciuto, ma ne esistono molti altri come Bing Translator (<http://www.microsofttranslator.com>), SYSTRANet (<http://www.systranet.com/translate>), Prompt (<http://www.prompt.com>), SDLFreeTranslation (<http://www.freetranslation.com>),

Linguec (http://www.linguec.de), Reverso (http://www.reverso.net), Worldlingo (http://www.worldlingo.com) e il più recente DeepL (https://www.deepl.com/translator).

Oggi la traduzione automatica non è più limitata ai testi scritti. Google Translate per Android, per esempio è utilizzabile come traduttore simultaneo live per tradurre in tempo reale le frasi pronunciate dall'utente.

Se agli albori della TA ci si aspettava che con il tempo si sarebbero potuto tradurre tutti i tipi di testi, in modo, totalmente automatizzato e con alta qualità di output, poi questa fede crollò e si pensò che la TA era stata solo una grande illusione. Successivamente si diffuse la percezione che, poiché non sembrano esistere sistemi di TA capaci di tradurre qualsiasi testo con risultati di alta qualità, che la totale automazione del processo e una buona qualità dell'output fosse possibile, ma non per tutti i tipi di testi. Si raggiunse un certo consenso sul fatto che la traduzione automatica non può sostituire in toto quella umana, ma può raggiungere buoni livelli di accettabilità per testi chiusi, cioè testi tecnici o testi informativi con una struttura lineare e culturalmente neutri, con o senza interventi di postediting o beneficiare di pratiche di scrittura o linguaggio controllato. Il recente avvento della traduzione basata su reti neurali, ovvero sull'intelligenza artificiale, ha provocato una nuova impennata nelle aspettative, anche per quanto riguarda generi testuali prima ritenuti non adatti, quali per esempio la saggistica o addirittura la narrativa.

## **II. Traduzione automatica, sì o no?**

Nel 1965 Italo Calvino scrisse un celebre articolo («l'antilingua») in cui si scagliò contro il dilagare della cattiva lingua della burocrazia, il burocratese. E verso la fine formulava queste previsioni:

«ogni lingua si concentrerà attorno a due poli: un polo di immediata traducibilità nelle altre lingue con cui sarà indispensabile comunicare, tendente ad avvicinarsi a una sorta di interlingua mondiale ad alto livello e un polo in cui si distillerà l'essenza più peculiare e segreta della lingua, intraducibile per eccellenza, e di cui saranno investiti istituti diversi come l'argot popolare e la creatività poetica della letteratura...»

Mezzo secolo dopo sembra che le cose stiano proprio così. Una parte della lingua, quella più espressiva e intima, poetica continua a sottrarsi alla traducibilità, mentre il primo polo «di immediata traducibilità», sembra aver trovato il suo strumento principe nella traduzione automatica. Questa da qualche anno è entrata nella vita quotidiana di agenzie governative, aziende, uffici, studi e redazioni, degli studenti di ogni grado e di tutti gli utenti che hanno la necessità di tradurre siti di lingue sconosciute, articoli, istruzioni, testi di email ecc. Oggi quasi tutti i grandi global player di internet come Google, Microsoft, Yahoo, Amazon, Facebook e Twitter sviluppano propri sistemi o offrono servizi di traduzione automatica di testi o siti in tutte le lingue.



Tra i professionisti c'è chi è entusiasta di questi sviluppi, ma non mancano posizioni di segno opposto. «Non usiamo la traduzione automatica, perché ci hanno detto di non farlo», affermano molti studenti di mediazione linguistica o di traduzione. Infatti, tra molti traduttori e soprattutto docenti di traduzione sono diffusi tuttora atteggiamenti, essenzialmente emotivi, di timore, di diffidenza, di sospetto o di superiorità. C'è chi teme, per esempio che la traduzione automatica disincentivi l'apprendimento delle lingue e favorisca così la divisione dei popoli. Altri considerano questi sistemi dei simpatici giocattoli, buoni tutt'al più per farsi una risata, ma assolutamente non in grado di sostituire il traduttore umano, altri invece temono che diventando troppo efficienti essi diventino un strumento che gli studenti usano per fare esercitazioni e esami di traduzione con poco sforzo. Molti studenti si chiedono quale sarà il loro ruolo se il lavoro viene fatto dalle macchine, oppure si vedono degradati a un ruolo di posteditori, cioè correttori di traduzioni automatiche. Ma poi quando si fanno delle esercitazioni con uno dei traduttori automatici disponibili in rete, come Google Traduttore, molti appaiono affascinati dalla semplicità e dall'efficacia dello strumento e dalle possibilità che la macchina offre di prendersi in carico l'incombenza di tradurre le parti più noiose. «Lavorare così non è tradurre, non è un lavoro da traduttore,» insistono invece gli scettici.

Un illustre portavoce degli TA-scettici è stato Umberto Eco. Il suo saggio sulla traduzione «Dire quasi la stessa cosa» si apre con il capitolo «I sinonimi di Altavista» (Eco, 2007, p. 25 sgg.) che presenta una serie di esempi di traduzioni fatte alla fine degli anni '90 dal traduttore automatico Babel Fish, ai tempi usato dal motore di ricerca Altavista:

<b>Originale inglese</b>	<b>Traduzione italiana di Babel Fish</b>
The Works of Shakespeare	Gli impianti di Shakespeare
Harcourt Brace	sostegno di Hartcourt
Speaker of the chamber of deputies	Altoparlante dell'alloggiamento dei delegati
Studies in the logic of Charles Sanders Pierce	Studi nella logica delle sabbiatrici Peirce del Charles

Successivamente sono riportate le ancor più esilaranti «traduzioni automatiche» dell'inizio del *Genesi* in diverse combinazioni linguistiche. Eco (Eco, 2007, p. 31) spiega che poiché la traduzione non dipende solo dal contesto linguistico, ma anche da qualcosa che sta al di fuori del testo, cioè informazione circa il mondo, i sistemi di traduzione automatica a cui manca tale informazione circa il mondo non possono che fornire risultati mediocri.

Ma, nel 2016, riprovando con Google traduttore, i risultati sono diventati questi:

Originale inglese	Traduzione italiana di Babel Fish	Traduzione italiana di Google Traduttore
The Works of Shakespeare	Opere di Shakespeare	Le opere di Shakespeare
Harcourt Brace	Harcourt Brace	Harcourt Brace
Speaker of the chamber of deputies	Speaker della camera dei deputati	Presidente della Camera dei deputati
Studies in the logic of Charles Sanders Pierce	studi nella logica del Charles Sanders Pierce	Studi nella logica di Charles Sanders Peirce

Osserviamo un netto miglioramento, al punto che per queste espressioni anche un traduttore umano non avrebbe potuto fare di meglio. E se proviamo con molti altri testi, vediamo che quest'impressione è confermata. Cos'è successo, quindi in questi pochi anni? E' successo che i sistemi di traduzione automatica, da qualche anno usano sistemi statistici e da ultimo anche neurali. Ciò significa che attingono le loro soluzioni non più soltanto da sistemi di regole di corrispondenza linguistica, più o meno 1 a 1. Essi, invece si basano su corpora di testi multilingui «umani» raccolti in rete e costantemente aggiornati in cui cercano testi e traduzioni uguali o possibilmente simili, già esistenti, oppure addirittura in grado di apprendere attraverso l'esperienza, come i più recenti sistemi neurali. In sintesi oggi sembrerebbe dunque che dei sistemi di traduzione automatica (statistica o neurale) si possa dire che:

- più si usano (bene), meglio funzionano.
- più stanno online («nella cloud» o in altre basi dati condivise), più diventano accessibili a un largo numero di persone, e più saranno usate.
- più passa il tempo, più imparano e migliorano le loro performance

In altre parole maggiore è l'accessibilità, maggiore è il potenziale uso, e maggiore è la probabilità che il sistema lavori bene. In breve, queste caratteristiche dovrebbero creare un circolo virtuoso.

Siamo dunque dinanzi a una sorta di rivoluzione? Varie ricerche (cfr. per esempio Plitt e Masselot 2010 e Pym 2012) hanno già dimostrato un sensibile aumento della produttività, sia in ambito didattico sia in ambito professionale. Questa rivoluzione non riguarda quindi tanto le tecnologie stesse, quanto il ruolo, l'uso e la funzione sociale della traduzione.

Ovviamente, come in tutte le rivoluzioni, non sempre le cose vanno esattamente come ce lo si aspetta. Nel momento in cui la TA viene usata da tutti e, come nel caso di Google traduttore, si alimenta di testi provenienti da chiunque e da ovunque. Utenti poco competenti possono pubblicare traduzioni approssimative e in questo modo perpetuare traduzioni scorrette che ritornano nelle basi dati sulle quale i sistemi statistici lavorano. Quindi da virtuoso, rapidamente il

circolo si fa vizioso. Per farlo tornare virtuoso, la soluzione potrebbe essere di permettere l'uso solo a utenti selezionati, oppure di permettere ai singoli di sviluppare i propri sistemi di traduzione automatica e le proprie basi dati. Una soluzione più generale potrebbe invece essere quella di prevedere brevi corsi sull'uso della TA, che potrebbero essere utili a tutti.

Possiamo quindi dire che per gli operatori della traduzione è passato il periodo in cui chiedersi traduzione automatica sì o no. La domanda caso mai, è quando sì e quando no. Oppure come sì e come no?

Si tratta di una rivoluzione che riguarda la tecnologia, ma soprattutto anche il ruolo, l'uso e la funzione sociale della traduzione stessa.

Jaap van der Meer, fondatore di TAUS, organizzazione che ha come scopo lo studio e la diffusione della traduzione automatica, ha coniato il termine Convergence Era, significando che la TA non è fine a se stessa, ma un mezzo essenziale in un'epoca in cui i contenuti dovranno essere disponibili in ogni lingua, in ogni momento: la traduzione diventa un'utilità, come l'acqua, l'elettricità o internet, inglobata in qualsiasi cosa come app, strumenti di ricerca, social media, Internet delle cose. Non sarà perfetta ma in molti casi le esigenze di comunicazione in tempo reale diventano più importanti dell'eccellenza linguistica (cfr. van der Meer 2014a).

Quali sono le conseguenze per la professionalità e la formazione dei traduttori? C'è da aspettarsi che in molti ambiti la traduzione automatica, con le sue numerose varianti, sia effettivamente destinata a trasformare presto molti traduttori (anche) in post-editori, ma anche in pre-editori e gestori di sistemi di traduzione. Diventa pertanto urgente ripensare la configurazione di base dei nostri programmi di formazione nonché rivedere i nostri modelli di competenza traduttiva.

