

HELLMUT RIEDIGER – GABRIELE GALATI

TECNOLOGIE PER LA TRADUZIONE

SOMMARIO

1.	L'INDUSTRIA DELLE LINGUE	4
	DALLA TRADUZIONE ALLA LOCALIZZAZIONE	4
	INTERNET E LE LINGUE	4
	TENDENZE DEL MERCATO	5
2.	STRUMENTI PER LA TRADUZIONE	7
	IL COMPUTER COME TRADUTTORE	7
	IL COMPUTER COME ASSISTENTE	7
	IL RITORNO DELLA MT	8
	MODELLI DI SISTEMI DI MT	8
	GLI STANDARD	9
3.	OPERE DI CONSULTAZIONE ONLINE	10
	DAL CD-ROM AL WEB	10
	DAL WEB AGLI SMARTPHONE	11
4.	GLOSSARIO ESSENZIALE	12
	Allineamento	12
	Banca dati terminologica	12
	Concordancer	12
	Corpus elettronico	12
	Corrispondenza esatta	13
	Corrispondenza parziale	13
	Déjà Vu	13
	Desktop-publishing	13
	Dizionario online	13
	Estrazione terminologica	14
	File bilingue	14
	Fuzzy match	14
	Grado di corrispondenza	14
	Google Traduttore	14
	IATE	15
	Internazionalizzazione	15
	LISA	15
	Localizzazione	16
	Logiterm	16
	Memoria di traduzione	16
	Multiterm	17
	OmegaT	17

OSCAR.....	17
Post - editing.....	17
Pre-editing	17
Pretraduzione.....	18
Revisione.....	18
Riconoscimento di termini.....	18
Ripetizioni esterne.....	18
Ripetizioni interne.....	18
Scheda terminologica (anche: voce terminologica).....	18
SDL TRADOS.....	18
Segmento.....	18
Tag.....	19
Termine.....	19
Term-mining.....	19
TMX (Translation Memory eXchange format).....	19
Traduzione assistita.....	19
Traduzione automatica.....	20
Trattamento automatico del linguaggio.....	20
XML (eXtensible Markup Language).....	20

Finito di redigere il 21.11.2012

1. L'INDUSTRIA DELLE LINGUE

L'industria delle lingue è il settore dedicato alla realizzazione e commercializzazione di prodotti, strumenti, servizi e tecnologie per la traduzione, la localizzazione di software, siti web e prodotti multimediali, la sottotitolazione e il doppiaggio, la gestione terminologica, la terminografia, la lessicografia e l'insegnamento delle lingue.

La traduzione come attività professionale e imprenditoriale è un fatto relativamente recente; solo a partire dalla prima metà del XX secolo, l'aumento degli scambi internazionali e la diffusione di prodotti industriali, tecnologie e servizi hanno via via accresciuto l'importanza dei servizi linguistici nella comunicazione tra aziende e istituzioni di paesi diversi.

Oggi con internet e i media digitali il multilinguismo rappresenta un'esperienza quotidiana per milioni di persone. Le situazioni di contatto aumentano e una delle poche barriere rimaste è proprio quella linguistica. Poiché i marchi si globalizzano, ma gli utenti si identificano meglio con essi quando sono adattati alle loro specificità linguistiche e culturali, il multilinguismo, la traduzione e/o la localizzazione hanno assunto un significato strategico.

DALLA TRADUZIONE ALLA LOCALIZZAZIONE

Di industria della lingua si inizia a parlare dagli anni '90. Lo sviluppo dell'informatica, la diffusione dei PC prima e l'esplosione di internet poi favorivano le occasioni di comunicazione interlinguistica e dall'altro cominciavano a porre problemi di adattamento e di traduzione per i produttori di software e hardware. Si sviluppò il cosiddetto settore GILT (Globalizzazione, Internazionalizzazione, Localizzazione, Traduzione): la *globalizzazione* richiede l'*internazionalizzazione*, cioè la progettazione dei prodotti per un potenziale uso al di fuori del mercato per cui sono stati progettati, per facilitarne la *localizzazione*, vale a dire l'adattamento alle esigenze e abitudini di utenti/clienti di altri paesi e regioni e la *traduzione* nelle loro lingue.

I prodotti oggetto di tali processi sono oggi i più diversi: oltre ai software e ai siti web questi vanno dalla pubblicità (televisiva, editoriale, ecc.) ai manuali d'uso di ogni genere, alle pubblicazioni mediche e scientifiche, agli strumenti medicali, ai farmaci, alle etichette dei prodotti venduti sul mercato internazionale e tanti altri.

Oltre alla traduzione di grandi volumi di testi in tempi stretti, i servizi e prodotti dell'industria della lingua comprendono la localizzazione di applicazioni informatiche, media digitali, prodotti multimediali (CD-ROM, DVD, videogiochi) e siti web, la gestione di eventi multilingui, la gestione di banche dati terminologiche e di piattaforme per l'apprendimento linguistico. A questi si accompagnano lo sviluppo di dizionari elettronici, strumenti e programmi per la traduzione assistita e automatica, la revisione linguistica, la localizzazione, la terminologia/terminografia, il doppiaggio, la sottotitolazione scritta e quella orale attraverso sistemi di riconoscimento vocale (*respeaking*).

INTERNET E LE LINGUE

Nell'aprile 2011, secondo le rilevazioni dell'internet services provider inglese Netcraft (<http://news.netcraft.com/archives/category/web-server-survey>), nel mondo risultavano circa 313 milioni di siti, di cui oltre un terzo attivi, con varie decine di miliardi di pagine web accessibili ai motori di ricerca.

Nell'era digitale, senza dubbio, l'inglese ha accentuato il suo carattere di lingua franca, soprattutto per la comunicazione tecnica e scientifica. Tuttavia uno dei fenomeni più rilevanti è l'espansione del traffico in linea in ambiti geografici in cui la lingua inglese (o americana) è meno conosciuta e molti segnali lasciano intravedere che la posizione dominante dell'inglese su internet ha iniziato a erodersi (cfr. myGengo, 2009, p.2). Mentre nel 1998 le pagine in lingua inglese ammontavano al 75% del totale, nel 2005 erano già scese sotto il 45% (Unione latina, 2005).

Secondo la rilevazione periodica della società di ricerca Internet World Stats (IWS) (<http://www.internetworldstats.com/stats7.htm>) sulla suddivisione per lingua madre della popolazione online,

a metà del 2010 a livello mondiale risultavano connessi quasi 2 miliardi di utenti. Di questi «solo» il 27,3 % è di lingua inglese, il 22,% di lingua cinese, il 7,8% di lingua spagnola, il 5,0% di lingua giapponese, il 4,2% di lingua portoghese, il 3,8% di lingua tedesca, il 3% di lingua francese, il 2,6% di lingua araba, il 2% di lingua coreana. La percentuale relativa alla lingua italiana che ancora nel 2008, con il 2,4% figurava tra le prime 10 a livello mondiale è scesa sotto il 2,0%.

