



Dal pre-editing al post-editing: regole e impostazioni del sistema neurale alla base di una buona traduzione automatica

Speakers: Verena-Kristin Bickel, Davide Osta

because – il partner italiano di SYSTRAN



Redazione tecnica

L'outsourcing del ruolo del redattore tecnico offre un processo snello e un ciclo produttivo garantito. Vi offriamo la produzione della vostra documentazione tecnica o possiamo fornire supporto.



Intelligent Information

Vi aiutiamo a gestire la documentazione tecnica in maniera strutturata e modulare per creare, tradurre e pubblicare in ambito Industria 4.0. Offriamo dei tool all'avanguardia.



Training & Consulting

Vi offriamo consulenza e corsi di formazione nell'ambito della documentazione tecnica. Analizziamo insieme a voi le vostre esigenze per una efficace progettazione della consulenza.



Traduzioni tecniche

La documentazione tecnica (manuali, software, presentazioni aziendali, brochure, ecc.) necessita di traduzione per i produttori di macchine che esportano. La localizzazione fa sì che la traduzione rispecchi gli aspetti culturali e l'utilizzo della documentazione tecnica multilingue.



Terminologia

La terminologia fa parte della documentazione tecnica e del marketing. Una terminologia aziendale monolingue e multilingue dà più chiarezza nella documentazione aziendale. Vi offriamo delle soluzioni per implementare un sistema di terminologia nei vari reparti.



SYSTRAN

Soluzioni di traduzione automatica su misura

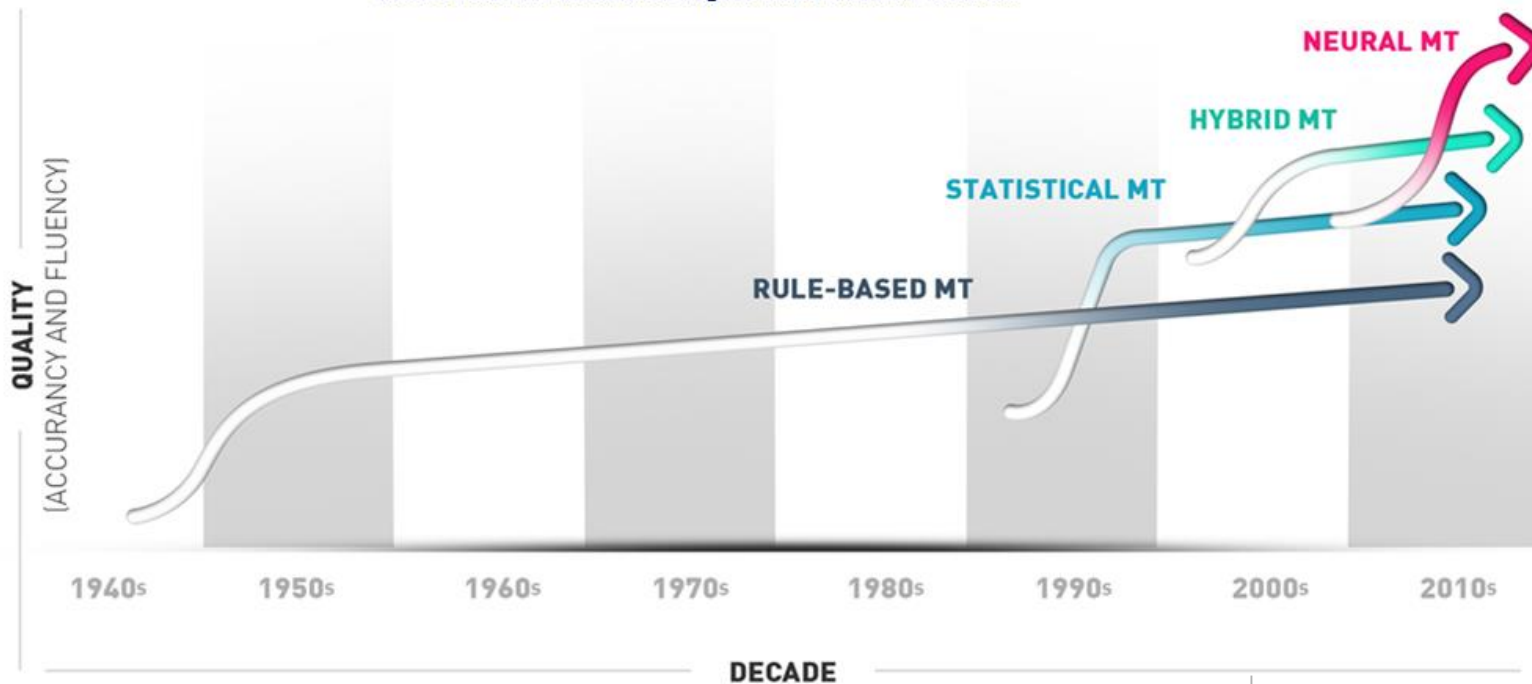


because Italy S.a.s.
Via Biliani 39, 15020 Mombello Monf.to (AL)
(+39) 0142 944555 - info@because-italy.com