### **III.L'uso della traduzione automatica nella pratica professionale**

I sistemi di traduzione automatica sono largamente usati da fornitori di servizi linguistici e servizi di traduzione di istituzioni pubbliche e private e sempre più i sistemi di traduzione automatica sono integrati nei sistemi di traduzione assistita impiegati dai traduttori nella loro pratica professionale.

Infatti già nel 2009 quasi il 50% degli operatori della traduzione (servizi linguistici e traduttori) dichiarava di far uso più o meno frequentemente di sistemi di traduzione automatica (State of the Translation Industry 2009, 2009). Nel 2014 il valore di mercato della TA è stato stimato pari a 250 milioni di dollari e il trend di crescita appare irreversibile (cfr. van der Meer 2014). Secondo un recente studio di Lommel e DePalma (2016), le aziende tendono a ridurre gli investimenti in traduzione umana a favore della cosiddetta PEMT Post Edited Machine Translation Post Editing, cioè l'uso della traduzione automatica per produrre una prima bozza

della traduzione da correggere successivamente, ed entro il 2019 tradurranno con la TA circa il 60% dei loro contenuti.

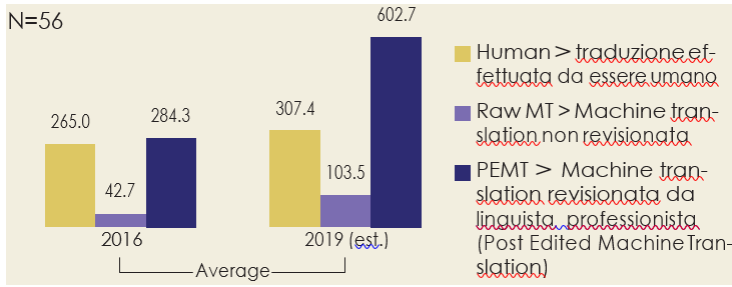


Fig. 1- Stima delle parole tradotte e metodo di traduzione usato nel 2016 e proiezione sul 2019. A livello europeo, nel 2016, sono state tradotte circa 592.000.000 parole, mentre nel 2019 si stima che le parole complessivamente tradotte saranno 1.010.200.000, con un aumento percentuale pari al 70,5% (Lommel e DiPalma 2016)

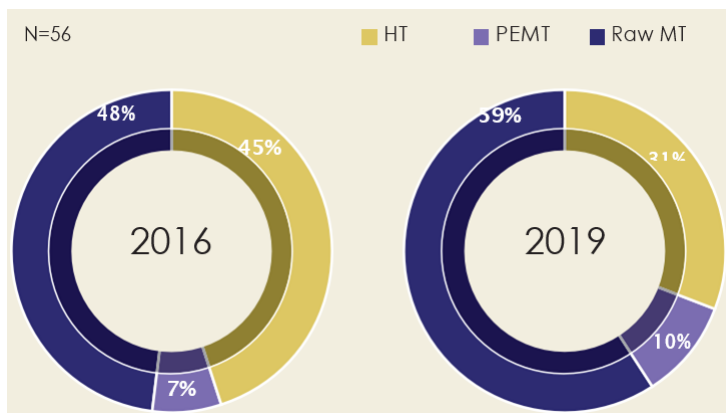


Fig. 2- Si prevede un progressivo aumento delle parole tradotte tramite PEMT/Raw MT, con conseguente decrescita relativa della HT (traduzione umana) (Lommel e DiPalma 2016)

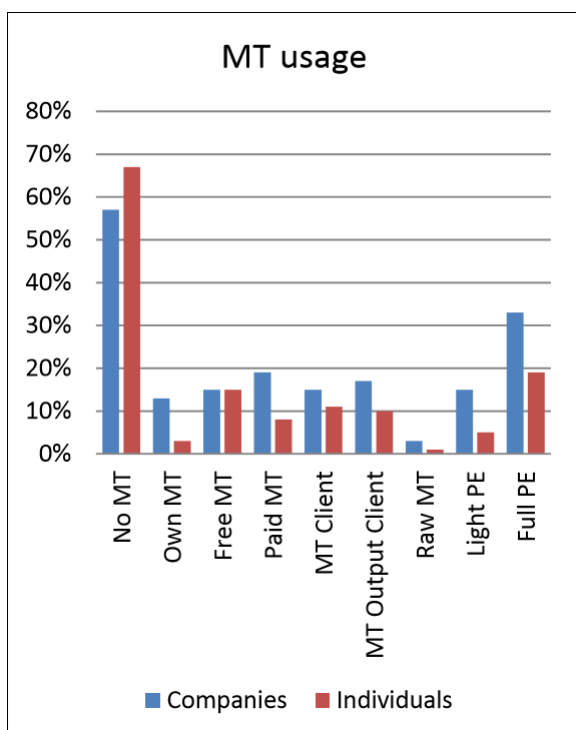


Fig. 3- L'interesse per la TA è maggiore per le aziende che per gli individui (DGT, 2017)

Tutto questo è accompagnato da un significativo aumento di studi e ricerche sulle possibilità d'uso della traduzione automatica da parte dei traduttori<sup>1</sup>.

Alcune cifre. Quante lingue ci sono nel mondo? Quasi 7000. Quanti documenti vengono tradotti in più di 100 lingue? Molto pochi. E qual'è la ragione principale che impedisce la traduzione di un documento in tutte le lingue che potrebbero trarre beneficio dalla traduzione? L'elevato costo della traduzione. Servizi di traduzione che usano la traduzione automatica possono offrire più alternative ai loro clienti.

Senza software di traduzione, un traduttore professionista è in grado, a seconda del tempo impiegato per le ricerche terminologiche e le riletture, di tradurre circa tra le 1500 e le 3000 parole al giorno. Con l'ausilio della TA questa quantità cresce in modo esponenziale.

La produttività può essere modellata sulle esigenze specifiche, poiché è importante tenere presente che la traduzione automatica può essere usata a diversi livelli che richiedono diversi tipi di intervento. I principali sono:

- **gisting**: una traduzione approssimativa generata tramite le funzioni di base della TA senza alcun intervento umano, utilizzata generalmente per le e-mail e per la messaggistica istantanea, per avere un'idea generale del contenuto;

<sup>1</sup> cfr. Bersani Berselli 2011

- **FAUT (Fully Automated Useful Translation)**: una traduzione generata da motori di TA personalizzati grazie all'intervento di linguisti, ma la traduzione viene utilizzata direttamente senza ulteriore rielaborazione;
- **post-editing leggero**: traduzione e revisione umana di livello più basso, spesso definita come «sufficientemente buona» o «adatta allo scopo» (cfr. TAUS-Evaluation 2010); p.es. il cliente richiede una traduzione standard per usi interni, e quindi il traduttore umano corregge solo gli errori che pregiudicano la comprensione del contenuto del testo;
- **post-editing completo**: traduzione e revisione umana di alta qualità» (anche detta «qualità pubblicabile»); p.es. il cliente richiede una traduzione standard destinata alla pubblicazione, e quindi il traduttore umano corregge anche tutti gli errori terminologici e grammaticali e inoltre interviene con ritocchi stilistici;
- **pre-editing**: si sottopone il testo a interventi per ridurre l'ambiguità e complessità del linguaggio naturale, semplificandone la grammatica e restringendone il lessico. (vedi par. scrivere per la TA);
- **integrazione della TA nei CAT-Tool**: molti CAT-Tools offrono funzioni di (pre-) traduzione automatica dei segmenti, che una volta corretti (posteditati), vengono memorizzati e resi disponibili per traduzioni successive;

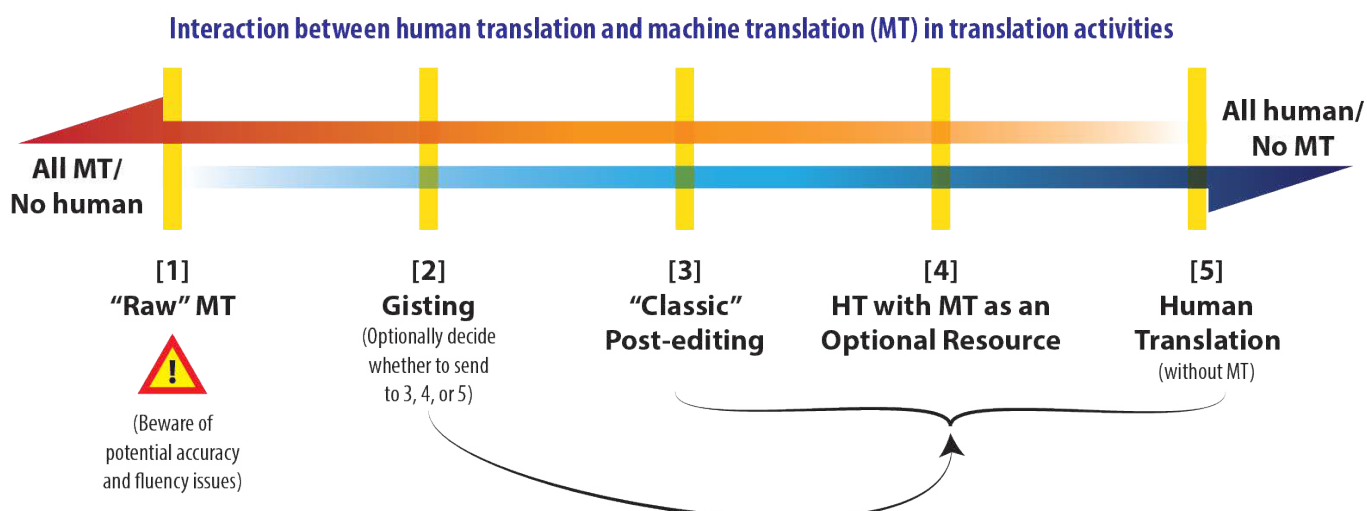


Fig. 4- Schema di Alan Melby (Brigham Young University) in Chen (2016)