La tabella seguente riassume i dati delle rilevazioni di Internet World Stats relative al numero degli utenti delle prime dieci lingue di internet dal 2004 al 2010 e comprende anche il T-index, indice elaborato dalla società di traduzioni Translated.net, basato sulla combinazione tra la popolazione internet e il relativo PIL pro capite, che aiuta a identificare le lingue utilizzate dagli utenti interessati ad acquistare servizi, prodotti o spazi pubblicitari su un determinato sito.

	Utenti di Internet suddivisi per lingue				T-Index
	2004	2006	2008	2010	2010
Totale	785.710.022	1.022.863.307	1463.632.361	1.966.514,816	
inglese	37,3 %	↓ 30,6 %	↓ 29,4%	↓ 27,3%	35,1%
cinese	12,5 %	↑ 13,0 %	↑ 18,9%	↑ 22,6%	10,7%
spagnolo	6,8 %	↑ 7,9 %	↑ 8,5%	↓ 7,8%	7,0%
giapponese	8,2 %	↑ 8,5 %	↓ 6,5%	↓ 5,0%	7,0%
francese	4,2 %	↓ 4,0 %	↑ 4,7%	↓ 3,0%	4,6%
tedesco	6,7 %	↓ 5,0 %	↓ 4,2%	↓ 3,8%	5,9%
arabo	nd	nd	↑ 4,1%	↓ 3,3%	2,6%
portoghese	2,8 %	↑ 3,2 %	↑ 4,0%	↑ 4,2%	3,5%
russo	nd	↑ 2,3 %	nd	↑ 3,0%	3,3%
coreano	3,7%	↓ 3,3 %	↓ 2,4%	↓ 2,0%	2,4%
italiano	3,6%	↓ 2,8%	↓ 2,4%	nd	2,5%
nederlandese	1,7%	nd	nd	nd	1,6%

Le tendenze più chiare sono insieme alla crescita esponenziale del numero degli utenti, il netto calo della percentuale dell'inglese (-10% dal 2004), la diminuzione più o meno contenuta delle lingue degli altri principali paesi industrializzati (giapponese, tedesco, francese e italiano), e la netta crescita del cinese (+10%).

TENDENZE DEL MERCATO

Dagli anni '90 il settore dell'industria della lingua cresce a una media annua del 10 % circa. Nel 2010, p. es., le oltre 25 000 imprese fornitrici di servizi linguistici (*Language service providers, LSPs*) attive nel settore hanno generato a livello mondiale un fatturato di oltre 26 miliardi di dollari (oltre 43 % in Europa) e si prevede una crescita a oltre 38 miliardi nel 2015 (commonsenseadvisory, 2010). Uno studio, invece, condotto per conto della Commissione europea ha previsto in Europa un aumento da 8,4 miliardi di euro nel 2008 (5,77 miliardi per traduzione, interpretazione, localizzazione di software e globalizzazione di siti web, 568 milioni per tecnologie linguistiche, 633 milioni per il settore della sottotitolazione e il doppiaggio, 1,6 milioni per l'insegnamento delle lingue e 143 milioni per l'organizzazione di conferenze multilingui) a 16,5 - 20 miliardi di euro nel 2015 (Commissione europea, 2009, p.3). L'Italia copre all'incirca un decimo del fatturato europeo.

Dal 2006 l'attività del settore è normata dalla norma ISO 15038, ovvero norma UNI EN 15038 «Servizi di traduzione - Requisiti del Servizio» nella versione italiana.

Un fenomeno rilevante, legato alla diffusione del web 2.0 e dei social media è lo sviluppo di un universo della traduzione volontaria e gratuita a fianco di quella della traduzione commerciale. In particolare in contesti come quelli dei social network, del settore no profit e dell'open source dove gli utenti hanno un interesse personale a contribuire a migliorare il servizio, si diffondono anche nel campo della traduzione pratiche di crowdsourcing.

Sitografia

myGengo, *State of the translation industry 2009*.

<http://mygengo.com/report/translation-industry-2009> (23.10.2012)

Unione latina, *Langues et cultures sur la Toile - étude 2005 - Analyse détaillée de l'espace anglophone, 2005*

<http://dtil.unilat.org/LI/2005/fr/espace%20anglophone.htm> (23.10.2012)

Commonsenseadvisory. The Language Services Market: 2010, 2010, <http://www.commonsenseadvisory.com> (23.10.2012)

Commissione europea, Direzione Generale Traduzione. *Study on the size of the language*.

http://bookshop.europa.eu/is-bin/INTERSHOP.enfinity/WFS/EU-Bookshop-Site/en_GB/-/EUR/ViewPublication-Start?PublicationKey=HC8009985 (23.10.2012)

2. STRUMENTI PER LA TRADUZIONE

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.

IL COMPUTER COME TRADUTTORE

L'applicazione della tecnologia informatica al campo della traduzione ha una lunga storia. Nel 1949 Warren Weaver pubblicò un celebre memorandum in cui delineò le prospettive della MT e negli anni '50 si diffuse una grande euforia. Eventi come il *Georgetown experiment* del '54, quando con un vocabolario di sole 250 parole e sole 6 regole grammaticali fu tradotto in inglese un campione selezionato di frasi russe, fecero pensare che la sostituzione dei traduttori umani con i computer fosse solo una questione di tempo. Ma nel giro di pochi anni gli entusiasmi si smorzarono poiché ci si rese conto che era impossibile raggiungere una traduzione automatica di alta qualità (*fully automatic high quality translation*). Nel 1966 un famoso rapporto dell'ALPAC (Automatic Language Processing Advisory Committee) considerò i sistemi di MT costosi e fondamentalmente inutili perché la macchina non era in grado di imitare le procedure del traduttore umano. La ricerca rallentò e la MT rimase confinata nell'ambito di grandi istituzioni come l'Aeronautica degli Stati Uniti che a partire dal 1970 adottò il sistema di traduzione Systran o la CEE che nel 1975 approvò un piano che oltre all'uso di Systran prevedeva la progettazione a lungo termine del sistema di MT su larga scala EURO-TRA.

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 nel primo decennio del 2000, videro la luce diversi software economici oppure gratuiti come Wordfast (1999), Omega T (2002), Memo Q (2009) o la Translator's Toolkit di Google (2008).

Oggi i sistemi di traduzione assistita sono diventati uno degli strumenti fondamentali dell'industria delle lingue. I più diffusi sistemi CAT sono SDL Trados, Star Transit, Across, Atril Déjà Vu, incalzati da WordFast, Omega T e Memo Q. Essi comprendono programmi per la gestione di memorie di traduzione, gestione della terminologia, l'allineamento dei file, la gestione del progetto. A questi si aggiungono speciali applicazioni per la localizzazione di software come SDL Passolo.