NEURAL MACHINE TRANSLATION -VS- LEGACY TECHNOLOGIES

Sviluppo della tecnologia MT

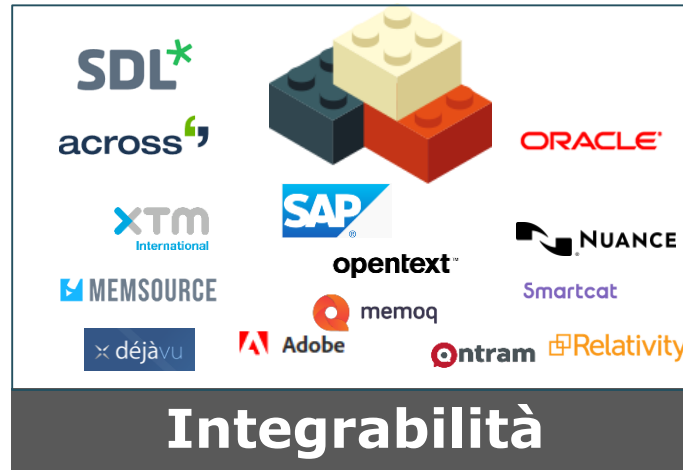
dalla soluzione basata su regole alla soluzione neurale



Valore aggiunto



- Sicurezza dei dati
- Server Linux in Europa

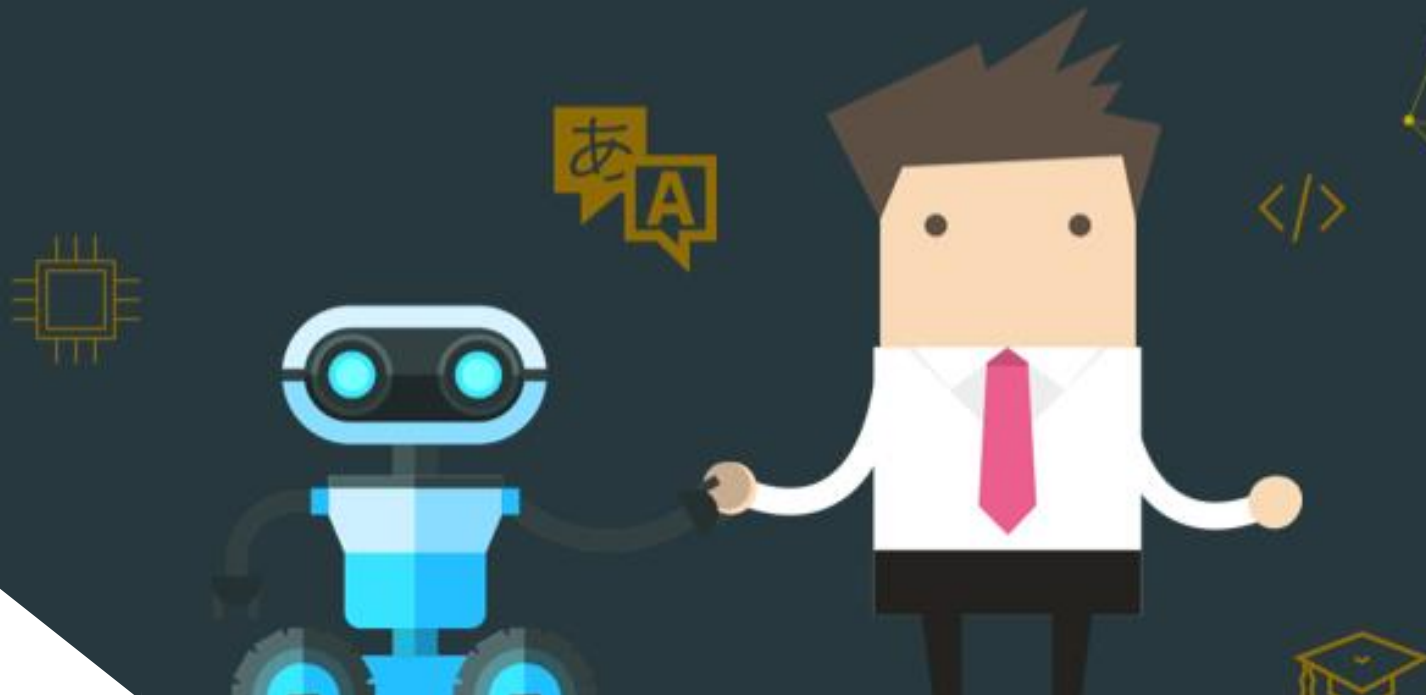


- Plug-in in CAT e CMS
- API per l'integrazione di Systran in Intranet o altri sistemi



