

Alma Mater Studiorum Università di Bologna

SCUOLA DI LINGUE E LETTERATURE, TRADUZIONE E
INTERPRETAZIONE
Sede di Forlì

**Corso di Laurea magistrale in
Traduzione specializzata (classe LM - 94)**

TESI DI LAUREA

in Metodi e Tecnologie per la Traduzione

*Le collocazioni in traduzione e interpretazione
tra italiano e inglese: uno studio su EPTIC_01_2011*

CANDIDATO:

Manuela Santandrea

RELATORE:

Adriano Ferraresi

CORRELATORE

Silvia Bernardini

Anno Accademico 2013/2014

Sessione II

Indice

1	Introduzione.....	1
2	Background teorico.....	3
2.1	Introduzione al capitolo.....	3
2.2	Corpora.....	3
2.2.1	Corpus-based Translation Studies.....	3
2.2.2	Corpus-based Interpreting Studies.....	5
2.2.3	Tipologie di corpus.....	7
2.3	Collocazioni.....	10
2.3.1	Definizioni.....	10
2.3.2	Collocazioni e corpora.....	15
2.3.2.1	Le collocazioni dei parlanti non nativi.....	17
2.3.2.2	Le collocazioni in traduzione e interpretazione.....	19
3	EPTIC_01_2011.....	21
3.1	Introduzione al capitolo.....	21
3.2	Il corpus.....	21
3.3	Struttura del corpus.....	23
3.4	Preparazione del corpus.....	26
3.4.1	Sottocorpora di interpretazione.....	27
3.4.2	Sottocorpora di traduzione.....	29
3.4.3	Convenzioni di editing.....	29
3.4.3.1	Parole troncate.....	30
3.4.3.2	Disfluenze.....	30
3.4.3.3	Pause.....	32
3.4.3.4	Numeri.....	33
3.4.3.5	Problemi di comprensione.....	34
3.4.3.6	Punteggiatura.....	34
3.4.3.7	Applausi e parole inventate.....	35
3.4.3.8	Lettere maiuscole e appellativi.....	36
3.4.4	Metadati.....	39
3.4.4.1	Metadati relativi al testo.....	40

3.4.4.2 Metadati relativi all'oratore.....	43
3.4.4.3 Metadati relativi al testo source	44
3.4.4.4 Metadati relativi all'interprete.....	45
4 Metodo.....	49
4.1 Introduzione al capitolo	49
4.2 Estrazione e classificazione dei bigrammi.....	49
4.3 Analisi quantitativa	52
4.4 Analisi qualitativa.....	53
5 Risultati.....	57
5.1 Introduzione al capitolo	57
5.2 Analisi quantitativa	57
5.2.1 Bigrammi non frequenti/non attestati.....	57
5.2.1.1 Inglese	57
5.2.1.2 Italiano	58
5.2.2 Alto t-score.....	59
5.2.2.1 Inglese	59
5.2.2.2 Italiano	60
5.2.3 Alto MI.....	61
5.2.3.1 Inglese	61
5.2.3.2 Italiano	62
5.3 Analisi qualitativa.....	63
5.3.1 Bigrammi non frequenti/non attestati.....	67
5.3.1.1 Inglese	67
5.3.1.2 Italiano	78
5.3.2 Alto t-score.....	91
5.3.2.1 Inglese	91
5.3.2.2 Italiano	97
5.3.3 Alto MI.....	108
5.3.3.1 Inglese	108
5.3.3.2 Italiano	119
5.3.4 Riassunto risultati analisi qualitativa	130

6 Discussione.....	141
6.1 Introduzione al capitolo	141
6.2 Discussione analisi quantitativa	141
6.3 Discussione analisi qualitativa	142
6.4 Prospettive future	145
7 Conclusione.....	149
Bibliografia e sitografia	151
Appendice	I
EPTIC_01_2011 Guidelines	I

1 Introduzione

Il presente lavoro ha come scopo uno studio sulle differenze nell'utilizzo di pattern fraseologici da parte di traduttori e interpreti attraverso il corpus intermodale EPTIC_01_2011, pensato come uno sviluppo dei corpora EPIC ed EPTIC.