IL RITORNO DELLA MT

Negli anni '90 si registrò un rinnovato interesse anche per la MT quando, con ridotte aspettative, i sistemi furono convertiti in prodotti per PC (p. es. SYSTRAN, Globalink, Intergraph e Logos) a prezzi accessibili al vasto pubblico. La diffusione del web portò i traduttori automatici on-line, soprattutto per soddisfare il bisogno di traduzioni immediate per le applicazioni di Internet, come la posta elettronica e le pagine web, 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 Babel Fish seguito poi da vari altri servizi come Babylon (<http://italiano.babylon.com>), Google Traduttore (<http://translate.google.com>), Bing Translator (<http://www.microsofttranslator.com>), SYSTRANet (<http://www.systranet.com/translate>), Reverso (<http://www.reverso.net>) e Worldlingo (<http://www.worldlingo.com>).

Servizi di MT sono oggi integrati negli elaboratori di testi come Word e disponibili anche per i telefoni cellulari. Google Translate per Android, per esempio è utilizzabile come traduttore simultaneo live per tradurre in tempo reale le frasi pronunciate dall'utente.

Oggi i sistemi di MT sono sempre più usati da fornitori di servizi linguistici e servizi di traduzione di istituzioni pubbliche e private. Spesso si sottopone il testo da tradurre al pre-editing ricorrendo a un linguaggio controllato, semplificando la grammatica e restringendo il lessico, oppure si sottopone a revisione umana il testo tradotto in modo automatico (post-editing). Sempre più frequentemente i sistemi di MT sono integrati nei sistemi CAT.

MODELLI DI SISTEMI DI MT

I principali paradigmi della MT sono:

- sistemi basati su regole
- sistemi statistici
- sistemi ibridi

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 MT basata su regole fornisce una buona qualità nelle situazioni generiche (cioè, non specifiche di un dominio) ed è per sua natura prevedibile.

Nei sistemi statistici, come Google traduttore, si fa ricorso a parametri statistiche che derivano dall'analisi di corpora monolingui o bilingui per l'attribuzione dell'ordine delle parole e del migliore candidato per la traduzione. Non vengono utilizzate regole grammaticali, ma si usano le probabilità di occorrenza e di frequenza delle parole. 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. La traduzione è fluida, e i risultati dei sistemi di impronta statistica risultano nel complesso migliori rispetto a quelli che fanno ricorso a regole.

I più recenti sistemi di MT sono spesso ibridi (per esempio Systran o Bing Translator), e combinano le metodologie basate su regole con quelle statistiche con lo scopo di sviluppare diverse strategie di risoluzione a seconda dello specifico problema traduttivo da affrontare.

GLI STANDARD

Lo standard più diffuso è quello per l'interscambio delle memorie di traduzione TMX (*Translation Memory eXchange*), supportato da quasi tutti gli strumenti CAT/TM. XLIFF (*XML Localisation Interchange File Format*), invece, è uno standard specifico per il settore della localizzazione.

Per quanto riguarda database lessicali e terminologici, il panorama è più confuso. Coesistono, tra gli altri, i seguenti standard: TBX (*Term Base eXchange*), XLT (*XML representation of Lexicons and Terminologies*), OLIF (*Open Lexicon Interchange Format*) e MARTIF (*MAchine-ReadableTerminology Interchange Format*).

Sebbene alcuni di questi standard siano supportati da alcuni strumenti CAT/TM, siamo lontani dalla definizione di uno standard unico generalmente accettato e utilizzato.

Sitografia

<http://italiano.babylon.com>

<http://translate.google.com>

<http://www.microsofttranslator.com>

<http://www.systranet.com/translate> <http://www.reverso.net>

<http://www.worldlingo.com>

3. OPERE DI CONSULTAZIONE ONLINE

In rete sono disponibili numerose opere di consultazione quali enciclopedie, dizionari generali e specialistici, banche dati terminologiche, testuali e statistiche ecc. Sono realizzate da case editrici, istituzioni pubbliche o autori privati, con caratteristiche molto eterogenee, sia per quanto riguarda la struttura, sia per quanto riguarda le modalità di funzionamento.

DAL CD-ROM AL WEB

Negli anni '90 del secolo scorso, con l'affermarsi del computer come strumento di lavoro, di studio e di svago si impennò la domanda di contenuti veicolati attraverso i monitor. In rete comparirono i primi dizionari e le prime enciclopedie online, tra cui merita di essere ricordato il progetto Encarta di Microsoft. Internet cresceva, ma era ancora ostacolata dalla lentezza delle connessioni, e quindi, allorché, a metà degli anni '90, apparvero le prime enciclopedie multimediali, il supporto privilegiato dall'editoria tradizionale per le opere di consultazione divenne il Cd-Rom. Tra la fine degli anni '90 e i primi anni del nuovo secolo con la diffusione dei DVD si pensò di aver risolto anche il problema dello spazio disponibile, mentre contemporaneamente sembrava che internet, nel frattempo velocizzata dall'avvento delle connessioni ADSL, potesse diventare un'estensione dei supporti ottici.

Tuttavia con l'aumento vertiginoso degli utenti della rete, le case editrici, benché in molti casi disorientate su come fare per garantirsi un ritorno economico, si resero conto che né i CD-rom né i DVD-rom avrebbero più avuto un futuro, ma che questo apparteneva al web, visto che ormai risultava molto più agevole consultare un'enciclopedia o un dizionario online, piuttosto che inserire un disco nel computer.

Nel web di lingua italiana, il primo a percorrere questa strada fu l'editore De Agostini, che, con il portale Sapere, mise a disposizione un vasto corpus organico di conoscenze in lingua italiana. Ma poiché i tempi di internet sono sempre più veloci di quelli dell'editoria, mentre l'editoria tradizionale si stava impegnando a sviluppare opere di consultazione online con accesso a pagamento, ecco partire la sfida del web 2.0 e delle opere collaborative, nonché degli stessi motori di ricerca generalisti come Google, che iniziavano a sviluppare propri dizionari e servizi di ricerca enciclopedica. Tra il 2003 e il 2004 si assistette allo sviluppo vertiginoso di Wikipedia, presto seguita da altri simili progetti ad accesso gratuito che invitano gli utenti a contribuire alla compilazione.