- Traduzioni di alta qualità grazie a sistemi neurali specializzati
- Gestione terminologica (morfologia, semantica, sintassi)
- Miglioramento continuo della qualità della traduzione grazie a DeepLearning

Esempio pratico



Traduzione di documenti

Chi?



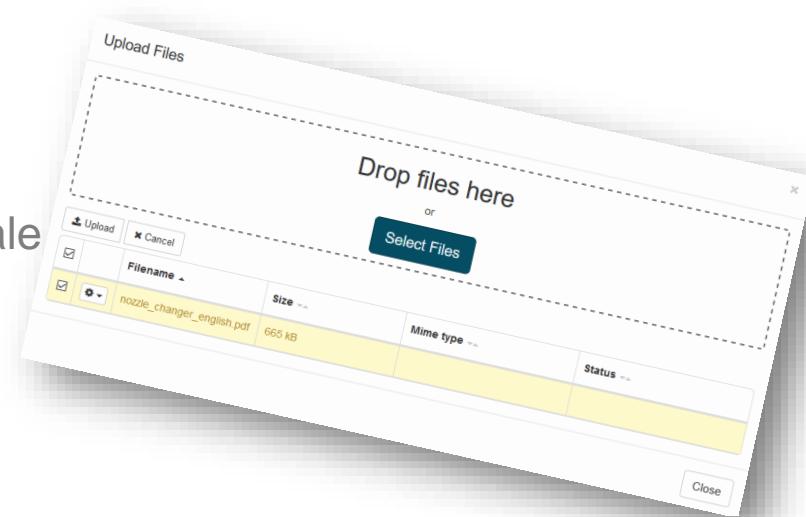
Redattrice
tecnica



Documentazione
tecnica

Scenario

1. La redattrice tecnica vuole capire un manuale francese.
2. Dovrebbe tradurre il manuale in tedesco. Ci vorrà del tempo.



Come potrebbe andare piu' veloce?

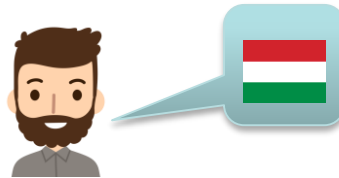
- Utilizza il portale SYSTRAN Pure Neural Server 9. Con il drag and drop carica il documento e con un click il manuale viene tradotto.

Integrazione API

Chi?



Servizio
clienti



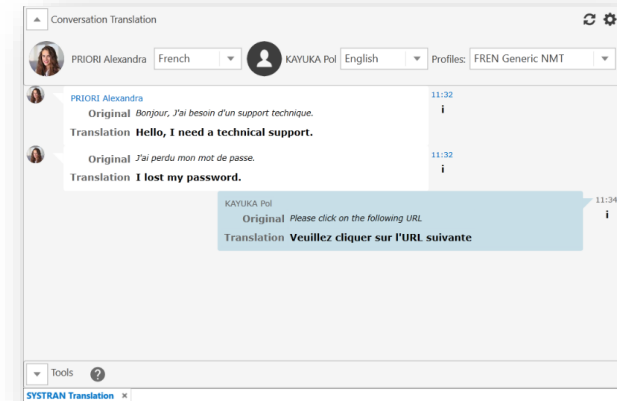
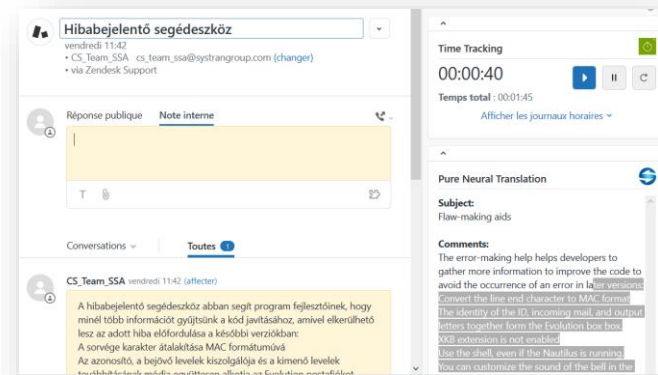
Tecnico

Scenario

1. Il servizio clienti vuole capire il fascicolo del tecnico ungherese.
2. Dovrebbe tradurre il fascicolo
3. E la comunicazione diretta su Skype o via e-mail sarà difficile.