Il capitolo 1 passa in rassegna i principali studi e le principali teorie relativi alla linguistica dei corpora e alle collocazioni. Dopo aver proposto una definizione di corpus, fornisce una panoramica sulle varie tipologie, soffermandosi in particolare sul tipo più recente: il corpus intermodale. Nella seconda parte, si concentra sulle collocazioni, offrendone le principali descrizioni proposte dagli studiosi più in vista nel campo della linguistica. Infine, presenta alcuni casi in cui la ricerca e lo studio delle collocazioni si sono avvalsi della linguistica dei corpora.

Il capitolo 2 introduce EPTIC_01_2011, descrivendone forma e struttura. Il capitolo riserva particolare attenzione al processo di preparazione dei testi inclusi nella versione finale del corpus, riportando dettagliatamente le convenzioni di editing applicate e analizzando i metadati che accompagnano ciascun testo.

Il capitolo 3 riporta la metodologia impiegata per la ricerca delle collocazioni e per un'analisi delle stesse da un punto di vista sia quantitativo che qualitativo. Tale metodologia è stata ripresa e adattata da quella esposta nello studio delle collocazioni effettuato da Durrant e Schmitt nel 2009.

Il capitolo 4 riporta i risultati delle analisi di cui sopra corredati da esempi e grafici. Buona parte del capitolo è incentrata sui fenomeni osservati durante l'analisi qualitativa di alcuni

bigrammi contigui aggettivo+nome/nome+aggettivo estratti applicando la metodologia presentata nel capitolo 3.

Il capitolo 5, infine, riassume e classifica i fenomeni presentati nel capitolo 4. Inoltre, ha come scopo la discussione delle cause che possono presumibilmente aver dato origine a tali fenomeni. L'ultima parte del capitolo contiene idee per ulteriori analisi basate su EPTIC_01_2011 e per possibili sviluppi futuri dello stesso.

2 Background teorico

2.1 Introduzione al capitolo

Il presente capitolo ha come obiettivo quello di fornire il background teorico su cui si basa la ricerca oggetto di questa tesi. Innanzitutto si fornirà una panoramica sui principali studiosi che, grazie alle loro teorie, hanno dato vita ai Corpus-based Translation Studies (CTS). In seguito, si passeranno in rassegna i momenti salienti che hanno caratterizzato la seppur breve storia dei Corpus-based Interpreting Studies (CIS), per poi procedere a una classificazione delle diverse tipologie di corpus. Nella seconda parte del capitolo si passerà a un'analisi dei diversi modi di intendere il concetto di collocazione e dei diversi studi condotti su corpora per descriverne l'utilizzo, così da fornire un quadro completo delle teorie applicate alla ricerca effettuata sul corpus EPTIC_01_2011.

2.2 Corpora

2.2.1 Corpus-based Translation Studies

I Corpus-based Translation Studies costituiscono un fenomeno relativamente recente, dovuto principalmente al tardo sviluppo della linguistica dei corpora. Questa infatti nasce negli anni '60 con la creazione dei primi corpora, che sono tuttavia ancora di dimensioni ridotte e analizzabili solo manualmente. Naturalmente, si tratta di primi esperimenti che conosceranno un'espansione graduale negli anni a venire, con massima fioritura negli anni '90, grazie alla messa a punto di strumenti e software per la costruzione e la consultazione dei corpora. Infatti, "to have corpora at one's disposal is only the first step. As long as one has

not developed the right instruments for dealing with them, they remain silent” (Teubert, 1996:242).