## IV. Altri usi della TA

### 1. La TA come dizionario o come banca dati terminologica

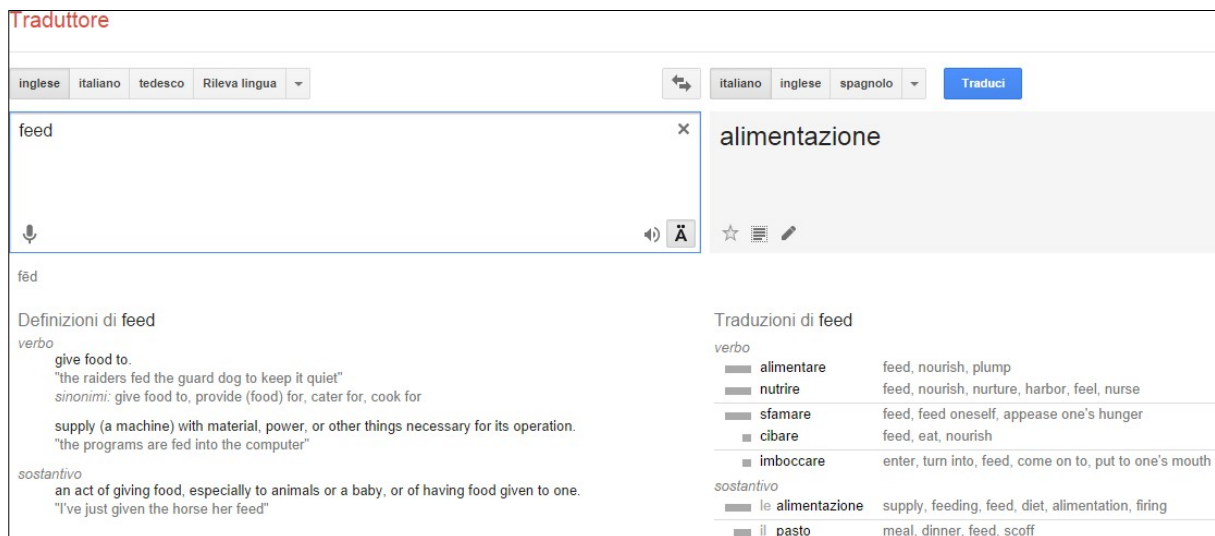


Fig. 3- Traduzione automatica di un termine con Google Traduttore

Una pratica sempre più diffusa è quella di usare i traduttori automatici come dizionario. Google Traduttore, per esempio, contiene direttamente nel suo sistema vari dizionari bilingui. Se un termine è presente, viene mostrata la voce relativa. Inoltre Google offre agli utenti la possibilità di creare propri dizionari personalizzati da usare con la TA. Vari altri dizionari online strumenti offrono servizi simili (p.es. Glosbe, Mymemorynet, Sensagent, BaBla)

Inoltre i traduttori automatici (statistici) sono usati anche come strumento di ricerca di equivalenze in più lingue di termini specialistici. Un recente studio (cfr. Badesso e Bariatine 2016) ha mostrato che per ricerche di equivalenti terminologici, la TA abbinata a tecniche di pre- (vedi appendice) e post-editing permette di ottenere in tempi infinitamente brevi risultati analoghi o migliori di strumenti offline o online, come memorie di traduzione con contesto, testi paralleli, dizionari e banche dati terminologiche ecc. In questi casi la TA inverte l'ordine della classica procedura di ricerca terminologica: invece che ricercare, nelle diverse fonti, definizioni, contesti di utilizzo e equivalenze terminologiche verificandone subito la validità, si fa dapprima tradurre la macchina per poi eseguire approfondimenti concettuali (ed eventuali post-editing) nelle suddette fonti delle diverse lingue.

## 2. La TA per la ricerca di testi paralleli

La TA può essere usata con strumenti come 2lingual.com che traduce in automatico la stringa di ricerca inserita dall'utente e propone in parallelo i risultati di Google in due lingue sulla stessa pagina.

The screenshot shows the 2lingual search interface. At the top, the search query "How does a hybrid car work" is entered in a text box. Below the search bar, there are dropdown menus for "Search pages written in: English" and "and Italian". The search results are displayed in two columns. The left column shows English results for "How does a hybrid car work", including links to "Hybrid" (BMW), "Hybrid Technology" (A-Maier), and "Auto Hybrid" (Peeplo). The right column shows Italian results for "Come funziona un'auto ibrida", including links to "Auto ibride in commercio: cosa sono e come funzionano" (QNM), "Auto ibride: Come funzionano? | ELETTRONICA OPEN SOURCE" (emceletronica.com), and "Auto ibride: quale scegliere - Guide - GreenStyle" (GreenStyle).

Fig.4 - Risultato di ricerca di 2lingual

### 3. La TA per tradurre sottotitoli

Sistemi di traduzione automatica sono integrati, per esempio, in programmi di sottotitolazione come SUMAT o quello di Youtube oppure di riconoscimento vocale e sottotitolazione dal vivo come SAVSA.

## V. Sistemi di traduzione automatica

La sfida dei moderni sistemi di traduzione automatica è come sviluppare un sistema che «comprenda» un testo come fa una persona e che «crei» un nuovo testo nella lingua di destinazione che «sembri» scritto da una persona. Questo problema può essere affrontato in diversi modi. Quella che segue è una descrizione dei principali paradigmi della TA.

### 1. Sistemi basati su regole/ architetture «rule-based»

I sistemi basati su regole, per esempio Babel-Fish, si basano su un numero elevato di regole linguistiche organizzate in moduli diversi che interagiscono a diversi livelli e su milioni di dizionari bilingui per ogni combinazione linguistica. La traduzione automatica basata su regole fornisce una buona qualità nelle situazioni generiche (cioè, non specifiche di un dominio) ed è per sua natura prevedibile.



### 5.1.1 Approccio diretto

I primi sistemi di TA, sviluppati tra gli anni '50 e '60, usavano il cosiddetto approccio diretto. Il processo di traduzione consisteva semplicemente nella sostituzione delle parole della lingua di partenza con supposti equivalenti in quella di arrivo. Nel caso di parole polisemiche (o omografe) nel testo in LP (es. solo, pesca o partita) o di diversi possibili traducanti in LA (es. meta=goal/destination/limit/aim), generalmente nell'output veniva scelta la parola corrispondente all'accezione primaria, selezionando cioè la parola più frequente in generale per aumentare le probabilità di ottenere una traduzione soddisfacente, ma ignorando qualsiasi considerazione contestuale.

#### *Sistemi rule-based ad approccio transfer*

Si cercò quindi di limitare i problemi dell'approccio diretto affinando il software di TA con un approccio transfer. La prima parte del processo consisteva, come nell'approccio precedente, nel trasferimento delle parole dalla lingua di partenza a quella di arrivo. In seguito l'ordine delle parole veniva adattato secondo alcune regole morfosintattiche. In questo modo si riuscì a ottenere un miglioramento nella scorrevolezza e nella naturalezza dell'output.

#### *Sistemi con un approccio basato sull'interlingua*

L'ultimo approccio delle architetture rule-based è quello più complesso. Dato un testo di partenza si cerca di creare una rappresentazione astratta del suo contenuto, per poi generare un testo d'arrivo. Con interlingua si intende, infatti, un linguaggio artificiale di rappresentazione semantica che rimane indipendente sia dalla lingua di partenza sia da quella di arrivo. Il processo si suddivide quindi in due fasi: nella prima alcune regole trasformano l'input nell'interlingua, dalla quale, attraverso nuove regole, viene creato un output.

Il punto di forza nello sviluppo di questo tipo di software sta nel fatto che non bisogna creare delle regole diverse per ogni combinazione linguistica (come accade per gli approcci transfer o diretto). L'interlingua fa da collegamento fra tutte le lingue di partenza e di arrivo. Questo vantaggio teorico si scontra però con uno svantaggio più pratico. Infatti, è molto difficile ottenere una rappresentazione semantica astratta e allo stesso tempo accurata e neutrale rispetto a ogni lingua di partenza e di arrivo, poiché essa dovrebbe contenere una quantità enorme di informazioni, che dovrebbero mantenersi equidistanti da qualsiasi altra lingua.

La creazione di un'interlingua da parte di uno studioso, o di un gruppo di studiosi, rappresenta sempre inoltre un'inferenza della loro cultura. Questo problema potrebbe rivelarsi marginale nell'uso dell'interlingua per la traduzione di due lingue occidentali, che condividono molti aspetti semantici. Traducendo invece, per esempio, da una lingua occidentale in una lingua orientale (o viceversa), il problema diventa rilevante.

### Limiti dei sistemi rule-based

Le architetture rule-based più tradizionali presentano notevoli problemi di sviluppo, oltre che di resa stilistica dell'output. Creare regole esplicite di natura lessicale, morfo-sintattica e semantica che consentano il passaggio da una lingua all'altra richiede un'elaborazione lunga e dispendiosa da parte di personale con competenze elevate sia dal punto di vista linguistico che da quello informatico.

Inoltre la direzione traduttiva delle regole è vincolata: la creazione di un sistema transfer che traduca dal tedesco all'italiano non è reversibile, e quindi non permette anche la traduzione nella direzione opposta (dall'italiano al tedesco), per la quale sarebbe necessario investire nuove risorse per creare nuove regole di elaborazione. Tutti questi fattori contribuiscono a restringere la rosa di lingue analizzabili.

## 2. Sistemi statistici/Approcci statistici

Il sistema di TA statistico (definito anche data-driven o example-based), sviluppato a partire dagli anni '90, è del tutto alternativo rispetto a quello *rule-based*, in quanto opera seguendo delle informazioni non categorizzate sotto forma di regole. Le corrispondenze e i traduttori vengono ricavati per mezzo dell'analisi probabilistica di numerosi testi già tradotti.

Il programma funziona grazie a un sistema di algoritmi che analizza un enorme insieme di raccolte di testi originali con le rispettive traduzioni umane allineate tra loro frase per frase (corpora bilingue paralleli). Più i corpora sono numerosi, più facilmente gli algoritmi del programma riusciranno a proporre una o più versioni tradotte della parola o delle parole che compongono il testo della lingua di partenza. I possibili abbinamenti trovati vengono poi filtrati sulla base di uno o più testi della lingua d'arrivo. In questo modo tra tutti i possibili traduttori si possono selezionare quelli che nella lingua d'arrivo risultano più plausibili. Procedendo esclusivamente in base a considerazioni probabilistiche la TA statistica arriva a fornire un output completo senza dover ricorrere alle regole lessicali, morfo-sintattiche e semantiche delle lingue coinvolte ed è quindi in grado di rendere anche fenomeni traduttivi come trasposizioni, permutazioni ecc. La traduzione è fluida e i risultati dei sistemi di impronta statistica risultano nel complesso migliori rispetto a quelli che fanno ricorso a regole.