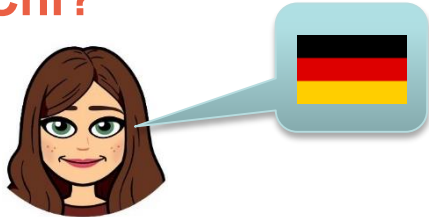
Soluzione?

- Utilizza il portale SYSTRAN Pure Neural Server 9.
- La comunicazione viene tradotta direttamente. Ognuno può scrivere nella propria lingua.



Internet Browser

Chi?



Marketing



Concorrente coreano

Scenario

1. La responsabile di marketing vuole controllare il sito web di un concorrente coreano.
2. Deve tradurre il sito web in tedesco.

Come potrebbe farlo?

- SYSTRAN traduce il sito web in tempo reale.
- Basta inserire il link del sito web nell'interfaccia SYSTRAN.



Audio

[Text Translation](#) [File Translation](#) [Speech Translation](#) [Web Translation](#)

Choose translation setting and click on "Upload" to select the files to translate. Once translation finished, click on the download button to retrieve your translated file.

Auto-Detect Target

Upload

Search

<input type="checkbox"/>	Filename	Status	Upload date	Languages	Profile selection	Size	TI
--------------------------	----------	--------	-------------	-----------	-------------------	------	----

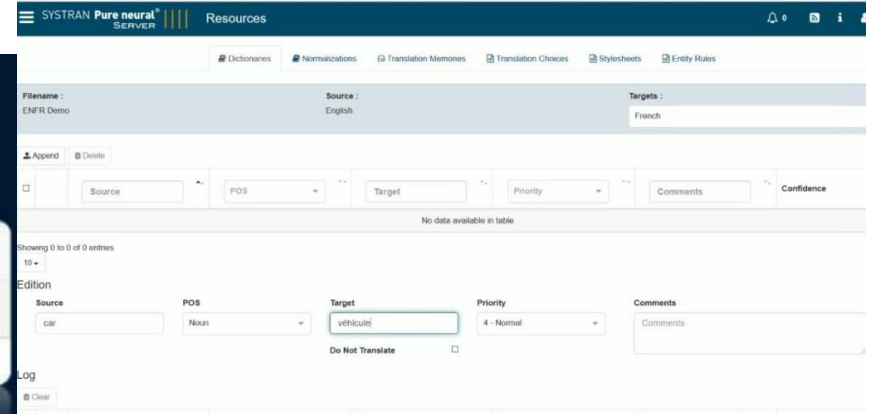
No data available in table

Terminologia

SPN9 – NMT with User Dictionary When Coding Engine Met Neural Networks

<input type="checkbox"/>		sentence	Noun	verdict	Inflections
<input type="checkbox"/>		to sentence	Verb	condamner	condamner il condamne condamné

