



Uso dei corpora linguistici per l'interpretazione: generazione di un glossario bilingue

MÍRIAM PÉREZ CARRASCO

Universidad de Málaga

miriamperez@uma.es

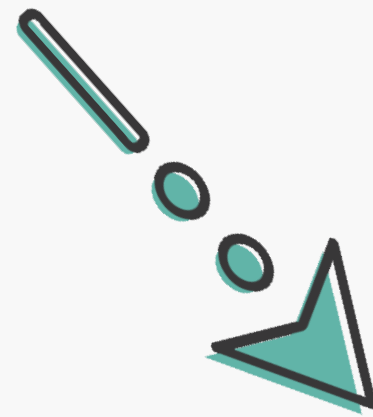
Indice

- I Introduzione
- II I corpora linguistici
- III Disegno e compilazione del nostro corpus
- IV Estrazione terminologica con AntConc
- V Generazione del glossario bilingue EN-ES per interpreti
- VI Importazione del glossario in un gestore terminologico per interpreti

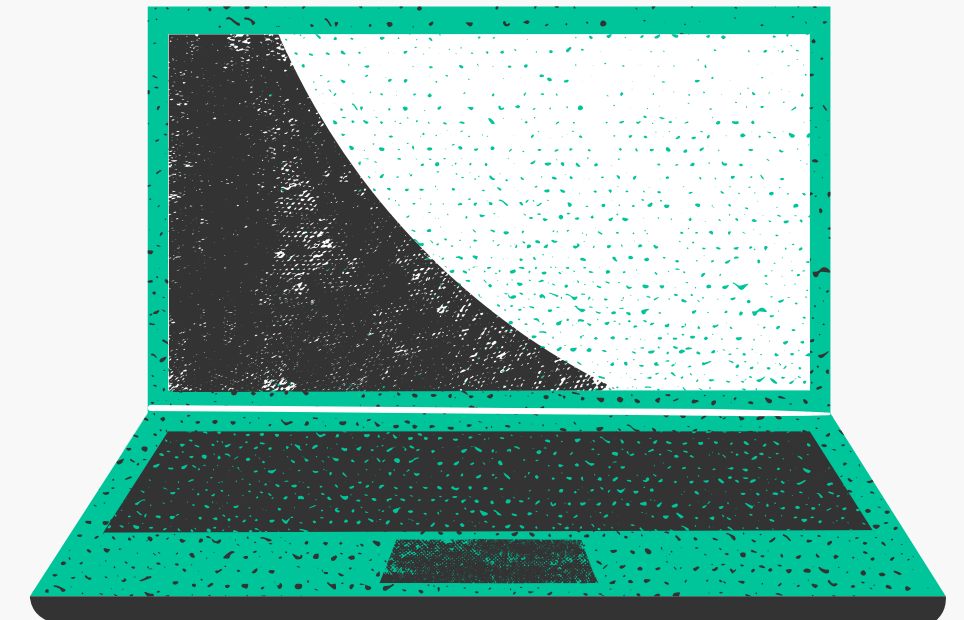
1 INTRODUZIONE



Competenze
linguistiche e
comunicative



INTERPRETE



Fase di preparazione,
competenza
documentale

I INTRODUZIONE

Lavoro contro il tempo

Temi altamente specializzati

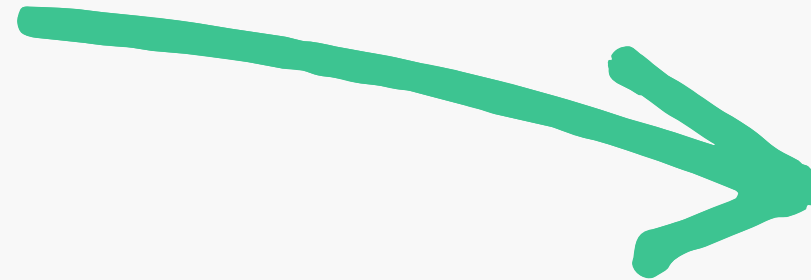
Convegni o incontri tra esperti

Temi sempre più vari e costantemente aggiornati



I INTRODUZIONE

Possiamo incorporare all'interno del flusso di lavoro dell'interprete **diversi strumenti tecnologici** che **semi-automatizzano** alcune delle fasi proposte da Moser-Mercer (1992: 508-509).