La creazione dei modelli di traduzione statistica è un processo rapido, ma la tecnologia fornisce una buona qualità solo se sono disponibili corpora estesi e qualificati. Questo è il motivo, perché i sistemi statistici hanno registrato un'impennata con la comparsa dei servizi di traduzione automatica dei colossi del Web come Google e Bing (Microsoft), Facebook, il motore di ricerca russo Yandex e quello cinese Baidu.

Prima di internet, la fonte principale di queste traduzioni era costituita da corpora come i documenti delle Nazioni Unite che erano stati tradotti in più lingue. Ma il Web ha prodotto una vera miniera di testi in tutte le lingue e l'indicizzazione di Google, p.es., ha semplificato il

controllo e il recupero delle informazioni tra miliardi di documenti utili per la traduzione di contenuti web, documenti o post da una lingua a un'altra. Gli algoritmi impiegano principi analoghi a quelli usati per l'identificazione dei siti web e sono grado di stabilire un ranking qualitativo per capire quali siano le migliori traduzioni.

Uno dei maggiori punti di forza della TA statistica, è che essa non necessita di massicci investimenti per sviluppare degli algoritmi che analizzino i testi secondo delle regole linguistiche esplicite. Per costituire un sistema di TA statistica completo e in grado di funzionare è sufficiente un'adeguata quantità di corpora paralleli allineati (ovvero di testi con le relative traduzioni, in cui a ogni frase del testo di partenza corrisponde una frase del testo d'arrivo). Tuttavia, mentre questo tipo di risorse è ampiamente disponibile nel caso delle lingue che godono di maggiore interesse a livello commerciale, esso scarseggia se consideriamo le lingue meno conosciute o meno studiate. Ciononostante, anche in questi casi l'investimento per ottenere un sistema di TA statistico è molto minore rispetto a qualsiasi altro sistema rule-based.

### *Limiti dei sistemi statistici*

Nonostante questi vantaggi, anche la TA statistica non è priva di problemi. Innanzitutto, per via del suo approccio, questo sistema non effettua traduzioni per così dire ex novo, ma crea un output basandosi su testi già tradotti da persone, quasi come una grande memoria di traduzione. Il problema per la TA statistica sta nel ricombinare questi elementi in modo da salvaguardarne la qualità complessiva dell'output.

Una soluzione al problema consiste nell'implementazione di sistemi ibridi, aggiungendo cioè alcune regole linguistiche e traduttive esplicite, che possono ovviare ad alcune delle possibili imperfezioni dovute all'approccio statistico. Questa tecnica si rivela utile specialmente nei casi in cui i corpora impiegati per l'addestramento iniziale del sistema siano di ridotte dimensioni, o l'input non risulti simile, per quanto riguarda il dominio o il genere, ai dati sui quali il sistema calcola le sue corrispondenze traduttive.

### *Sistemi addestrabili dall'utente*

Inoltre, ogni software statistico deve essere «addestrato» su vari corpora paralleli. Quelli maggiormente disponibili sono limitati a ben determinate tipologie testuali: leggi, manuali, contratti ecc. Una volta effettuato l'addestramento sui corpora di una determinata tipologia testuale, gli algoritmi estrarranno corrispondenze traduttive legate a quello specifico campo semantico e linguistico. Questa caratteristica rende i sistemi statistici utili essenzialmente per testi analoghi a quelli presenti nel loro database. Se infatti si fornisce un input di una determinata tipologia testuale a un software di TA statistica addestrato per altri campi semantici, il sistema ha difficoltà a fornire una traduzione chiara e adeguata al nuovo testo di partenza in questione.

Sempre più diffuso l'uso di software o piattaforme che permettono agli utenti di costruire e addestrare sistemi di traduzione automatica statistica, anche in combinazioni linguistiche poco frequenti, a partire da raccolte di traduzioni umane (corpora), ovvero testi paralleli, memorie di traduzione e raccolte terminologiche selezionate e caricate dall'utente stesso. Tra questi sono da citare MOSES, KANTAN e Microsoft Translator Hub. Il limite di questi sistemi è che per dare risultati utili, hanno bisogno di essere «nutriti» con quantità molto ingenti di dati. Si prestano quindi per organizzazioni di dimensioni medio-alte che dispongono di molte traduzioni su un certo tema e meno per realtà piccole o singoli.

Per ovviare agli inconvenienti dell'uno e dell'altro sistema la ricerca si è mossa in direzione di sistemi (per esempio Bing Translator), e combinano le metodologie basate su regole, abbinati alle singole lingue, con quelle statistiche con lo scopo di sviluppare diverse strategie di risoluzione a seconda dello specifico problema traduttivo da affrontare.

#### **4. Sistemi neurali**

La più recente evoluzione della TA è la cosiddetta Neural Machine Translation (NMT) introdotta nel 2016 da Google, Systran, Microsoft e Facebook, seguiti da Amazon e più recentemente dalla promettente DeepL. Ogni volta che un utente richiede una traduzione in una delle sopra citate lingue, questi sistemi fanno appello alle reti neurali, seguendo un percorso logico dinamico variabile in base al contesto della frase e del significato che viene dato alla stessa.

Una rete neurale artificiale, normalmente chiamata solo «rete neurale», è un modello matematico/informatico di calcolo basato sulle reti neurali biologiche. Si tratta pertanto di un insieme di elementi semplici interconnessi in parallelo in larga misura, la cui organizzazione gerarchica è progettata in modo da interagire con gli oggetti del mondo reale come il sistema nervoso umano. L'aspetto più tipico di una rete neurale è la capacità di acquisire «esperienza», nel senso di autorganizzarsi sulla base di esempi del compito che deve svolgere. Si tratta quindi di sistemi che si avvicinano molto alla logica umana a differenza del costruito programmatico dei sistemi informatici statici basati su codice operativo. La grammatica dovrebbe, perciò, migliorare notevolmente, consentendo un'analisi completa del costruito sintattico e la traduzione di più paragrafi. Un sistema quindi, che non è semplicemente chiuso in sé stesso ma riesce, col tempo, a perfezionarsi attraverso un'intelligenza basata su capacità di autoapprendimento (machine learning). In grado di ottenere, quindi, traduzioni sempre più precise. Infatti il sistema prende in esame la frase nel suo insieme, ma non trascura i singoli elementi e tutte le loro accezioni, senza considerare connotazioni linguistiche o etimologiche, ma basandosi esclusivamente su matematica e statistica.

E dove la statistica dovesse risultare poco utile, per esempio nel caso di parole rare, il sistema è in grado di creare dei collegamenti e delle associazioni utili per il loro riconoscimento.

Dal momento che si parla di autoapprendimento grazie a milioni di iterazioni, verrebbe superato anche il problema dell'obsolescenza dei termini, poiché il sistema segue le lingue nel loro sviluppo.

## VI. Metodi di valutazione della traduzione automatica

Come dire che una traduzione è «buona» o che un sistema è «migliore» di un altro?

Non è facile dare una risposta univoca. Infatti, come nella traduzione umana, potrebbero essere accettabili molte diverse traduzioni.

Per la valutazione della TA vengono usate diverse metriche (criteri): valutazioni eseguite in modo soggettivo da valutatori umani, valutazione se il messaggio passa, valutazioni della mole di post-editing necessaria per ottimizzare il testo, metriche di valutazione automatiche.

Schematicamente i sistemi di valutazione possono essere suddivisi in sistemi di valutazione umani centrati sull'utente (cfr. p. es. Van Slype (1979), Lehrberger/Bourbeau (1988), Arnold (1994) e Hutchins (1997), Koehn (2012) preferiti per valutare la qualità dell'output della TA e sistemi di valutazione automatizzati (BLEU, Meteor, Nist) usati per misurare l'efficienza dei software di TA.

Un sistema centrato sull'utente è il modello «Adequacy/Fluency» di Philipp Koehn. Secondo questo chi esamina la traduzione deve chiedersi se essa è adeguata, cioè trasmette lo stesso significato della frase del testo di partenza, oppure se vi sono omissioni, aggiunte o distorsioni e poi se la lingua d'arrivo è «fluente», cioè corretta per quanto riguarda la grammatica e le scelte sintattiche, idiomatiche, lessicali e terminologiche. Per la valutazione si usano due scale con punteggio da un massimo di 5 a un minimo di 1.

Adequacy		Fluency	
5	all meaning	5	flawless English
4	most meaning	4	good English
3	much meaning	3	non-native English
2	little meaning	2	disfluent English
1	none	1	incomprehensible

Questa griglia indubbiamente permette all'utente di valutare la qualità e l'utilità di una traduzione automatica, anche se le diverse gradazioni non sempre risultano univoche. P. es. può essere difficile distinguere tra «most meaning» e «much meaning» o tra «non-native English» e «disfluent English», e in più in generale collocare con precisione il confine tra la «adequacy» e «fluency» (se una frase è scritta male, spesso diventa anche poco comprensibile).

Tra i vari altri modelli, il più funzionale (usabile) da parte degli utenti è quello di Arnold (1994). Considerando solo la comprensibilità (intelligibility) questo propone una scala di quattro punti, ovvero una frase può essere completamente comprensibile (1), in maggior parte comprensibile (2), solo dopo un'analisi approfondita (3), oppure completamente incomprensibile. Questa è stata adattata da me e Galati sostituendo il criterio della

comprensibilità (intelligibility) con quello dell'accettabilità, depurandola di elementi che potevano generare incertezza nella valutazione. L'accettabilità è stata valutata quindi secondo la seguente scala di valori:

Accettabilità
3 - traduzione perfettamente accettabile. Non richiede revisione
2 - traduzione comprensibile, ma richiede revisione stilistica.
1 - traduzione comprensibile, ma con errori grammaticali, linguistici, e stilistici
0 - traduzione incomprensibile. Richiede riscrittura.

Questo modello assolutamente centrato sulle esigenze del traduttore/redattore, permette di valutare speditamente se una traduzione è di livello umano (pronta per essere stampata p.es.) (livello 3), «sufficientemente buona» per essere capita, ma non pubblicata (livello 2) o se necessita di interventi di pre-editing o di post-editing più o meno approfondito per arrivare al livello 2 0 3 a seconda dello scopo (livelli 2, 1, 0).