Showing 1 to 2 of 2 entries

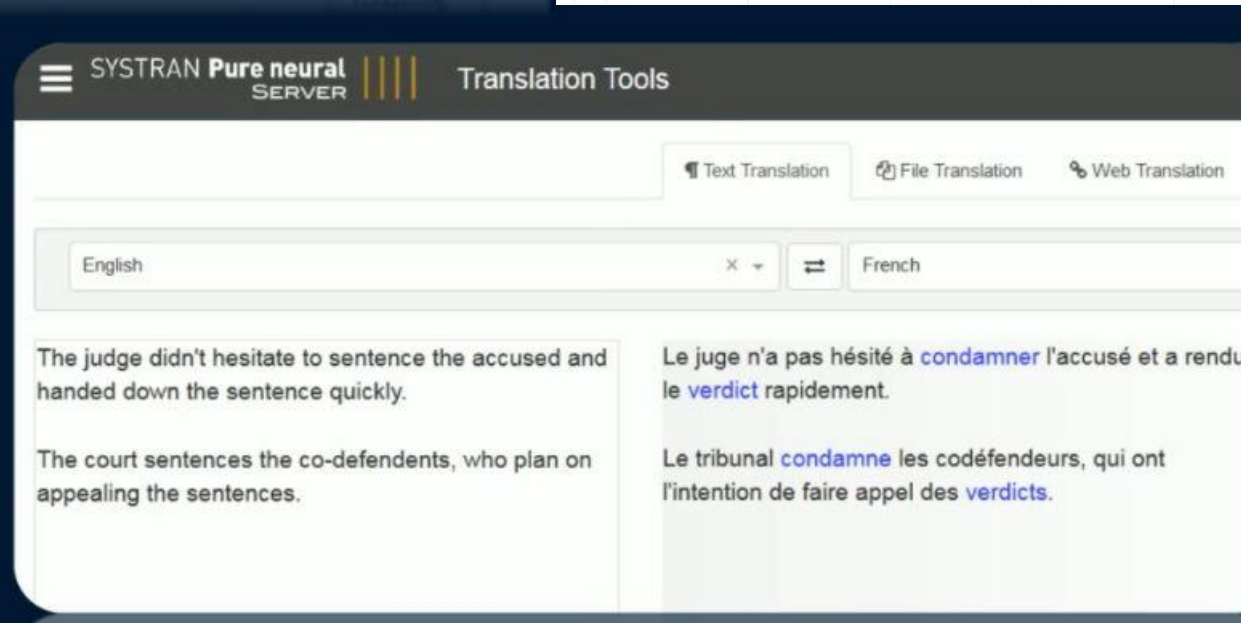


The screenshot shows the SYSTRAN Pure neural SERVER interface. At the top, there are tabs for Dictionaries, Normalizations, Translation Memories, Translation Choices, Stylesheets, and Entry Rules. Below this, the 'Resources' section is active, showing a table with columns for Filename, Source, and Targets. The current entry is 'ENFR Demo' with Source 'English' and Targets 'French'. Below the table, there are fields for 'Append', 'Delete', 'Source', 'POS', 'Target', 'Priority', 'Comments', and 'Confidence'. The 'Target' field is set to 'vehiculel'. Below this, there is a section for 'Edition' with a table for 'Source', 'POS', 'Target', 'Priority', and 'Comments'. The 'Source' is 'car', 'POS' is 'Noun', and 'Target' is 'vehiculel'. There is also a 'Log' section with a 'Clear' button.

The Translation Engine:

Determines the correct part of speech (with homograph support)

Inflects words with morphology support



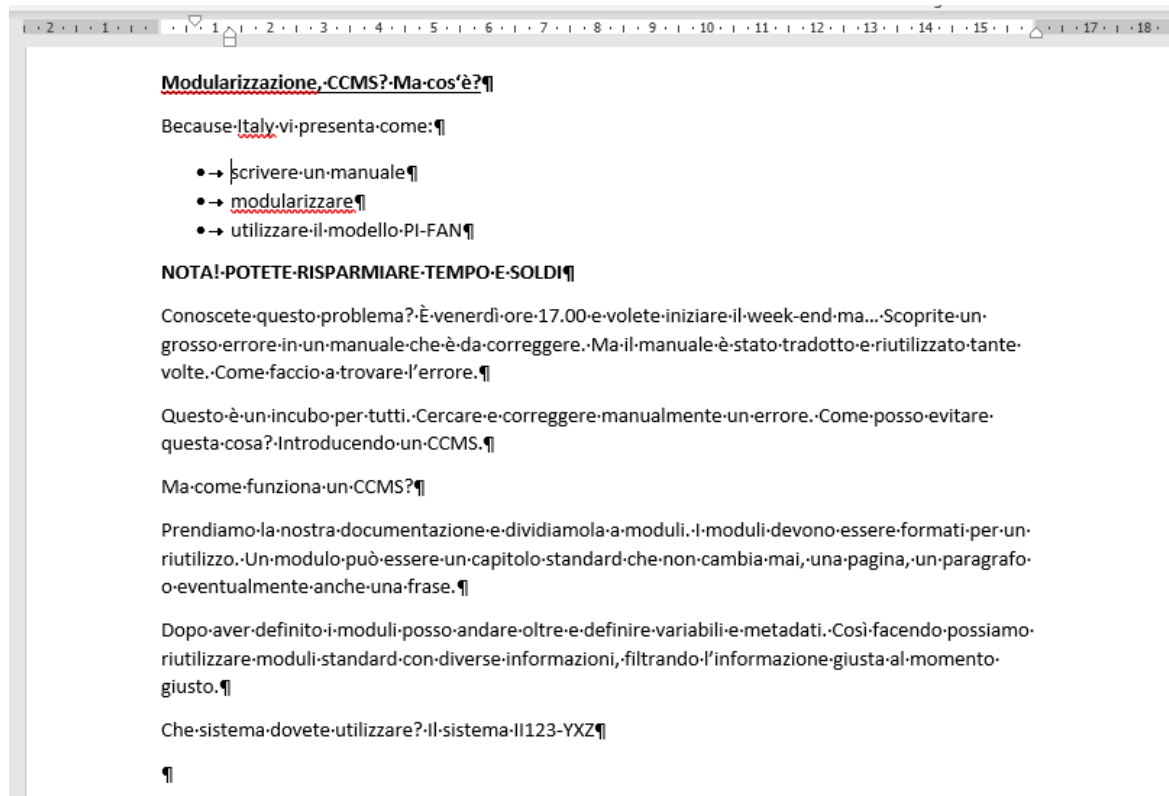
The screenshot shows the SYSTRAN Pure neural SERVER Translation Tools interface. At the top, there are tabs for Text Translation, File Translation, and Web Translation. Below this, there are fields for 'English' and 'French' with a bidirectional arrow between them. The main area shows two text boxes. The left box contains the English text: 'The judge didn't hesitate to sentence the accused and handed down the sentence quickly.' and 'The court sentences the co-defendants, who plan on appealing the sentences.' The right box contains the French text: 'Le juge n'a pas hésité à condamner l'accusé et a rendu le verdict rapidement.' and 'Le tribunal condamne les codéfendeurs, qui ont l'intention de faire appel des verdicts.' The word 'condamner' is highlighted in blue in the French text.