Questi sviluppi rendono l'attuale panorama delle opere di consultazione online alquanto variegato. Vi sono editori che consentono la consultazione online gratuita, integrale o parziale, dei loro dizionari e delle enciclopedie per ottenere una maggiore visibilità per i loro prodotti su carta, su CD-ROM o DVD, ma anche per i siti che li ospitano (per esempio di dizionari della Rizzoli e della Hoepli sul sito del Corriere della Sera e della Repubblica). Sul versante della consultazione a pagamento, invece, si sviluppano offerte sempre più articolate come per esempio quella della piattaforma eLexico che attraverso l'acquisto di un abbonamento permette di consultare online oppure di scaricare numerosi dizionari degli editori Hoepli, RCS e Le Monnier oppure quella di Zanichelli, il cui abbonamento consente di consultare una ventina di opere del catalogo. Poi ci sono case editrici, che invece raccogliendo la sfida del web 2.0, hanno scelto di aprirsi a forme di partecipazione dal basso e di collaborazione da parte degli utenti come l'Enciclopedia Britannica, la francese Larousse e l'italiana Enciclopedia Treccani. Quest'ultima ha trasformato il suo sito in un grande portale in cui offre gratuitamente dizionari ed enciclopedie, fino a poco tempo fa accessibili solo in volumi cartacei, con l'aggiunta di altre risorse informative e servizi per sperimentare nuove forme di comunicazione partecipativa, come la possibilità di inserire commenti e tag nelle voci enciclopediche o nei contenuti pubblicati sul sito, condividere contenuti con altri utenti, proporre nuove voci da inserire, segnalare neologismi e sinonimi regionali ecc.

Strumenti che sfruttano appieno le caratteristiche del web, per lo più realizzati da docenti, professionisti o appassionati, sono invece i cosiddetti dizionari multisorgente, metadizionari o aggregatori, come Onelook (<http://www.onelook.com/>), Term-minator (<http://term-minator.it>) o Webster Rosetta Edition (<http://www.websters-online-dictionary.org>). Queste risorse sono in genere costituite da interfacce, che per

esempio consentono la contemporanea consultazione di diverse fonti, quali vocabolari, enciclopedie o corpora testuali.

DAL WEB AGLI SMARTPHONE

Visto che negli ultimi anni il web dal computer ha raggiunto anche i dispositivi mobili, ecco che anche per le opere di consultazione si avvia una nuova metamorfosi. Gli smartphone in generale e i dispositivi mobili di Apple in particolare, stanno diventando una nuova piattaforma di sviluppo per gli editori di opere di consultazione. Nell'editoria di lingua italiana, Zanichelli, che ha già un buon catalogo di opere per Kindle (l'eBook di Amazon), offre ora numerose opere create per iPhone/iPad. Le Monnier invece ha pubblicato *Il Devoto-Oli* per iPhone, come ha fatto anche Garzanti con la *Garzantina*.

4. GLOSSARIO ESSENZIALE

Allineamento

L'allineamento di file crea da coppie di file di testo simili o uguali, ma in lingue diverse, un unico file bilingue.

I programmi di allineamento procedono alla segmentazione dei testi e al successivo confronto dei singoli segmenti nelle due lingue. L'operazione viene svolta in modo automatico dal programma, che propone una prima associazione di segmenti in base al loro formato, struttura e posizione nel testo. La persona del traduttore ne valuta la correttezza e può decidere per la conferma o per la modifica delle unità traduttive proposte. Alla fine del lavoro di revisione si avrà un unico file bilingue, che potrà essere utilizzato per creare una memoria di traduzione. I più recenti programmi propongono il salvataggio del file bilingue in formato standard XLIFF, basato su XML, che ne garantisce l'interoperabilità.

Banca dati terminologica

Raccolta di termini, provenienti da settori diversi, curati da organismi nazionali e internazionali, aziende, università ecc., consultabili via internet. Solitamente i termini sono presentati in schede terminologiche che contengono informazioni relative alla gestione delle schede e al termine.

Per la gestione delle schede si ha, di solito, nome del compilatore, data di compilazione o dell'aggiornamento. Per il termine, invece, oltre al termine stesso, si hanno informazioni sulla lingua, la categoria grammaticale, settore di appartenenza, eventuali sinonimi, dominio di impiego, definizione, contesto definitorio, note, fonti e grado di affidabilità. Queste raccolte terminologiche sono database informatici, in genere gestiti e alimentati all'interno dell'intranet di un'organizzazione, e successivamente resi pubblici. In internet, sono disponibili numerosi servizi per la consultazione di importanti raccolte terminologiche in ambito specialistico. IATE, per esempio, è la banca dati dell'Unione europea, con schede in 23 lingue. Il rapido sviluppo di questo tipo di offerta testimonia del valore strategico delle risorse terminologiche come capitale intellettuale e fattore di sviluppo nell'economia globale. Per assicurare l'interoperabilità dei dati nella gestione dei database terminologici, lo standard raccomandato è TBX, basato su XML.

Concordancer

È un programma che serve per la ricerca e l'analisi linguistica in un corpus, ovvero a creare in modo automatico concordanze. È usato, cioè, per verificare le relazioni tra le parole di un testo e dare accurate informazioni circa il modo in cui sono usate nel loro contesto.

Può essere usato per l'analisi di corpora paralleli, bilingui, o anche come programma all'interno di un sistema di memoria di traduzione, per verificare come una parola o parte di una frase sono state tradotte in contesti simili. Un concordancer, di norma, analizza un corpus (un'ampia raccolta di testi su un argomento, di solito salvata in formato solo testo) dando la possibilità di cercare combinazioni di parole all'interno di una gamma specificata o anche solo parti di parole. Una volta terminato, propone una lista di parole ordinate alfabeticamente, inserite in una frase di contesto, o altri dati di carattere linguistico. Questi dati possono essere usati in vario modo, come, per esempio, per studiare le collocazioni (gli abbinamenti frequenti di parole), verificare l'uso delle preposizioni o più semplicemente la frequenza delle occorrenze di una data parola. Tra i più comuni concordancer, usati in ambito linguistico, ricordiamo: AntConc, ApSIC Xbench, BootCat, CorpusEye, GlossaNet, MonoConc e WordSmith.

Corpus elettronico

Collezione di testi digitalizzati raccolti per lingue, per ambiti settoriali (letteratura, filosofia, scienze, stampa ecc.) o periodi storici, selezionati e organizzati per facilitare le analisi linguistiche. Attraverso strumenti di analisi quantitativa e statistica, un c.e. permette di indagare le regolarità linguistiche che emergono dai testi e che costituiscono la base per la descrizione della struttura del linguaggio. La disciplina che studia il linguaggio attraverso i corpora è nota come →Linguistica dei corpora. Oltre che dagli studiosi di linguistica, i corpo-

ra sono usati frequentemente nell'ambito della lessicografia, della traduzione o dell'insegnamento delle lingue. Esempi noti di c.e. consultabili via web sono il →British National Corpus per l'inglese o il corpus riferito al quotidiano "La Repubblica" per l'italiano.

Corrispondenza esatta

Indica il massimo grado di somiglianza tra due segmenti. durante la traduzione interattiva con programmi di →traduzione assistita (CAT). E' il risultato della ricerca nella →memoria di traduzione di una →corrispondenza al →segmento da tradurre tra i segmenti già tradotti.