Per favorire una stima del lavoro necessario o per confrontare l'output di varie traduzioni (p.es prima o dopo il pre-editing), abbiamo realizzato una scheda preimpostata per il numero di segmenti analizzati, che alla fine restituisce la percentuale complessiva dell'accettabilità del testo, laddove 100% si raggiunge con 3 punti in tutti i segmenti mentre 0% si ottiene con 0 punti in tutti segmenti.

I sistemi automatizzati come BLEU (Bilingual Evaluation Understudy), il più diffuso, o Meteor e Nist, si basano invece su algoritmi che valutano la qualità di un testo tradotto da una macchina confrontandolo con una traduzione di riferimento effettuata da un traduttore umano. L'ipotesi alla base di questi sistemi è che una traduzione di buona qualità condivida un elevato numero di parole con la traduzione di riferimento. Quanto più elevata è la corrispondenza tra le due traduzioni, più la qualità è giudicata migliore. In realtà il confronto viene fatto con un insieme di traduzioni fatte da più traduttori per evitare giudizi troppo soggettivi, e si calcola il giudizio medio. Va notato come in questo metodo il testo di partenza scompare, e il confronto è fatto solo con le traduzioni di riferimento realizzate dai traduttori umani. L'algoritmo di BLEU, p.es., fornisce un numero compreso tra 0 e 1. Più alto è il valore, migliore viene giudicata la qualità della traduzione. Tuttavia il valore 1 è irraggiungibile in quanto richiederebbe la perfetta identità alla traduzione di riferimento.

## VII. Scrivere per la TA – Lingue naturali, sotto-linguaggi, linguaggi controllati

Purtroppo ai sistemi di TA manca il buon senso (chissà se con la TA neurale cambierà?). Nel caso di ambiguità un (almeno discreto) traduttore umano traduce in modo ragionevole. La macchina non necessariamente, e il risultato sono quelle traduzioni palesemente assurde, esilaranti o semplicemente incomprensibili, molto frequenti nei sistemi basati su regole, ma presenti, anche se meno, anche nei sistemi statistici. Tra le principali fonti di errore vi sono le asimmetrie e le ambiguità lessicali:

- una certa parola non può essere tradotta sempre con un unico equivalente (il *wood* inglese in italiano a seconda del contesto si traduce con *bosco*, *legno* o *legna*).
- molte parole di uso comune possono appartenere a diverse categorie grammaticali (sostantivo, verbo, aggettivo ecc.):
  - My team won the first *round* and got through to the quarter finals
  - Cowboys used to *round* up the cattle
  - King Arthur sat with his knights at the *Round* Table

Oppure vi sono le ambiguità nei riferimenti. Nella frase «Old men and women were evacuated», *old* si riferisce solo agli uomini o anche alle donne? Può trovarsi in difficoltà un traduttore umano, figuriamoci la macchina. Lo stesso discorso vale per ambiguità strutturali dei complementi. P. es.: «I saw Michael on the roof with my eyes», potrebbe essere interpretato come scena surreal-horror tipo «Ho visto Michael sul tetto con i miei occhi in mano», perché il sistema TA non possiede strumenti per cogliere le differenze in gioco.

Altro problema è quello dei nomi propri, omografi di parole comuni. Pensiamo a George Bush, Bill Gates o Donald Trump, Windows, Apple, spesso tradotti letteralmente dai traduttori automatici. Problematica è poi la resa dei pronomi (p. es. l'inglese *it*)

In generale possiamo dire che sono a rischio tutte le traduzioni che dipendono dal contesto, cioè da considerazioni extra-testuali che un traduttore umano è in grado di fare, mentre la macchina no, perché le sue traduzioni derivano esclusivamente dalle corrispondenze nei propri dizionari (i sistemi basati su regole) o dai corpora (i sistemi statistici) con cui il software è stato addestrato.

Un modo efficace per limitare il rischio di errori della TA è quello di esercitare un controllo sul processo di scrittura dell'input, attraverso una serie di restrizioni linguistiche.

### 1. Sotto-linguaggi



### 1.1 Cos'è un sotto-linguaggio?

Una lingua naturale è un sistema di comunicazione parlato o scritto proprio di una comunità umana formato da oggetti arbitrari quali segni, suoni e norme che disciplinano quei segni. Questo grande insieme tende a essere suddiviso in linguaggi speciali o sotto-linguaggi (in inglese *sublanguages*) più piccoli a seconda della situazione comunicativa in cui si è coinvolti. Un sotto-linguaggio è l'insieme dei mezzi linguistici (di tipo lessicale, morfologico, fraseologico e sintattico) adottato in modo convenzionale e consensuale dagli individui che operano in uno stesso settore, per lo scambio e la divulgazione di informazioni e conoscenze a esso relative.

I sotto-linguaggi sono caratterizzati dalla presenza di terminologie specifiche, di risemantizzazioni (cioè parole comuni con significati diversi, p. es. mouse, bit, bus, browse, ecc.) e cosiddetti tecnicismi collaterali, vale a dire termini tipici di un certo settore o espressioni stereotipiche non legati a esigenze comunicative (il farmaco che *antagonizza* invece di *contrastare*; *accendere* un mutuo invece di *fare*), di determinate forme fraseologiche, particolarità sintattiche e la predilezione per determinate forme grammaticali (p.es. stile nominale in testi amministrativi, assenza di domande dirette in manuali d'istruzione, frasi imperative nelle ricette, uso del tempo futuro dei verbi nelle previsioni meteo ecc.).

Un sotto-linguaggio può essere definito anche come una sequenza di frasi le cui restrizioni lessicali e grammaticali riflettono l'insieme di oggetti e relazioni presenti in un determinato ambito del discorso (cfr. Kittredge e Lehrberger 1982, 2).

Tuttavia anche all'interno dello stesso sotto-linguaggio è possibile incontrare una grande varietà di tipi di testo. Alcuni sono determinati dal mezzo (parlato o scritto), dall'autore, dal contenuto, dalla funzione e dall'obiettivo del testo. Per esempio, nel campo automobilistico possiamo imbatterci in opuscoli pubblicitari e manuali d'uso destinati ai clienti finali oppure in informazioni tecniche o istruzioni di riparazione destinate a ingegneri o meccanici. Ognuna di queste tipologie ha caratteristiche particolari, p. es. terminologie e strutture sintattiche, che generano una sorta di sotto-sotto-linguaggi o (un sotto-linguaggi o con particolari caratteristiche all'interno del sotto-linguaggi o automobilistico più generale).

### 1.2 Sotto-linguaggi e TA?

In effetti, esiste un chiaro legame tra l'uso di sotto-linguaggi e la TA. È noto che i sistemi di TA possono trovarsi in difficoltà con testi di tipo generalistico e che i risultati sono molto migliori quando invece i testi sono in qualche modo «ristretti» e nel caso dei sotto-linguaggi queste restrizioni sono presenti naturalmente. Problemi lessicali come il mancato riconoscimento di omonimie, omografie e polisemie si riducono in presenza di una buona gestione della terminologia che determina in modo preciso la categorizzazione e sottocategorizzazione degli elementi lessicali. Poiché di solito questi sono i principali portatori del significato testuale, se vengono tradotti correttamente, ci sono grandi possibilità che il



significato del testo venga trasferito, anche se le strutture sintattiche non sono perfette o non suonano naturali nella lingua di destinazione.

L'uso di sotto-linguaggi in combinazione con la TA fu esplorato per la prima volta alla fine degli anni '60 da un gruppo di ricercatori della New York University guidato dallo studente di linguistica Naomi Sager.

In seguito un'intensa attività di ricerca è proseguita fino agli anni '90. Furono ottenuti risultati significativi con diversi sotto-linguaggi nell'ambito biomedico, ma il maggior successo fu il sistema TAUM-METEO. Questo fu sviluppato negli anni '70 presso l'Università di Montreal per tradurre automaticamente i bollettini meteo dall'inglese in francese per il governo canadese fino al 2001. Il sistema usava un sotto-linguaggio controllato per migliorare la qualità dell'output della TA. Un altro sistema che utilizzò un sotto-linguaggio controllato (il Langage Documentaire canonique) fu TITUS. Questo fu progettato negli anni '70 presso l'Institut Textile de France per il trattamento multilingue di abstract in un database on-line.

Tuttavia, non sempre i sotto-linguaggi sono adatti per l'uso della TA. P. es., lo stesso gruppo che progettò il TAUM-METEO lavorò anche al TAUM-AVIATION, un sistema sperimentale per la traduzione dall'inglese al francese del sotto-linguaggio o dei manuali di manutenzione tecnica. Tuttavia questo sistema ebbe meno successo, perché presto ci si rese conto che i testi di questi manuali non costituivano un dominio sufficientemente limitato. Oppure, un altro sotto-linguaggio o che presenta evidenti difficoltà per i sistemi di TA è il linguaggio amministrativo (burocratese), costellato com'è di nominalizzazioni, arcaismi, aggettivi sostantivati, tecnicismi più o meno necessari e lunghe serie di subordinate.

Kittredge (1985: 159) distingue le seguenti caratteristiche che rendono i sotto-linguaggi più o meno adatti alla TA:

- *Dimensioni del lessico.* Infatti, a seconda del dominio queste possono variare enormemente. Il sublanguage dei bollettini meteorologici si basa su un lessico di meno di 1000 parole, esclusi i toponimi. Un manuale di manutenzione degli aeromobili contiene oltre 4000 denominazioni differenti solo per la parte idraulica;
- *Complessità.* Un sotto-linguaggio può essere molto vasto per quanto riguarda il lessico, ma può usare strutture delle frasi prevedibili, che favoriscono la TA. All'altro estremo può avere un vocabolario ridotto ma utilizzare, frasi o strutture piene di ellissi che invece la ostacolano;
- *Uso sistematico.* Il grado di sistematicità è dato dall'utilizzo sistematico dei modelli distributivi di parole che definiscono il sotto-linguaggio. Quanto più il sotto-linguaggio è sistematico, tanto più sarà adatto alla traduzione automatica.

I sotto-linguaggi nascono dalla necessità di esprimere determinati contenuti specialistici in modo essenziale e comprensibile. Al fine di trasmettere questi contenuti informativi in modo

preciso e inequivocabile c'è stato sempre un interesse a controllare il vocabolario e le strutture grammaticali e sintattiche utilizzate in questi sotto-linguaggi, specialmente nella comunicazione scritta. Non è difficile imbattersi in guide di stile, norme e raccomandazioni che regolano comunicazione di determinati ambiti specialistici.