I corpora linguistici

II I CORPORA LINGUISTICI

EAGLES (Expert Advisory Group on Language Engineering Standards, 1996: 4)

"a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of language"

- Consentono di **estrarre automaticamente i termini.**
- Permettono di accedere a un **numero infinito di testi specializzati**, sia manualmente a partire dai testi forniti dal cliente stesso, sia da testi scaricati da Internet o tramite ricerche automatiche sul web.
- Permettono di verificare dove e come (**contesto**) e quante volte (**frequenza**) i termini sono stati utilizzati nei testi compilati.
- Possiamo fare uso di database o gestori terminologici in cui i glossari possono essere **archiviati, localizzati rapidamente e facilmente aggiornati o modificati.**

II I CORPORA LINGUISTICI



- Corpus virtuale e bilingue (ES-EN).
- Corpus parallelo (composto da testi originali e le loro traduzioni).

III DISEGNO E COMPILAZIONE DEL NOSTRO CORPUS

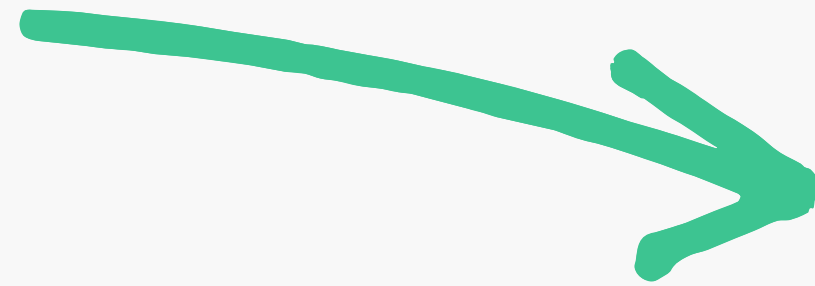
The logo for MOTOCOR, consisting of the word "MOTOCOR" in a bold, black, sans-serif font, enclosed within a thick, hand-drawn green oval border.

Per la compilazione del corpus abbiamo seguito, innanzitutto, i criteri stabiliti da **Bowker e Pearson (2002)**.

- **SCOPO:** l'estrazione della terminologia per la generazione di un glossario specializzato bilingue (EN/ES).
- **DIMENSIONE:** dieci testi per ogni lingua (82520 tokens EN; 139567 tokens ES).
- **MEZZO:** testi scritti, scaricabili facilmente e gratuitamente dal sito web della casa motociclistica.
- **TEMATICA:** motociclismo (moto leggera, scooter).
- **GENERE TESTUALE:** manuale del proprietario.
- **ATTRIBUZIONE:** la casa motociclistica.
- **DATA DI PUBBLICAZIONE:** manuali prodotti tra il 2014 e il 2021.
- **LINGUE:** inglese e spagnolo.

IV ESTRAZIONE TERMINOLOGICA CON ANTCONC

Identifichiamo (semi)automaticamente
i termini più rilevanti in entrambi i
corpus per identificare potenziali
termini e i loro traducenti



Creazione di un
glossario
terminologico
bilingue



AntConc
(Laurence Anthony)

IV ESTRAZIONE TERMINOLOGICA CON ANTCONC

Due possibilità per
estrarre l'elenco dei
termini

KEYWORDS

Il programma confronta il corpus caricato con un corpus di riferimento ed estrae i termini che sono relativamente frequenti nel corpus specializzato e poco frequenti nel corpus di riferimento.

WORD LIST

Calcola la frequenza con cui ogni termine appare nel nostro corpus e li ordina in base a questi risultati.

Occorre caricare una **stoplist** per evitare i termini che hanno soltanto un significato grammaticale.



IV ESTRAZIONE TERMINOLOGICA CON ANTCONC

Rank	Freq	Lemma	Lemma Word Form(s)
1	1912	check	check 1215 checked 217 checking 393
2	1513	brake	brake 1297 brakes 73 braking 143
3	1389	engine	engine 1387 engines 2
4	1278	oil	oil 1256 oils 22
5	1256	vehicle	vehicle 1213 vehicles 43
6	985	light	light 760 lighting 26 lights 199
7	933	fuel	fuel 933
8	761	maintenance	maintenance 761
9	728	switch	switch 612 switched 25 switches 89 swi
10	695	warning	warning 684 warnings 11
11	678	scooter	scooter 646 scooters 32
12	662	remove	remove 482 removed 72 removing 108
13	626	key	key 626



Termini più frequenti in inglese

IV ESTRAZIONE TERMINOLOGICA CON ANTCONC

Rank	Freq	Lemma	Lemma Word Form(s)
1	1348	motor	motor 1341 motores 7
2	1329	freno	freno 1039 frenos 290
3	1192	vehículo	vehículo 1147 vehículos 45
4	1154	aceite	aceite 1132 aceites 22
5	1023	comprobar	comprobadas 2 comprobado 1 compro
6	981	sistema	sistema 913 sistemas 68
7	965	gasolina	gasolina 965
8	960	luz	luces 167 luz 793
9	845	mantenimiento	mantenimiento 845
10	700	líquido	líquido 695 líquidos 5
11	681	utilizar	utilice 323 utiliza 72 utilizado 24 utilizar
12	649	indicador	indicador 357 indicadora 234 indicador
13	646	delantero	delantera 192 delanteras 3 delantero 4
	