Una corrispondenza esatta indica, quindi, la perfetta identità tra due segmenti in lingua originale. La relativa traduzione, proposta dal programma, non necessita, quindi, di modifiche. In alcuni programmi, la corrispondenza esatta viene ulteriormente qualificata da indicazioni di contesto. Se, per esempio, la corrispondenza esatta, trovata nella memoria, riguarda anche il segmento immediatamente precedente, allora aumenta il grado di affidabilità della traduzione proposta e viene chiamata ICE, In Exact Context match, oppure Context match o Guaranteed match.

Corrispondenza parziale

Indica una parziale somiglianza tra due segmenti. durante la traduzione interattiva con programmi di →traduzione assistita (CAT). E' il risultato della ricerca nella →memoria di traduzione di una →corrispondenza al →segmento da tradurre. È il contrario di →corrispondenza esatta.

Una corrispondenza parziale (fuzzy match) tra due segmenti in lingua originale viene indicata generalmente con un valore in percentuale: man mano, che scende il valore diminuisce la somiglianza. Viene data entro il valore di soglia, impostato dall'utente. I segmenti possono differire, per esempio, per alcune parole, per un ordine diverso delle stesse, per dei dati numerici in esso contenuti o, anche, solo per diversa formattazione. La relativa traduzione, proposta dal programma, necessita di un adattamento, se non di una riscrittura, nei casi di percentuali di corrispondenza troppo basse.

Déjà Vu

Applicazione per la traduzione assistita (CAT) della società Atril di Madrid. Non molto diffuso tra le agenzie, ma molto apprezzato tra i freelancer, per la compatibilità con →SDL Trados, leader del mercato, e per la facilità d'uso.

Offre un ambiente di lavoro unificato per la gestione di database di memoria, terminologici e di progetto. Progetti complessi, costituiti da un gran numero di file, anche in formati diversi, possono essere visualizzati in un unico file in una unica finestra, dove si possono applicare operazioni di ricerca e sostituzione, filtro e propagazione. Con la funzione EBMT (example-based machine translation, traduzione automatica basata su esempi) un →fuzzy match può essere integrato con il supporto fornito dalle basi dati terminologiche, arrivando, in molti casi, alla traduzione corretta.

Desktop-publishing

Editoria elettronica; impiego dei computer per le attività di composizione e impaginazione di testi e grafica. Le più avanzate applicazioni software per il DTP consentono, tra l'altro, di realizzare testi combinando diversi font e misure dei caratteri.

Dizionario online

Raccolte di vocaboli, termini di settori specifici o nozioni relative a particolari materie consultabili in rete. Queste risorse possono essere molto eterogenee, sia per quanto riguarda la struttura, sia per quanto riguarda le modalità di funzionamento e di consultazione. Si va dalla semplice scansione di opere cartacee o elenchi di termini in ordine alfabetico, alla trasposizione online di dizionari su cd-rom, alle →banche dati terminologiche fino ai cosiddetti metadizionari, motori di ricerca che consultano simultaneamente numerosi altri dizionari. Molti sono gratuiti, mentre altri sono consultabili previo acquisto di un abbonamento (per esempio

Zanichelli o la piattaforma eLexico). Sull'onda del web 2.0 si diffondono sempre più i cosiddetti dizionari collaborativi che, seguendo la strada tracciata da Wikipedia, invitano gli utenti ad aggiungere lemmi, definizioni o traduzioni.

Estrazione terminologica

Estrazione di un elenco di termini potenziali, o candidati, da un insieme di testi in formato digitale allo scopo di creare o allargare un glossario terminologico. Si realizza in modo semi-automatico con applicativi specializzati, con l'attenta revisione di un terminologo o traduttore.

L'estrazione di termini può essere sia monolingue sia bilingue. Nel primo caso si analizzano testi in una sola lingua, nel secondo caso vengono analizzati testi bilingui, a partire da quelli in lingua di partenza, con lo scopo di identificare i potenziali termini e i loro equivalenti, con una successiva procedura di →allineamento. Gli applicativi di tipo linguistico cercano i termini candidati in combinazioni di parole che corrispondono a strutture del discorso (per es., sostantivo + aggettivo o sostantivo + sostantivo), lavorando su una sola lingua o su lingue estremamente affini. In quelli di tipo statistico s'identificano, invece, le sequenze ripetute di segmenti lessicali. Si ha, però, una gran quantità di noise (segnalazioni non idonee) e di silence (termini non identificati). In genere i due tipi vengono combinati in applicativi ibridi. Difficoltà possono sorgere dall'estrarre termini senza un ampio contesto, che possa fornire informazioni sul concetto in questione. Un'ulteriore difficoltà deriva dal fatto che l'estrazione terminologica parte dal termine per descriverne il concetto legato ad esso (approccio semasiologico), mentre nella terminologia si segue un approccio onomasiologico, cioè, partendo da un dato concetto, se ne descrivono i termini.

File bilingue

È un tipo di file intermedio creato durante il processo di traduzione con programmi di traduzione assistita (CAT). Contiene i segmenti in lingua di partenza e d'arrivo e include, nella maggior parte dei casi, la struttura del documento originale.

I file da tradurre, che possono essere in origine nei più vari formati di file, sono convertiti in formato bilingue per essere letti dagli strumenti CAT. Dopo la traduzione, di solito, si preferisce provvedere alla revisione della traduzione sul file bilingue, prima di produrre il file tradotto e importarlo nella memoria di traduzione definitiva. I formati di file bilingui di tipo proprietario, a lungo ostacolo alla compatibilità tra i vari prodotti, sono stati sostituiti dallo standard XLIFF, basato su XML.

Fuzzy match

Corrispondenza parziale tra due segmenti di testo, ossia corrispondenza non esatta tra un segmento di partenza da tradurre ed un segmento di partenza archiviato nella memoria di traduzione.

Grado di corrispondenza

Grado di somiglianza, espresso in percentuale, tra un segmento del testo di partenza che deve essere tradotto ed una unità di traduzione recuperata dalla memoria di traduzione.

Google Traduttore

Interfaccia in italiano di Google Translate, il servizio di →traduzione automatica offerto dalla società Google Inc., per tradurre in modo istantaneo parole, termini, frasi o intere pagine web tra decine di coppie linguistiche.

Dal 2007 G.T. usa un software di traduzione proprio basato su un approccio di traduzione chiamato →traduzione automatica statistica. Ciò significa che il sistema cerca modelli in centinaia di milioni di documenti per scegliere la migliore traduzione da offrire all'utente. Individuando i modelli in documenti che sono già stati tradotti da esseri umani, G.T. è in grado di fare congetture intelligenti sulla presumibile correttezza di una traduzione.