E ora vediamo il sistema!



Facciamo dei test!

Vediamo come si comporta il sistema con diversi testi.



Falsa segmentazione



Ieri pomeriggio mi sono
comprato un libro

Ieri pomeriggio mi sono comprato un libro.



Yesterday afternoon I was
bought a book

Yesterday afternoon I bought myself a book.

... dobbiamo fare un pre-editing?





Norma ISO 18587:2017



Processi di preproduzione

Il fornitore di servizi linguistici dovrà avere un **processo per determinare**, se necessario in collaborazione con il cliente, **se il testo sorgente è adatto alla traduzione automatica e al conseguente post-editing**, dal momento che l'efficienza della lavorazione machine translation / post-editing dipende dal sistema utilizzato, dalla combinazione linguistica, dal settore e **dallo stile del linguaggio utilizzato** nel testo sorgente.

È possibile eseguire un pre-editing del testo prima di mandarlo in traduzione automatica in modo da facilitare il processamento, migliorare l'output del sistema e diminuire di conseguenza il tempo necessario per il post-editing.

Modi di dire



Hai capito Roma per toma; parlare con te è come lavare la testa all'asino.

È inutile parlare con te, non comprendi quello che ti dico!



You understand Rome for toma. talking to you is like washing your head off the donkey.

It's no use talking to you, you don't understand what I'm saying!

Frasi spezzate in elenchi puntati

Because Italy vi spiega come:

- scrivere un manuale
- modularizzare i contenuti
- utilizzare il modello PI-FAN



Because Italy vi spiega:

- **come** scrivere un manuale
- **come** modularizzare i contenuti
- **come** utilizzare il modello PI-FAN

Because Italy explains how:

- to write a manual
- modularisation of content
- use IP-FAN



Because Italy explains:

- **how** to write a manual
- **how** to modulate content
- **how** to use the PI-FAN model

Acronimi e abbreviazioni



Fate il TQA evt. direttamente nel CAT.

Do the TQA Ev. directly to the CAT.



Eventualmente, eseguite il **controllo di qualità** direttamente tramite **SDL Trados Studio**.

Where appropriate, perform **quality control** directly through **SDL Trados Studio**.

Omissioni di articoli e preposizioni (HMI)

movimentazione leva sinistra cambio posizione riposo -> lavoro



Litigation move move rest position change -> work

Movimentazione **della** leva sinistra **del** cambio **dalla** posizione **di** riposo **alla** posizione di lavoro

Movement of the left lever of the change from the rest position to the working position

Sostantivizzazione di verbi (linguaggio legale)



La sostantivizzazione delle espressioni verbali complica la traducibilità del testo

Per facilitare la traduzione, usa verbi semplici al posto di sostantivi astratti che derivano da verbi



Subversive verbal expressions prevent fluid translation of text

To facilitate translation, use simple verbs instead of abstract names that come from verbs



La batteria dell'automobile si è fusa
e ora ci tocca suonare la batteria in
automobile.



The car battery melted and now we have
to play the battery in the car.

Utilizzo incoerente di termini

Utilizzare verbi precisi al posto di generici

Agire sulla vite di un quarto di giro.

Act on the lives of a quarter of a turn.



Ruotare la vite di un quarto di giro.