Termini più frequenti in spagnolo

V GENERAZIONE DEL GLOSSARIO BILINGUE EN-ES PER INTERPRETI

Una volta estratti i termini più frequenti sia in inglese che in spagnolo, **confrontiamo i risultati in ciascuna lingua per abbinare ogni termine al suo equivalente.**



Microsoft Excel

L'interprete ha la possibilità di includere tutti i termini estratti da AntConc nel glossario o di scartare quelli che non sono di suo interesse

	A	B
1	ESPAÑOL	ENGLISH
2	abrazadera	clamp
3	ABS	ABS
4	aceite	oil
5	aceleración	acceleration
6	acelerador	throttle
7	acoplador	coupler
8	advertencia	warning
9	ajustes	adjustment
10	amortiguador	shock absorber
11	amperaje	amperage
12	anticongelante	antifreeze
13	apretar	tighten
14	arrancar	start
15	asa de agarre	grab bar
16	asiento	seat
17	atención	notice
18	avería	damage, failure, fault, malfunction
19	bastidor	frame
20	batería	battery
21	bocallave	keyhole
22	bocina	horn
23	bomba	pump
24	bombilla	bulb
25	bujía	spark plug
26	cable	cable
27	caja	box
28	calentar	warm up
29	calzada	road
30	cambiar	replace



Microsoft Excel

VI IMPORTAZIONE DEL GLOSSARIO IN UN GESTORE TERMINOLOGICO PER INTERPRETI

Esistono sistemi di gestione terminologica appositamente disegnati per soddisfare le esigenze dell'interprete



InterpretBank



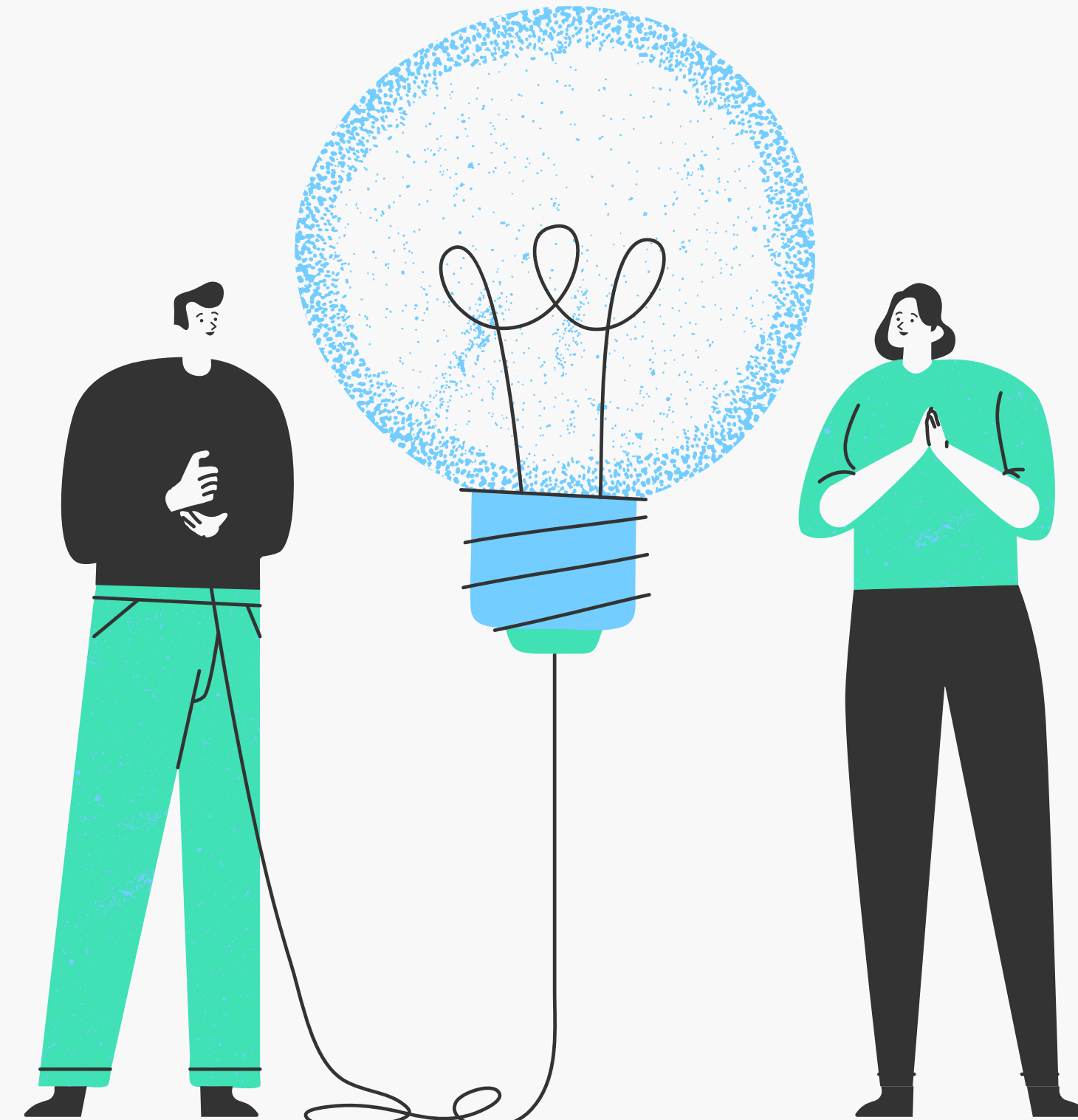
InterpretBank permette di importare il glossario creato con Excel