Di G.T. fa uso anche Google Translator Toolkit, un servizio di →traduzione assistita, lanciato nel 2009, progettato per permettere a chiunque di correggere o integrare più facilmente le traduzioni che G.T. genera automaticamente ed essere di ausilio a chi traduce. Google Translator Toolkit permette ai traduttori di organizzare il loro lavoro e usare come basi una memoria di traduzione globale, in cui confluiscono le traduzioni validate dai singoli utenti, →memorie di traduzione personali in formato →TMX e glossari. Il sistema permette di caricare e tradurre documenti nei formati Microsoft Word, OpenOffice, RTF, HTML, testi in formato .txt e articoli di Wikipedia e di Knol.

HTML (Hypertext Markup Language)

Linguaggio di marcatura utilizzato per definire il formato dei documenti destinati ad essere pubblicati sul World Wide Web. Le pagine web sono costruite con i tag HTML, ossia simboli di marcatura inseriti nel testo. Il linguaggio HTML definisce il layout della pagina, i font, gli elementi grafici ed i collegamenti ipertestuali con altri documenti disponibili in rete.

IATE

È la banca dati terminologica multilingue dell'Unione europea, redatta in tutte le 23 lingue ufficiali. Raccoglie la terminologia e il gergo specifico delle istituzioni comunitarie, nonché di settori più generali, come per es. diritto, agricoltura, tecnologie dell'informazione, ecc.

Ha iniziato ad operare nel 1999 con l'obiettivo di fornire un'infrastruttura basata sul web per tutte le risorse terminologiche comunitarie e favorirne l'uso e la standardizzazione. Incorpora i dati delle precedenti banche dati terminologiche delle istituzioni europee: EURODICAUTOM (Commissione), EUTERPE (Parlamento), TIS (Consiglio), EUROTERMS (Centro traduzioni), CDCTERM (Corte dei conti). Contiene 8,4 milioni di termini, di cui circa 540000 abbreviazioni e 130000 frasi. Costantemente aggiornata, la banca dati è liberamente consultabile in Rete.

Internazionalizzazione

È l'adattamento o ingegnerizzazione di prodotti digitali per un potenziale utilizzo al di fuori del mercato per cui sono stati progettati, rendendo tecnicamente facile la →localizzazione degli stessi. Spesso indicato con *I18n*, dove 18 sono le lettere sottintese per *Internationalization*.

Normalmente si procede separando i dati testuali dai dati dipendenti dall'ambiente di utilizzo, ovvero dal codice del programma. In questo modo la futura attività di localizzazione riguarderà solo una parte delle risorse del programma, separate dal codice del programma. In fase di sviluppo, per es. di un' applicazione, si deve prevedere che in fase di esecuzione dell'applicazione stessa vengano selezionate le corrette risorse linguistiche richieste dall'utente. Una applicazione internazionalizzata deve poter modificare il suo aspetto non solo per la lingua utilizzata nell'interfaccia, ma anche il suo comportamento (modalità di formattazione dell'output), in funzione dell'area linguistica in cui il programma viene mandato in esecuzione. Per es., per quelle lingue, dove il verso di scrittura è inverso rispetto alle lingue occidentali, anche le finestre di interfaccia dovranno adattarsi graficamente a questa necessità. Allo stesso modo, specialmente per le lingue dell'est e quelle asiatiche, l'ambiente di sviluppo deve prevedere un corretto supporto della codifica Unicode, per una corretta rappresentazione dei caratteri. In genere i sistemi operativi e gli ambienti di sviluppo prevedono alcune librerie particolari per il supporto internazionale.

LISA

Acronimo per Localisation Industry Standards Association. Fondata nel 1990 e chiusa nel 2011, LISA è stata il principale forum internazionale per le organizzazioni operanti nel mercato globale. È fonte autorevole per le procedure di localizzazione e pubblicazione di documenti multilingue per il settore IT.

Vi aderivano oltre 240 importanti aziende ICT, fornitori di soluzioni globali e un numero crescente di società, che operano su mercati verticali, ma con una forte presenza strategica internazionale orientata al web, oltre a enti governativi, ministeri e organizzazioni del commercio. Tra gli aderenti si contavano aziende co-

me, per es., Adobe, Hewlett Packard, Xerox, Nokia ma anche Coca Cola, McDonald, oltre a istituzioni e organizzazioni come Banca mondiale, Unione europea, ISO, ma anche enti di formazione e associazioni di traduttori, ecc. Piuttosto che agire come un ente di normazione tradizionale a sé, i membri si consideravano come un gruppo orientato all'attuazione, concentrandosi sul miglioramento dei risultati e dei costi per le attività di localizzazione. Quattro volte l'anno, il Forum LISA riuniva i professionisti della localizzazione per discutere di costi di prodotto e di traduzione, controllo di qualità e procedure di internazionalizzazione. Gruppi di speciale interesse (detti SIG) erano incaricati di elaborare, per settori individuati, tecnologie standard. Tra questi ricordiamo →OSCAR, per l'elaborazione di standard per tecnologie di traduzione.

Localizzazione

Rendere un prodotto digitale linguisticamente e culturalmente appropriato per la versione locale di destinazione (paese/regione e lingua), in cui sarà utilizzato e venduto. Il termine deriva dall'inglese *locale* ed è spesso indicato come *L10n*, dove 10 sono le lettere sottintese per *Localization*.

È un'attività legata all'espansione dei mercati internazionali globali, in cui un identico servizio o prodotto deve essere reso fruibile da parte di un ampio parco di utenti. L'adattamento di un servizio o di un prodotto digitale (sito web, videogame, programma o anche un impaginato) avviene per un bacino di utenza geolinguistico (in inglese, *locale*). Nella localizzazione gli aspetti linguistici e tecnici sono interdipendenti. Gran parte del processo di localizzazione, difatti, riguarda la traduzione dell'interfaccia utente, i messaggi di sistema e la documentazione. L'adattamento traduttivo deve garantire, quindi, il perfetto funzionamento del prodotto. Per questo, già in fase di realizzazione del prodotto originale, viene prevista l'→internazionalizzazione, che è la premessa tecnica delle future localizzazioni di prodotto. Le scelte di localizzazione coinvolgono:

- lingua e scelta della sua varietà regionale
- sistema di scrittura e codifica dei caratteri
- convenzioni di scrittura (formati di data, ora, valuta, numeri, numeri di telefono, indirizzi, nomi, regole di punteggiatura)
- convenzioni culturali legate a colori, forme, immagini, suoni.

L'indicazione di *locale* è data da un identificatore di lingua (in minuscolo) e da un identificatore di regione (in maiuscolo) secondo gli standard ISO 639 (lingua) e ISO 3166 (Paese): per es., it/IT sta per italiano/Italia, mentre it/CH per italiano/Svizzera.