Turn the screw by a quarter of a turn.

Verbi impersonali - pronomi

Si raccomanda di recarsi puntualmente sul lavoro, anche se si fa fatica a rispettare ciò.

La direzione richiede ai dipendenti di arrivare puntualmente al lavoro, anche se questi fanno molta fatica a rispettare questa regola.



It is recommended that people go to work on time, even if it is difficult to respect this.

Management requires employees to get to work on time, even if they struggle to follow this rule.



Sublime è il senso di sgomento che l'uomo prova di fronte alla grandezza della natura sia nell'aspetto pacifico, sia ancor più, nel momento della sua terribile rappresentazione, quando ognuno di noi sente la sua piccolezza, la sua estrema fragilità, la sua finitezza, ma, al tempo stesso, proprio perché cosciente di questo, intuisce l'infinito e si rende conto che l'anima possiede una facoltà superiore alla misura dei sensi.

Sublime is the sense of dismay that man feels in the face of the magnitude of nature both in the peaceful way and, even more so, at the moment of his terrible representation, when each of us feels his small, extreme fragility, his finiteness, but, at the same time, because he is aware of this, he sees infinity and realises that the soul has a faculty that exceeds the measure of the senses.

**Frase molto lunga anche con subordinate, con
soggetti variabili**

Riassumendo: problemi più frequenti?

- Segmentazione
- Elenco puntato con frasi spezzate
- Tutto maiuscolo
- Nomi propri
- Acronimi
- Modi di dire
- Omissioni di articolo e preposizioni
- Slang
- Terminologia non coerente ...



Allegato C - Pre-editing

[...] Il pre-editing può coprire una gamma ampia o piuttosto ristretta di attività preparatorie. Nella sua accezione più ampia, può consistere nella **trasformazione del testo in un linguaggio controllato** (CNL), in cui il contenuto è scritto secondo regole fisse per rendere più facile la traduzione automatica. Questo è chiamato pre-editing linguistico. Ad esempio, utilizzando frasi relativamente brevi, evitando certe strutture sintattiche complesse o ambigue o assicurando che lo stesso termine sia usato in modo coerente, senza sinonimi.

Nella sua accezione più ristretta, il pre-editing si concretizza nel **ripulire il contenuto della lingua di partenza** attraverso il controllo ortografico, il controllo del formato e l'etichettatura di testi come indirizzi e nomi propri per impedire la traduzione. È possibile utilizzare **strumenti per facilitare il pre-editing**, ad esempio, suddividendo automaticamente ogni frase in brevi frasi naturali, che possono poi essere tradotte manualmente se necessario. Questo contributo umano può migliorare l'accuratezza della traduzione automatica e quindi ridurre il carico di lavoro successivo di post-editing.





Il pre-editing è altamente raccomandato se un documento deve essere tradotto in più di tre lingue, in quanto è più economico investire tempo in un ampio pre-editing linguistico per garantire la migliore resa possibile della traduzione automatica, e quindi minori spese di post-editing in ogni lingua di destinazione.

Il pre-editing è fondamentale anche se il documento deve essere tradotto in una lingua che funge da lingua di partenza per una seconda traduzione in un'altra combinazione linguistica.

Dopo aver ottimizzato il testo...

... creiamo un glossario.

10 ▾

Edition

Source: modulo

POS: Noun

Target: topic

Priority: 1 - High Over Rules & Ex... ▾

Comments: Comments

Do Not Translate:

Log

Clear

Status ▾	POS ▾	Target ▾	Priority ▾	Confidence ▾
No data available in table				

+

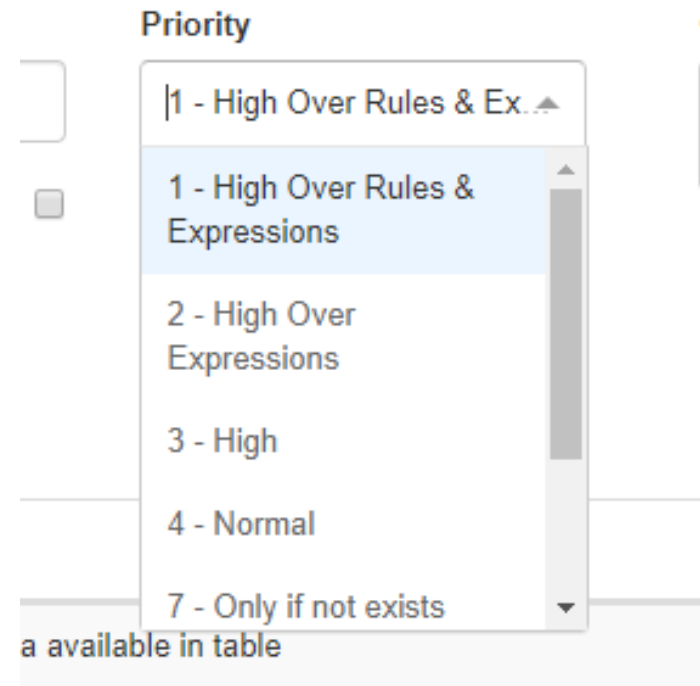
✎

i

Sistema neurale + glossario rule-based

Definendo l'attributo grammaticale il sistema è in grado di gestire la parola inserita.