## 2. Linguaggi controllati

Se un sotto-linguaggio rappresenta dunque un segmento della lingua naturale «spontaneamente» sottoposto a restrizioni del vocabolario e delle strutture grammaticali per ragioni di funzionalità comunicativa, un linguaggio controllato è invece, una forma di uso di una lingua secondo restrizioni applicate all'uso del lessico e di determinate strutture sintattiche, decise deliberatamente prima della stesura del testo. Ricorrere a un linguaggio controllato può far sì che i testi redatti secondo queste restrizioni siano più chiari e meno ambigui: una frase complessa e/o ambigua può costituire un problema sia per un lettore/traduttore umano che per un traduttore automatico. Pratiche di scrittura semplificata e di linguaggio controllato sono usate in molti ambiti per favorire la comprensibilità e la traducibilità dei testi, compresa la loro fruibilità da parte di lettori con scarse conoscenze linguistiche e persone disabili. Questo è il caso, per esempio nella comunicazione tecnica (technical writing) per la redazione di manuali, guide, libretti di istruzioni e in generale di testi e documenti, al fine di favorirne la chiarezza e la leggibilità da parte degli utenti e permetterne un corretto utilizzo grazie alla maggiore precisione e la riduzione di potenziali ambiguità interpretative.

Nella stessa direzione vanno anche le iniziative per la semplificazione della comunicazione delle pubbliche amministrazioni con i cittadini, intraprese negli ultimi decenni, come l'adozione della *Plain language* negli USA (Carter 1978-79, Clinton 1998, Obama 2010), le indicazioni per la semplificazione del linguaggio amministrativo contenute nel *Codice di stile* di Cassese del 1993, nel *Manuale di stile* di Bassanini del 1994-97 e nelle *Direttive Frattini* del 2002, i suggerimenti della Ue per favorire una comunicazione chiara sintetizzate nel manuale *Scrivere chiaro* pubblicato dalla Direzione Generale Traduzione nel 2012, o l'iniziativa *Leichte Sprache* (lingua facile) diffusa nei paesi di lingua tedesca. In particolare per quanto riguarda comunicazione web, indicazioni analoghe sono presenti in diverse raccomandazioni per l'usabilità dei siti web e delle applicazioni per dispositivi mobili e in parte anche nelle linee guida per l'accessibilità di soggetti disabili (Web Content Accessibility Guidelines – WCAG 2.0).

Nel tempo, vari sono stati gli autori che hanno classificato i linguaggi controllati. Huijsen (1998), per esempio, presenta una distinzione tra Human-oriented controlled language (HOCL) e Machine-oriented controlled language (MOCL): mentre i primi puntano a migliorare la leggibilità facilitando la comprensione del testo, nel secondo caso si tratta di redigere testi in modo da migliorare la loro traduzione in un sistema di traduzione automatica.

Il primo linguaggio controllato, sviluppato intorno al 1930, fu il British American Scientific International Commercial English, più conosciuto come BASIC English, secondo il quale poche centinaia di parole erano sufficienti per redigere un testo in inglese. Invece di usare tutte le 100.000 parole disponibili in un normale dizionario di lingua, il vocabolario da usare fu ridotto a 850 vocaboli e la grammatica drasticamente semplificata. Negli anni '60 l'idea penetra nell'industria, che pensa che la scrittura di testi e manuali in un inglese sufficientemente semplice, avrebbe potuto renderli più facilmente comprensibili anche da ingegneri e tecnici stranieri e renderne superflua la traduzione. Negli anni '80 l'uso di linguaggi controllati si diffonde prima nell'industria aeronautica e successivamente molte grandi aziende e organizzazioni sviluppano loro linguaggi quali per esempio il ASD-STE 100 (STE) della ASD, Boeing Technical English (BTE) della Boeing, Caterpillar Technical English (CTE) della Caterpillar Inc., Controlled Automotive Service Language (CASL) della General Motors, IBM Easy English, Multinational Customised English (Xerox Corporation), Standard Language della Ford, AECMA Simplified English della European Association of Aerospace, Controlled Language at Diebold della Diebold, ACE (Attempto Controlled English) dell'Università di Zurigo, Français Rationalisé del Groupement des Industries Françaises Aéronautiques et Spatiales - GIFAS, Scania Swedish della Scania, Siemens Dokumentationsdeutsch (SDD), Simplified Technical Russian dell'azienda russa aeronautica di costruzione di parti software AVIAOK, Controlled Chinese dell'Istituto di Linguistica Computazionale ZHANG Qinlong, Controlled Language Authoring Technology (CLAT) dell'Istituto di Ricerche Informatiche tedesco. Per quanto riguarda l'italiano nel 2016 Ilaria Gobbi ha pubblicato «ITS italiano tecnico semplificato».

Tutti questi linguaggi contengono indicazioni riconducibili a una serie di regole di base:

Con Gabriele Galati abbiamo riassunto i principi essenziali del linguaggio controllato nell'acronimo **BASICO**:

**Brevi**: non più di 25 parole per frase, non più di 500 battute in un paragrafo

**Attivi**: usare la forma attiva e se possibile l'indicativo, evitare gerundi, forme passive e forme impersonali

**Semplici**: usare vocaboli semplici, evitare parole rare o polisemiche e tecnicismi

**Incisivi**: una sola idea in ogni frase senza usare subordinate o incisi

**Chiari**: esprimere sempre il soggetto e l'oggetto, evitare pronomi e particelle pronominali; esprimere lo stesso concetto senza temere le ripetizioni e senza usare sinonimi

**Ottimizzati** per il medium di destinazione

### *Parole e termini*

- Usare vocaboli possibilmente semplici (vocabolario di base), evitare parole rare o polisemiche e usare solo tecnicismi indispensabili;
- Non usare sinonimi per definire lo stesso elemento;
- Spiegare il significato delle abbreviazioni;
- Fare attenzione che l'ortografia sia corretta;

### *Verbi*

- Usare verbi precisi al posto di espressioni sostantivali o verbi generici;
- Usare la forma verbale attiva al posto di quella passiva;
- Usare la forma attiva e se possibile l'indicativo, evitare gerundi, forme passive e forme impersonali;

### *Frase e periodi*

- Esprimere una sola informazione e una sola istruzione per ogni frase;
- Sviluppare un concetto complesso in elenchi verticali;
- Usare la stessa struttura per esprimere concetti uguali o simili;
- Usare la punteggiatura in modo corretto ed essenziale;
- Usare al massimo 30 parole nei segmenti descrittivi e al massimo 25 parole nei segmenti istruttivi e normativi;

## *1.3 Vantaggi*

In conclusione, migliorare il testo di partenza utilizzando un linguaggio controllato per la sua redazione porta a vari vantaggi:

- chiarezza dei contenuti: in particolare nel dominio tecnico è fondamentale che le istruzioni siano semplici e chiare per tutti gli utenti (siano essi tecnici specializzati o utenti finali di un dato prodotto). Inoltre è possibile utilizzare un linguaggio controllato per testi destinati a parlanti non nativi di una lingua o a persone con difficoltà cognitive;
- riduzione dei costi: per quel che riguarda la traduzione, la revisione è meno costosa della traduzione (affermazione vera solamente nel caso in cui il linguaggio controllato

utilizzato è ben sviluppato e creato appositamente per il sistema di traduzione automatica in uso);

- velocità di traduzione: per esempio nel campo tecnico-informatico è spesso difficile tenere il passo dei rapidi sviluppi e lanci di nuovi prodotti sul mercato con la traduzione multilingui eseguite da traduttori umani. Redigere i testi con un linguaggio controllato può migliorare sensibilmente la qualità dell'output dei sistemi di TA e ridurre drasticamente i tempi e i costi di postrevisione (postediting).

Il ricorso a un linguaggio controllato può porre, però, anche problemi connessi alla capacità espressiva che può risultare tanto più limitata quanto più sono restrittivi i vincoli imposti. La velocità di scrittura risulta ridotta se si scrive in un linguaggio controllato: l'uso delle regole richiede grande attenzione circa le parole e le costruzioni sintattiche da usare e molto tempo per non incorrere in violazioni. Infine, imparare a scrivere in un linguaggio controllato richiede un consistente tempo di addestramento, il che implica investimenti che non sempre le aziende sono disposte ad affrontare.

### *3. Pre-editing*

Il pre-editing comprende tutti gli interventi di «correzione» di un testo prima che esso venga inviato al sistema di traduzione automatica. Analogamente al post-editing esso può essere «leggero», ovvero limitarsi a una correzione di refusi, spaziature errate, imprecisione nell'interpunzione oppure essere «profondo», ovvero consistere in una vera e propria traduzione intralinguistica da linguaggio comune in linguaggio controllato.

### *Appendice: Come scrivere per la traduzione automatica (di Lara Bernardi)*

Ecco alcune regole su come scrivere e pre-editare i testi per la traduzione automatica in lingua italiana

1. Evitare errori ortografici e di punteggiatura (in particolare apostrofi e accenti)

Le macchine (per ora...) non pensano come gli uomini: se noi leggiamo «conqistare» possiamo dedurre senza particolari problemi che è presente un errore e che il verbo corretto è «conquistare». La maggior parte dei sistemi di traduzione automatica non ragiona in questo modo e non proporrà nessuna traduzione soddisfacente. La questione della punteggiatura, in particolare quella della virgola (dove si trova, davanti o dopo quale costruzione sintattica) è una questione ancora controversa e non soggetta a regole, se non quella di evitarne un uso troppo frequente.

2. Scrivere frasi con una struttura semplice, con il soggetto al primo posto e il verbo al secondo posto

Dalle mie analisi è emerso che qualsiasi tipo di frase con qualsiasi tipo di contenuto viene tradotta meglio se presenta una struttura con il soggetto al primo posto e il verbo in seconda posizione. Il soggetto e il verbo sono parti fondamentali di una frase e nella maggior parte dei casi, se presenti, vengono tradotti e posizionati in modo corretto.