VI IMPORTAZIONE DEL GLOSSARIO IN UN GESTORE TERMINOLOGICO PER INTERPRETI

Spanish		English	
Term	advertencia	Term	warning
NotePad	Warning - advertencia (if it is not heeded, it can cause death or serious injury) vs. Caution - precaución (if it is not taken, it may cause minor or moderate injury)		Term ID: 210 MOTOCOR_ Last: 2021-06-02 Creator: miria
Spanish ▲▼	English ▲▼	BoothInfo ▲▼	
abrazadera	clamp		
ABS	ABS	Siglas de «Anti-lock Braking System», en español «Sistema Antibloqueo de Frenos»	
aceite	oil		
aceleración	acceleration		
acelerador	throttle	/'θrɒtəl/	
acoplador	coupler	/'kʌplə/	
advertencia	warning	Warning - advertencia (if it is not heeded, it can cause death or serious injury) vs. Caution - precaución (if it is not taken, it may cause minor or moderate injury)	
ajustes	adjustment		
amortiguador	shock absorber		
amperaje	amperage	/'æmpərɪdʒ/	

CONCLUSIONI

- Anche gli interpreti possono beneficiare dei numerosi vantaggi che offrono gli strumenti informatici, in questo caso durante la fase di preparazione.
- Il processo descritto soddisfa le esigenze specifiche degli interpreti.
- La metodologia descritta, sia per la compilazione del corpus che per l'estrazione della terminologia e la creazione di un glossario, può essere applicata a qualsiasi argomento, lingua o modalità di interpretazione (o traduzione).



REFERENZE BIBLIOGRAFICHE

Bowker, L., J. Pearson (2002). *Working with Specialized Language: A practical guide to using corpora*. Londres: Routledge.

EAGLES (Expert Advisory Group on Language Engineering Standards) (1996). *Text corpora Working Group reading Guide*. EAGLES Document EAG-TCWG-FR-2. Versión de mayo de 1996. Recuperado de <http://www.ilc.cnr.it/EAGLES/corpintr/corpintr.html>.

Moser-Mercer, B. (1992). Banking on Terminology Conference Interpreters in the Electronic Age. *Meta*, 37(4), 507-522.

Seghiri, M. (2006). *Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad*. Málaga, Universidad de Málaga. Tesis doctoral.

Seghiri, M. (2011). Metodología protocolizada de compilación de un corpus de seguros de viaje: aspecto de diseño y representatividad. *Revista de lingüística teórica y aplicada*, 49(2), 13-30.

RINGRAZIAMENTI

Questa ricerca è stata condotta nell'ambito delle reti tematiche T3 e della rete didattica di eccellenza TACTRAD (Ref. 719/2018), tutte presso l'Università di Malaga, nonché nell'ambito dei progetti VIP II (Ref. PID2020-112818GB-I00), TRIAJE (Ref. UMA18-FEDERJA-067), PROFETA (Ref. PIE19-033, UMA) e INMOCOR (Ref. P20-00109, PAIDI-J.A.).

Questo studio è stato realizzato anche grazie a un contratto FPU (Formazione Universitaria degli Insegnanti), codice FPU18/00206, concesso dal Ministero della Scienza, dell'Innovazione e dell'Università della Spagna a Míriam Pérez Carrasco, che ha anche ricevuto una sovvenzione dal 1° Piano di Ricerca, Trasferimento e Divulgazione Scientifica dell'Università di Malaga per la partecipazione a conferenze scientifiche.



UNIVERSIDAD
DE MÁLAGA

Convegno dottorale 2022

Università di Verona



UNIVERSIDAD
DE MÁLAGA

**¡Muchas gracias
por su atención!**

MÍRIAM PÉREZ CARRASCO

Universidad de Málaga

miriamperez@uma.es