Definendo la priorità del termine il sistema sa come gestirlo all'interno della traduzione.



Creiamo il nostro profilo di traduzione

... aggiungendo il nostro glossario e le nostre normalizzazioni al profilo

- Profile Selection
- Resources
 - Dictionaries
 - CCMS IT-EN
 - SYSTRAN Main Dictionary
 - SYSTRAN Alternative Dictionary
 - SYSTRAN Industries Dictionary
 - SYSTRAN Business Dictionary
 - SYSTRAN Life Sciences Dictionary
 - SYSTRAN Colloquial Dictionary
 - SYSTRAN Science Dictionary
 - Modularizzazione
 - Normalizations

Ritraduciamo il testo sul sito

Choose translation setting and click on "Upload" to select the files to translate. Once translation finished, click on the download button to retrieve your translated file.

Italian x ≡ English x + Upload

Refresh Download Edit Cancel Delete
Search

<input type="checkbox"/>	Filename	Status	Upload date	Languages	Profile selection	Size	TI
<input type="checkbox"/>	<input type="checkbox"/> Modularizzazione.docx	translated	39 minutes ago	IT → EN	domain: Generic owner: Self size: M	14.5 kB	
<input type="checkbox"/>	<input type="checkbox"/> Informazione 4.0 – Open Day.docx	translated	an hour ago	IT → EN	domain: Generic owner: Self size: M	15.7 kB	
<input type="checkbox"/>	<input type="checkbox"/> Documentazione tecnica per Industria 4....	translated	an hour ago	IT → EN	domain: Generic owner: Self size: M	19 kB	

... o tramite CAT-Tool

Traduzione automatica

Memories and Server for machine translation for the language pair in this project.

Translation Memories and machine translation

Legen Sie ein neues Translation Memory an oder verwenden Sie ein bestehendes und fügen Sie nach Bedarf MU oder andere Ressourcen hinzu.

+ Anlegen... Verwenden... Einstellungen... ↓ ↑ ✕ Entfernen

Name

Dateibasiertes Translation Memory...
 Serverbasiertes Translation Memory...
 SDL BeGlobal Enterprise...
 Maschinelle Übersetzung mit SDL Language Cloud ...
 Google Cloud Translation API...
 AnyTM: beliebiges dateibasiertes TM...
 AnyTM: beliebiges serverbasiertes TM...
 AnyTM: beliebige andere Übersetzungsquelle...
 SDL WorldServer-Translation Memory...
 SYSTRAN...

Sprachpaare mit anderen...
 Für die folgenden Sprachen...
 Name

Alle Sprachpaare

Ankordanz Aktualisieren



SYSTRAN...

Connection Type
 Select the SYSTRAN Translation Provider connection type.

SYSTRAN Pure Neural Server/SYSTRAN Enterprise Server
 SYSTRAN Online Services (SYSTRAN.io)
 Use a saved connection

translationpartners-spn9.mysystran.com

< Back Next > Finish Cancel

SYSTRAN...

Available Language Directions
 The list of available language directions.

Server Address:

Server Type: V7 V9/V8

Username:

Password:

API Key:

< Back Next > Finish Cancel



SYSTRAN...

Profile
 Select multiple translation profiles. Please make sure that you select only one profile from a group.

Profile: Italian-English

ITEN
 ITEN
 CCMS_IT-EN
 CCMS_IT-EN
 Modularizzazione_Test

< Back Next > Finish Cancel

E controlliamo il risultato



Conclusione

- Dobbiamo scegliere il sistema AT secondo le nostre esigenze
- La norma ISO 18587:2017 definisce come trattare il processo
- Con un buon pre-editing e specializzando il sistema possiamo ottenere ottimi risultati
- Il post-editing è ridotto al minimo e ottimizzato
- La maggior parte del processo di traduzione è completamente automatizzato
- Il testo si legge bene sia nella lingua di partenza sia in quella di arrivo



European
Commission