3. Evitare frasi troppo lunghe che contengano più di due soggetti o più di due verbi

Le frasi troppo lunghe rappresentano un grande problema per i sistemi di traduzione automatica poiché, nella maggior parte dei casi, sono composte da coordinate e subordinate che presentano più soggetti e più verbi o verbi diversi che dipendono dallo stesso soggetto. Già un traduttore esperto potrebbe trovarsi in difficoltà nel riportare una frase lunga in un'altra lingua, ma la rilettura e la scomposizione mentale della frase permettono al traduttore di trovare una soluzione ideale. Il computer non svolge però nessun lavoro di scomposizione della frase. L'elenco puntato e la scomposizione della frase lunga in frasi semplici più corte possono aiutare.

4. Preferire frasi coordinate a frasi subordinate

In seguito alle mie ricerche sono giunta alla conclusione che le frasi adatte alla traduzione automatica sono frasi semplici al presente e al passato prossimo (eventualmente anche al futuro se è necessario) con il soggetto in prima posizione e il verbo in seconda posizione. Dato che è poco probabile riuscire a scrivere un intero testo composto unicamente da frasi semplici, per quanto riguarda le frasi coordinate i tempi verbali più adatti da utilizzare sono sempre il presente e il passato prossimo. È necessario ripetere l'intera forma verbale in tutte le frasi e non inserire troppe informazioni tra un verbo e l'altro se essi dipendono dallo stesso soggetto citato una volta sola. Tra le frasi subordinate, le finali e le concessive risultano migliori rispetto alle temporali e alle causali. Ovviamente, più le frasi sono corte, meglio verranno tradotte.

Questo comporta però uno svantaggio a livello di corposità del testo: un testo composto da frasi corte e semplici non avrà lo stesso effetto di un testo più articolato e che presenta strutture subordinate. Tuttavia, se lo scopo finale è la traduzione automatica, vale la pena penalizzare un po' la corposità del testo di partenza in favore di una sua più accurata traduzione.

5. Inserire sempre gli articoli davanti ai sostantivi, in modo tale che i sostantivi siano riconoscibili

L’inserimento degli articoli davanti ai sostantivi è fondamentale in modo tale che il sistema di traduzione automatica non confonda il sostantivo con un'altra parte della frase, come successo ad esempio nella traduzione della frase seguente.

	Testo originale	Testo modificato il 02.04.2015	Testo modificato il 02.04.2015
Italiano	[...] consentendone un maggior <b>utilizzo</b> durante l'anno grazie all'energia solare.	<b>Utilizzo</b> annuale maggiore grazie all'energia solare	L'energia solare permette <b>un utilizzo</b> annuale maggiore
Francese	[...] permettant une plus grande <b>utilisation</b> toute l'année grâce à l'énergie solaire.	<b>Utilisez</b> les accrue grâce à l'énergie solaire annuels	L'énergie solaire permet <b>une utilisation</b> annuelle supérieure
Tedesco	[...] die in verstärktem Maße Jahres dank Solarenergie überall.	<b>Verwenden</b> Sie die jährlichen erhöht dank Solarenergie	Solarenergie ermöglicht <b>eine</b> höhere jährliche <b>Nutzung</b>
Inglese	[...] allowing greater <b>use</b> throughout the year thanks to solar energy.	<b>Use</b> the annual increased thanks to solar energy	Solar power allows <b>a</b> higher annual <b>use</b>

6. Evitare i pronomi e inserire sempre il soggetto della frase, se possibile al primo posto

L’uso dei pronomi è altamente sconsigliato: per quanto possa essere utile in un testo originale per evitare la ripetizione di un nome magari appena citato nella frase precedente, nella fase di traduzione automatica il pronome viene considerato esattamente come una particella traducibile, ma che non riprende nessun tipo di significato e che nella maggior parte dei casi fa risultare incorretta la frase. È quindi molto importante inserire il soggetto, se possibile all’inizio della frase.

	Testo originale	Testo modificato 02.04.2015
Italiano	Particolarmente adatti per riscaldare piscine scoperte ad uso estivo, <b>aumentano</b> di alcuni gradi la temperatura dell'acqua, [...]	<b>I collettori solari</b> sono adatti per il riscaldamento di piscine scoperte utilizzate in estate: • La temperatura dell'acqua aumenta di alcuni gradi [...]
Francese	Particulièrement adapté pour le chauffage des piscines extérieures pour l'été, <b>ont augmenté</b> de degrés la température de l'eau quelques- uns, [...]	Les capteurs solaires sont adaptés pour le chauffage des piscines extérieures utilisées en été: • La température de l'eau augmente de quelques degrés [...]
Tedesco	Besonders geeignet für die Heizung Außenpools für den Sommer, <b>erhöht</b> um ein paar Grad Wassertemperatur, [...]	Die Solarkollektoren sind für die Beheizung von Außenpools im Sommer genutzt: • Die Wassertemperatur steigt ein paar Grad [...]
Inglese	Particularly suitable for heating outdoor swimming pools for summer use, <b>increased</b> by a few degrees water temperature. [...]	The solar collectors are suitable for the heating of outdoor pools used in summer: • The water temperature increases a few degrees [...]

7. Collegare sempre, e soprattutto visibilmente, elementi che dipendono tra loro (ad es. preposizione + sostantivo, verbo + preposizione, sostantivo + aggettivo ecc.), evitando così collegamenti logici solo tramite un ragionamento umano

Se più parole o parti della frase dipendono dalla stessa preposizione, è necessario collegare le varie parti in modo «visibile» e non solo in modo logico. È ad esempio utile ripetere la preposizione davanti ad ogni parola o parte che dipende da essa. Il sistema di traduzione automatica non è in grado di comprendere il collegamento logico, ovvio per una persona, tra una preposizione e le parole che dipendono da quest'ultima.

Nel mio corpus era ad esempio presente una frase che conteneva «tubo di gomma, polietilene o metallico». In primo luogo in questo caso va corretto «metallico» con «metallo» se vogliamo legarlo a «tubo di» e in secondo luogo è necessario inserire la preposizione «di» davanti a «polietilene» e a «metallo». Solo in questo modo il sistema di traduzione automatica sarà in grado di riportare nel migliore dei modi il significato in un'altra lingua.

8. Se una preposizione dipende da un verbo presente nella frase, è necessario inserirla subito dopo il verbo

Nella fase di pre-editing ho notato che se all'interno della frase la preposizione retta dal verbo viene inserita (subito) dopo il verbo, la traduzione risultava maggiormente corretta. Questo e altri tipi di legame devono essere evidenti e il più semplice possibile nella frase.

Testo 2.1	Testo originale (r. 1)	Testo modificato il 14.04.2015
Italiano	DA UN'INTUIZIONE NASCE SOLARTEC	Solartec nasce da un'intuizione
Francese	Intuition BORN SOLARTEC	Solartec vient de l'intuition
Tedesco	Intuition BORN SOLARTEC	Solartec stammt aus einer Intuition
Inglese	Intuition BORN SOLARTEC	Solartec comes from an intuition

9. Se più aggettivi definiscono un sostantivo, è necessario ripetere il sostantivo davanti a ogni aggettivo

È fondamentale legare ogni singolo aggettivo al sostantivo che esso definisce. Un elenco di più aggettivi separati da virgole, che dipende da un unico sostantivo, non sarà riconosciuto tale da Google translate. La possibilità che il sistema riesca a tradurre gli aggettivi in modo corretto è più elevata se si traduce in francese poiché la struttura italiana sostantivo + aggettivo, + aggettivo rimane tale. In tedesco e in inglese, al contrario, gli aggettivi vengono sempre scritti prima del sostantivo e quindi vi è un maggior rischio che il primo aggettivo venga posto correttamente davanti al sostantivo, ma che gli altri vengano tradotti come aggettivi a sé stanti o come sostantivi.

	Testo originale	Testo modificato il 07.04.2015



<i>Italiano</i>	Tutti i moduli sono sottoposti ad un controllo visivo, meccanico ed elettrico.	Tutti i moduli sono sottoposti a vari controlli: <ul style="list-style-type: none"> <li>• un controllo visivo</li> <li>• un controllo meccanico</li> <li>• un controllo elettrico.</li> </ul>
<i>Francese</i>	Tous les modules sont soumis à une inspection visuelle, mécanique et électrique.	Tous les modules sont soumis à différents contrôles: <ul style="list-style-type: none"> <li>• une inspection visuelle</li> <li>• une commande mécanique</li> <li>• un contrôle électrique.</li> </ul>
<i>Tedesco</i>	Alle Module sind mit einer visuellen Inspektion, mechanische und elektrische unterworfen.	Alle Module werden verschiedenen Prüfungen unterzogen: <ul style="list-style-type: none"> <li>• eine visuelle Inspektion</li> <li>• eine mechanische Steuerung</li> <li>• eine elektrische Prüfung.</li> </ul>
<i>Inglese</i>	All modules are subjected to a visual inspection, mechanical and electrical.	All modules are subjected to various checks: <ul style="list-style-type: none"> <li>• a visual inspection</li> <li>• a mechanical control</li> <li>• an electrical check.</li> </ul>

10. Evitare i sinonimi e non temere le ripetizioni: usare le stesse parole per esprimere lo stesso concetto

Il perché di questa regola è molto semplice da comprendere: già Muegge era giunto a tale conclusione. La ricerca di parole conosciute e presenti nel linguaggio comune permette di aumentare la probabilità che il sistema di traduzione automatica traduca correttamente una determinata parola. È sbagliato pensare che due sinonimi in una lingua di partenza vengano tradotti con un unico termine nella lingua di arrivo, come magari farebbe un traduttore umano per rendere il testo maggiormente comprensibile e privo di incertezze. Essendo una macchina, è molto probabile che Google translate traduca parole diverse in modo diverso, creando così confusione. La ripetizione renderà il testo poco originale, ma al contempo la traduzione maggiormente precisa e il processo di traduzione semplice.

11. Evitare il gerundio, le forme passive, i verbi riflessivi e i verbi impersonali

Il gerundio è una forma verbale molto particolare: non esiste in tutte le lingue ed è quindi necessario evitare il suo uso. Vi sono vari modi per sostituire il gerundio.

	<i>Testo originale</i>	<i>Testo modificato il 02.04.2015</i>
<i>Italiano</i>	«I collettori solari per piscine TS POOL, <b>lavorando</b> a bassa temperatura, sono [...]»	I collettori solari sono [...], perché lavorano a basse temperature.
<i>Francese</i>	Les capteurs solaires pour piscines TS POOL, <b>travaillant</b> à basse température, [...]	Les capteurs solaires sont [...], parce qu'ils travaillent à des températures basses.
<i>Tedesco</i>	Die Solarkollektoren für Schwimmbäder TS POOL, <b>arbeitet</b> bei niedrigen Temperaturen, [...]	Sonnenkollektoren sind [...], weil sie bei niedrigen Temperaturen zu arbeiten.
<i>Inglese</i>	The solar collectors for swimming pools TS POOL, <b>working</b> at low temperature, [...]	Solar collectors are [...], because they work at low temperatures.