Logiterm

Applicazione per la traduzione assistita (CAT) di Terminotix, piccola società di Montreal, Canada. Poco conosciuta in Italia, l'applicazione riunisce una serie di funzioni linguistiche: ricerca full-text, creazione di file bilingui e memorie di traduzione, l'estrazione di terminologia, ecc. Funziona per sistemi Windows e lavora con formati standard per la traduzione, basati su XML (TMX, TBX, XLIFF). Apprezzabile la parte relativa alla terminologia: offre funzioni di estrazione automatica di sequenze di parole da un testo o un gruppo di testi e di ricerca su fonti esterne, quali Termium, GDT, TransSearch, OnTerm e Google direttamente dall'applicazione. La stessa società commercializza anche SynchroTerm, un programma di estrazione terminologica bilingue, con supporto per greco, nederlandese, ungherese, norvegese, polacco e turco.

Memoria di traduzione

Nelle applicazioni di →traduzione assistita (CAT) è il database dove sono archiviate le coppie di →segmenti originali e tradotti da un traduttore umano. Nel corso della traduzione interattiva di un nuovo testo, con un'applicazione CAT, 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. In questo formato il DGT, direttorato di traduzione, dell'Unione europea ha reso disponibile sul web l'Acquis communautaire, l'intero corpo delle leggi europee nelle 22 lingue ufficiali.

Multiterm

Prodotto di →SDL Trados per la gestione terminologica, ovvero per la creazione e gestione di →glossari. Gestisce singoli termini o singole espressioni, come avviene normalmente con i glossari, interfacciandosi con gli altri programmi SDL Trados per la →traduzione assistita (CAT). Crea database terminologici, detti termbase: per ogni termine è possibile specificare fino a 20 lingue di destinazione ed è possibile eseguire la ricerca in qualsiasi lingua di destinazione. L'interfaccia di Multiterm può essere considerata come una raccolta di schede, ciascuna delle quali contiene tutte le informazioni relative a un termine. Il programma si interfaccia in tempo reale con la suite SDL Trados per la traduzione assistita (CAT), permettendo l'inserimento controllato della corretta terminologia durante la traduzione interattiva.

OmegaT

Applicazione open-source per la traduzione assistita (CAT), multiplatforma. Attualmente è l'unica soluzione che può essere installata su sistemi Windows, Macintosh OS X e Linux. Sviluppato da un gruppo informale internazionale di volontari, è un marchio registrato in Germania.

Le sue caratteristiche includono segmentazione personalizzabili tramite espressioni regolari, memoria di traduzione con fuzzy match e propagazione dei match, controllo dei termini su glossario e dizionario, traduzioni automatiche da Apertium, Belazar e Google Translate. Supporta preferibilmente file in formati basati su XML con codifica UTF-8. Consigliato l'uso con OpenOffice, che viene utilizzato convertire i file MS Word versione 1997-2003. Crea memorie di traduzione in formato standard TMX e file bilingui in XLIFF.

OSCAR

Acronimo per Open Standards for Container/content Allowing Reuse. È l'organismo →LISA, per gli standard aperti per l'industria della traduzione e della localizzazione. Sviluppa e mantiene gli standard tecnici relativi alle esigenze linguistiche di aziende internazionali.

Gli standard OSCAR sono relativi a →memoria di traduzione, gestione della terminologia, l'elaborazione della traduzione, conteggio delle parole, analisi del volume di lavoro e gestione di contenuti multilingue. Al riguardo gli standard elaborati sono: →TMX, formato per memorie di traduzione; SRX, regole per la segmentazione del testo da tradurre, TBX, formato per lo scambio di dati terminologici; xml:tm, per l'inclusione di memorie in documenti XML; GMX-V, standard per il conteggio dei caratteri e delle parole.

Post - editing

Revisione, eseguita da un traduttore o da un correttore di bozze, del testo prodotto da un sistema di traduzione automatica.

Pre-editing

Attività di preparazione del testo da sottoporre a traduzione automatica al fine di ottenere traduzioni di qualità buona o quanto meno accettabile. Essa consiste essenzialmente nel semplificare il linguaggio e conformarlo allo standard gestibile del programma (attività, quindi, che si inserisce tra la redazione e la traduzione vera e propria) oppure a "predisporre" il testo seguendo delle linee guida per redattori tecnici (manualistica).

Pretraduzione

Confronto tra un intero testo di partenza e una memoria di traduzione al fine di inserire automaticamente la traduzione di tutti i match esatti trovati all'interno della memoria. Il risultato di tale processo è un testo ibrido che contiene sia segmenti pretradotti sia segmenti non tradotti.

Revisione

Verifica dell'accuratezza di una traduzione, con particolare riferimento agli errori di trasposizione, alle omissioni e alle aggiunte, alla resa di nomi e titoli, alla presentazione di dati, cifre e tabelle, ecc.

Riconoscimento di termini

Processo di individuazione automatica, all'interno di una base terminologica, di tutti i termini che ricorrono nella parte di testo su cui il traduttore sta lavorando.

N.B.: Riconoscimento ed estrazione di termini sono due concetti diversi.

Ripetizioni esterne

Segmenti che compaiono più di una volta all'interno di due o più testi di partenza ma che non sono stati ancora tradotti, non sono cioè ancora stati archiviati nella memoria di traduzione sotto forma di unità di traduzione. Vedere anche: ripetizioni interne, match esatto e analisi statistica del testo di partenza.

Ripetizioni interne

Segmenti che compaiono più di una volta nel testo di partenza ma che non sono stati ancora tradotti, non sono cioè ancora stati archiviati nella memoria di traduzione sotto forma di unità di traduzione. Vedere anche: ripetizioni esterne, match esatto e analisi statistica del testo di partenza.

Scheda terminologica (anche: voce terminologica)

Voce che contiene tutta l'informazione terminologica relativa a un singolo concetto. Ciò significa che tutte le informazioni relative allo stesso concetto (compresi, per esempio, i sinonimi) sono raccolte in un'unica entrata terminologica. Le informazioni relative a concetti diversi devono essere invece archiviate sotto voci diverse.

SDL TRADOS

Azienda leader nel settore delle applicazioni di → traduzione assistita. Nasce dalla fusione tra Trados GmbH e SDL, che in passato avevano prodotti concorrenti: la suite Trados e SDLX. La suite SDL Trados è composta da più applicazioni, con diverse funzioni pensate per supportare il traduttore.

In SDL Trados Studio 2009, suite offerta in più versioni standalone o server, sono comprese in un'unica interfaccia il sistema di gestione delle → memorie di traduzione, un editor testuale e il sistema di gestione del progetto. Il formato di file delle memorie è SDLTMM, e quello dei → file bilingui è SDLXLIFF, ambedue basati su → XML. Include anche un tool specifico, SDL Passolo, per la gestione di file di progetti di localizzazione (file risorse, eseguibili, html della guida in linea). La gestione terminologica si realizza con → Multiterm, che interagisce in tempo reale con gli altri strumenti di traduzione assistita. La precedente versione del software, SDL Trados 2007, è inclusa nella licenza del nuovo prodotto per permettere l'accesso e la conversione di precedenti memorie e file bilingui, mentre Winalign, strumento per l'→ allineamento di file, sopperisce alla mancanza di un analogo strumento nella nuova versione.