I verbi al gerundio vengono spesso tradotti in modo corretto solo in inglese, poiché corrispondono alle forme -ing, usate molto frequentemente.

La problematica delle frasi impersonali è invece legata a quella del soggetto: il sistema di traduzione automatica non traduce nessuna frase senza collegare il verbo a un soggetto e, se nel testo originale il soggetto non c'è, il sistema di traduzione proporrà una soluzione sicuramente sbagliata legando il verbo a un altro sostantivo o aggiungendo un pronome casuale. È quindi necessario evitare le forme impersonali e inserire sempre il soggetto all'interno della frase (ancora meglio se al primo posto). Nella frase «[...] per ogni kWh prodotto si risparmiano circa 250gr. di olio combustibile.» si riscontra il problema della forma impersonale e nelle traduzioni Google translate propone dei verbi principalmente al futuro: in francese «[...] pour chaque kWh produit

permettra d'économiser environ 250g. de fuel-oil.», in tedesco «[...] für jede kWh produziert wird ca. 250g speichern. Heizöl.» e in inglese «[...] for each kWh produced will save approximately 250g. of fuel oil.». Nella fase di pre-editing ho inserito il soggetto «sistema» e

ho mantenuto il contenuto originale della frase: «Il sistema risparmia 250 grammi di olio combustibile per ogni kWh.».

12. Favorire elenchi a frasi lunghe che presentano vari elementi suddivisi da virgole.

L'elenco puntato è uno degli strumenti introdotti da Nielsen nelle sue ricerche sulla scrittura per il web che ho sfruttato maggiormente nella fase di pre-editing per rendere più semplici frasi lunghe e ricche di virgole presenti nei miei testi. L'uso delle virgole non è infatti uguale in tutte le lingue; l'elenco puntato permette invece di trasformare una frase lunga in frasi più corte o in frasi nominali.

	Testo originale	Testo modificato il 07.04.2015
Italiano	Moduli fotovoltaici policristallini con alte prestazioni, celle di silicio solare (156.5 mm <sup>2</sup> ) , efficienza del modulo fino al 15,2 %.	<ul style="list-style-type: none"> <li>• moduli fotovoltaici policristallini con alte prestazioni</li> <li>• celle di silicio solare (156.5 mm<sup>2</sup>)</li> <li>• efficienza del modulo fino a 15,2 %.</li> </ul>
Francese	Modules photovoltaïques polycristallins avec des cellules de silicium de haute performance solaires (156,5 mm <sup>2</sup> ), module de rendement allant jusqu'à 15,2%.	<ul style="list-style-type: none"> <li>• Modules photovoltaïques polycristallins à hauteperformance</li> <li>• cellules solaires en silicium (156,5 mm<sup>2</sup>)</li> <li>• Module efficacité jusqu'à15,2%.</li> </ul>
Tedesco	Polykristalline Photovoltaik-Module mit Hochleistungs-Silizium-Solarzellen (156,5 mm <sup>2</sup> ), Modulwirkungsgrad von bis zu 15,2%.	<ul style="list-style-type: none"> <li>• PolykristallinePhotovoltaik-Module mit hoher Leistung</li> <li>• Silizium-Solarzellen (156,5mm<sup>2</sup>)</li> <li>• Modulwirkungsgrad bis zu15,2%.</li> </ul>
Inglese	Polycrystalline photovoltaic modules with high-performance silicon solar cells (156.5 mm <sup>2</sup> ), module efficiency of up to 15.2%.	<ul style="list-style-type: none"> <li>• Polycrystallinephotovoltaic modules with high performance</li> <li>• silicon solar cells (156.5mm<sup>2</sup>)</li> <li>• module efficiency up to15.2%.</li> </ul>

13. Evitare di inserire locuzioni temporali al primo posto nella frase

Sempre per permettere al soggetto di essere posizionato al primo posto nella frase, consiglio di posizionare le locuzioni temporali in mezzo o alla fine di una frase.

	Testo modificato il 07.05.2015	Testo modificato il 07.05.2015	Testo modificato il 07.05.2015	Testo modificato il 07.05.2015
Italiano	Negli ultimi anni Solterra realizza impianti fotovoltaici.	Solterra realizza impianti fotovoltaici dal 2010.	Solterra realizza impianti fotovoltaici da diversi anni.	Solterra realizza impianti fotovoltaici a partire dal 2010.
Francese	Au cours des dernières années Solterra fabrique des systèmes photovoltaïques.	Solterra fabrique des systèmes photovoltaïques depuis 2010.	Solterra fabrique des systèmes photovoltaïques depuis plusieurs années.	Solterra fabrique des systèmes photovoltaïques depuis 2010.
Tedesco	In den letzten Jahren Solterra fertigt Photovoltaikanlage n.	Solterra fertigt Photovoltaikanlag en seit 2010.	Solterra fertigt Photovoltaikanlag en für mehrere Jahre.	Solterra fertigt Photovoltaikanlage n seit 2010.

<i>Inglese</i>	In recent years Solterra manufactures photovoltaic systems.	Solterra manufactures photovoltaic systems since 2010.	Solterra manufactures photovoltaic systems for several years.	Solterra manufactures photovoltaic systems since 2010.
----------------	---	--	---	--

14. Evitare una terminologia troppo ricercata per esprimere un concetto che può essere spiegato con parole più semplici ed evitare formulazioni letterarie o a effetto che però non riportano nessun significato rilevante

La frase «Albasolar affonda le sue radici nel 1968, [...]» presente nel mio corpus è un tipico esempio di ciò che non va proposto a un sistema di traduzione automatica. Tutto ciò che crea colore e ricchezza lessicale va portato a un livello inferiore e sostituito da una terminologia più comune. In questo caso la frase potrebbe essere semplificata in ampio modo per risultare maggiormente adatta alla traduzione automatica, per esempio così: «Albasolar nasce nel 1968, [...]».

15. Evitare qualsiasi parola straniera

Per ovvi motivi, se possibile, le parole straniere vanno tralasciate. Probabilmente in futuro, tramite l’inserimento sempre più elevato di testi di vario genere, Google translate sarà in grado di comprendere che parole come ad esempio «governance» sono ormai diventate di uso comune anche in altre lingue, ma per ora consiglio di tralasciarle e di trovare corrispettivi nella lingua originale, per evitare problemi supplementari.

16. Evitare il non espletivo

Il *non* espletivo o pleonastico rappresenta la presenza di un *non* in una frase senza che ne alteri il significato. Il *non* è infatti facoltativo e ha una funzione non del tutto chiara, a volte fa risaltare la frase. Si ritrova molto spesso in proposizioni temporali introdotte da «finché». Nelle altre lingue questo non viene però tradotto come una negazione. È quindi importante evitare di scrivere frasi di questo tipo.

	<i>Frase</i>
<i>Italiano</i>	La produzione avverrà in Germania finché gli stabili in Svizzera <b>non saranno pronti</b> .
<i>Francese</i>	La production aura lieu en Allemagne jusqu'à ce que les bâtiments en Suisse <b>ne sera pas prêt</b> .
<i>Tedesco</i>	Die Produktion wird in Deutschland bis in die Gebäude in der Schweiz stattfinden <b>wird nicht bereit sein</b> .
<i>Inglese</i>	The production will take place in Germany until the buildings in Switzerland <b>will not be ready</b> .

## ***Bibliografia***

- Arnold, D. e. (1994). *Machine Translation: an introductory guide*. . Manchester: NCC Blackwell.
- Badesso, C. e. (2016). *Verifica dell'utilità della traduzione automatica e degli interventi di pre-editing e post-editing nella traduzione terminologica*. Winterthur: Tesi di Master non pubblicata - ZHAW – Università di Scienze applicate Zurigo.
- Calvino, I. (1965). *L'antilingua. Il giorno*.
- Chen, A. (2016). Talent Shortage and Talent Convergence. *Keynotes Portland Winter 2016* (pp. 27-30). Amsterdam: TAUS Signature Editions.
- DGT. (2017). *2017\_language\_industry\_survey\_report- Expectations and Concerns of the European Language Industry*.
- Eco, U. (2007). *Dire quasi la stessa cosa - Esperienze di traduzione*. Milano: Bompiani.
- Hujsen, W. O. (1998). Controlled language – An introduction. . *roc. of the 2nd Int. Workshop on Controlled Language Applications* . CLAW .
- Kittredge, R. (1985). The Significance of Sublanguage for Automatic Translation (pp. *Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation* (pp. 154-166). New York: Colgate University: Hamilton.
- Lehrberger, J. (1982). Automatic Translation and the Concept of Sublanguage. Dans R. K. di), *Sublanguage: Studies of Language in Restricted Semantic Domains. Library Edition/Foundation of Communication* (pp. 81-106). Berlin: de Gruyter.
- Lommel A.R., e. D. (2016, maggio 17). *MT's Journey to the Enterprise - The Evolution of Machine Translation into a mainstream solution*. Récupéré sur Common sense Advisory:

<http://www.common senseadvisory.com/AbstractView/tabid/74/ArticleID/36513/Title/MTsJourneytotheEnterprise/Default.aspx>

Plitt, M. e. (2010). A Productivity Test of Statistical Machine Translation Post-Editing in a in Typical Localization Context. *The Prague Bulletin of Mathematical Linguistics* ., pp. 7-16.

Pym, A. (2013, 3). "Translation skill-sets in a machine-translation age". *Meta* 58(3) , pp. 487-503.

Selinker, L. (1972, 10). Interlanguage . *International Review of Applied Linguistics*, p. 209-231.

*State of the Translation Industry 2009*. (2009). Tratto il giorno 10 13, 2012 da Gengo:

<http://gengo.com/express/report/translation-industry-2009/>

van der Meer, J. (2014, marzo 27). «*The Convergence Era, Translation as a utility.*» . Consulté le gennaio 26, 2017, sur Brighttalk. : <https://www.brighttalk.com/webcast/9273/76923>

Weaver, W. i. (1955). *Machine translation of languages*. Cambridge: MIT Press.