Segmento

Nei programmi di → traduzione assistita (CAT) è l'unità minima di traduzione, che nella maggior parte dei casi coincide con una frase. Il programma opera la segmentazione, dividendo il testo in base alla punteggiatura o ai segni di a capo, tabulazioni o anche → tag nei linguaggi di marcatura.

Il traduttore può controllare questo processo in base al tipo di testo, aspetto grafico o formato di file del documento da tradurre. Successivamente ad ogni segmento originale, in lingua di partenza (o source), viene abbinato un segmento tradotto, in lingua d'arrivo (target). La coppia di segmenti forma l'unità di traduzione (translation unit, TU), che verrà salvata nella memoria di traduzione. La possibilità di riciclare segmenti già tradotti nasce dalla creazione di regole di segmentazione ragionevolmente efficaci per individuare porzioni di testo sufficientemente piccole, ma sintatticamente sensate.

Tag

Sorta di etichetta associata a un elemento e contenente informazioni relative allo stesso (per esempio informazioni riguardo alla natura dell'elemento stesso o al modo in cui dovrebbe essere visualizzato). Un insieme di tag viene talvolta denominato marcatura; nei linguaggi di marcatura come l'HTML e l'XML i tag vengono racchiusi tra i segni di minore e maggiore (per esempio <CORSIVO>).

Termine

Parola, locuzione, espressione o simbolo usato in un settore ben definito per designare un determinato oggetto o concetto.

Term-mining

Insieme delle attività di individuazione ed estrazione automatica o semiautomatica di termini specialistici e dei loro contesti da testi o →corpora testuali. Più in generale il T.M. comprende anche tutte le attività e le tecniche di ricerca terminologica nel web attraverso l'uso di motori di ricerca generali o specializzati, la consultazione di banche dati terminologiche, testi specialistici, →corpora elettronici ecc.

L'estrazione di termini può essere sia monolingue sia multilingue. L'estrazione monolingue ha come obiettivo l'identificazione e l'estrazione di termini contenuti in singoli testi o all'interno di corpora testuali creati in vari formati (.txt, .rtf, .html o file esportati con strumenti di traduzione assistita) con l'ausilio di specifici strumenti di estrazione terminologica o di →concordancer.

L'estrazione multilingue analizza invece i testi delle diverse versioni linguistiche con lo scopo di identificare i potenziali termini e i loro equivalenti.

TMX (Translation Memory eXchange format)

Formato standard per memorie di traduzione sviluppato dalla LISA (Localization Industry Standards Association) al fine di facilitare lo scambio delle memorie tra i vari sistemi software e/o tra i fornitori di servizi di localizzazione. Per ulteriori informazioni consultare la homepage del TMX.

Traduzione assistita

È l'insieme degli strumenti e delle tecnologie software, che aiutano il traduttore nel processo di traduzione con l'obiettivo della memorizzazione e del riciclo della traduzione e del controllo della terminologia, producendo significativi guadagni in termini di tempo ed efficienza. Non va confusa con la traduzione automatica o *machine translation* (MT), dove la traduzione è affidata esclusivamente al programma. Con la traduzione assistita (CAT) il fattore umano resta comunque fondamentale.

Secondo alcuni, nella dizione CAT, andrebbero inclusi anche dizionari elettronici, spell checker, OCR, strumenti per la concordanza e per la ricerca nel testo. In senso stretto, seppur diversi tra loro e con funzionalità specifiche, i programmi CAT condividono elementi fondamentali quali:

- la suddivisione in →segmenti del testo sorgente;
- l'associazione delle coppie di segmenti sorgenti e tradotti in un →file bilingue
- -la memorizzazione e il riutilizzo del materiale tradotto;

Questi programmi vengono offerti come suite ovvero raccolta di applicazioni specializzate, ognuna dedicata a un aspetto differente del lavoro traduttivo, che interagiscono tra di loro in tempo reale.

Le applicazioni riguardano principalmente:

- → memoria di traduzione (*translation memory*) e recupero di testi paralleli con l' → allineamento di file
- gestione della terminologia (*terminology management*)
- gestione del progetto (*project management*)

Il settore della traduzione, che più si avvantaggia di questi strumenti è quello della traduzione professionale, studi e agenzie di traduzione, dove si lavora in team contemporaneamente allo stesso progetto per ridurre i tempi di consegna. In questo settore i testi da tradurre sono di carattere tecnico-scientifico o normativo (manuali, documentazione tecnica, schede di prodotto, ecc.). Sono pubblicazioni soggette a frequenti aggiornamenti, i cui testi sono basati su un linguaggio standard, controllato, necessariamente ripetitivo. L'uso degli strumenti CAT permette di assicurare il controllo della terminologia e la coerenza stilistica all'interno degli stessi testi tradotti in caso di aggiornamenti e, più in generale, di tutti i testi relativi a quel cliente, creando per essi una memoria globale.

Traduzione automatica

Processo in cui è il computer provvede alla traduzione di un testo. Un traduttore umano può intervenire nel processo svolgendo compiti di pre- o post-editing, ma è il computer, e non il traduttore, a svolgere una prima bozza di traduzione.

Trattamento automatico del linguaggio

Settore della → linguistica computazionale che si occupa dello sviluppo e dell'applicazione di risorse, sistemi, tecnologie, modelli e metodologie relativi all'elaborazione automatica del linguaggio naturale scritto e parlato. Si pone obiettivi come la comprensione del linguaggio umano per servizi quali il reperimento di informazioni, la → traduzione automatica e l'interazione vocale uomo-computer.

Le aree di ricerca del TAL comprendono, per esempio, la comprensione del significato di testi scritti e l'estrazione di informazioni rilevanti anche in lingue diverse, l'acquisizione dinamica di nuovi elementi di conoscenza su un certo dominio (→ text mining), la gestione e l'organizzazione dei documenti trovati per una condivisione ottimale delle conoscenze, il riconoscimento del parlato e del parlante, la conversione di testi scritti in messaggi vocali (→ sintesi vocale), la produzione, il riassunto e la → traduzione automatica di testi.

XML (eXtensible Markup Language)

Linguaggio di marcatura che serve a migliorare la funzionalità del web fornendo modalità di identificazione dell'informazione più flessibili e adattabili. Viene detto extensible ("estensibile") perché non è un formato fisso come l'HTML, che è invece un linguaggio di marcatura chiuso e non modificabile). Con l'XML, che in effetti è un "metalinguaggio", un linguaggio usato cioè per descrivere altri linguaggi, è possibile al contrario definire una marcatura personalizzata che si adatti a un numero infinito di documenti diversi.