I CTS rispondono alla necessità di studiare la lingua nel suo uso corrente, sia esso generale o settoriale, basandosi su grandi raccolte di testi autentici e non solamente su ipotesi e introspezione. Come suggerisce Laviosa, infatti,

the patterning of language use is not accessible to the native speaker or linguist by introspection or through the analysis of individual stances, but can only be discovered from the direct examination of corpus-based word frequencies, concordances and collocations.

(Laviosa, 2002:9)

La nascita dei CTS viene fatta risalire a Mona Baker, la quale vede nello sviluppo dei corpora un metodo per scoprire le caratteristiche del testo tradotto inteso come evento comunicativo mediato (Baker, 1993) attraverso lo studio di fenomeni quali gli universali della traduzione, ossia “features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems” (Baker, 1993:243). Più in generale, la stessa autrice crede che lo studio dei corpora permetta di identificare i principi e i vincoli che regolano il comportamento traduttivo (Baker, 1993).

Ma cosa si intende con il termine “corpus”? Sulla scia di vari tentativi di definizione e in seguito all’evoluzione che i CTS hanno conosciuto nel corso degli anni, un corpus può essere descritto come

any open-ended body of machine-readable full texts analyzable automatically or semi-automatically and sampled in a principled way in order to be maximally representative of the translation phenomenon under examination.

Si passerà ora a parlare di quella che è stata l'evoluzione di un ulteriore ramo della linguistica dei corpora, che ha preso le mosse proprio dai CTS: i Corpus-based Interpreting Studies.

2.2.2 Corpus-based Interpreting Studies

Sulle orme di quanto proposto da Mona Baker per lo sviluppo dei CTS, nel 1998 Miriam Shlesinger auspica l'applicazione della linguistica dei corpora anche allo studio dell'interpretazione, fino a quel momento ostacolata dalla non disponibilità di grandi corpora di interpretazione in formato elettronico (Shlesinger, 1998). Secondo Shlesinger, la linguistica dei corpora applicata allo studio dell'interpretazione costituirebbe una svolta verso l'individuazione delle caratteristiche tipiche di questo genere di parlato mediato, caratteristiche che l'autrice riunisce sotto il nome di *interpretese* (Shlesinger, 2009) in contrapposizione con quelle tipiche del testo in traduzione scritta che Gellerstam nel 1986 aveva denominato *translationese*. Tuttavia, Shlesinger crede che "any developments towards a full-fledged paradigm of corpus-based interpreting studies will depend on our success in overcoming two main obstacles" (Shlesinger, 1998:487). I due principali ostacoli a cui Shlesinger fa riferimento sono la trascrizione che i corpora di interpretazione richiedono e le cosiddette *paralinguistic dimensions* (Shlesinger, 1998). La trascrizione viene considerata un ostacolo in quanto richiede tempo e sforzi, oltre a non rendere conto di alcuni elementi tipici della modalità orale. Quest'ultimo problema si ricollega all'ostacolo costituito dalle *paralinguistic dimensions*, in quanto è difficile rappresentare nel corpus elementi quali la prosodia tipica del parlato interpretato. Per questi motivi, Shlesinger inizialmente propone di limitare i corpora, in termini sia

di dimensioni che di caratteristiche passibili di trascrizione. Come nei CTS, anche nei CIS i primi corpora consistono in

‘manual’ corpora, i.e. sample data and transcripts that [cannot] be studied using corpus linguistics methods. Then, more steps [are] made towards fully-fledged machine-readable corpora [...]. However, general accessibility to these electronic corpora [is] limited and most projects [remain] isolated attempts.

(Bendazzoli e Sandrelli, 2009:3)

Con il passare degli anni, gli ostacoli descritti sopra si sono notevolmente ridotti, grazie allo sviluppo di risorse rese disponibili online e di software per il riconoscimento vocale che possono velocizzare la fase di trascrizione. Inoltre, si è assistito a una progressiva standardizzazione delle convenzioni di trascrizione e degli aspetti da trascrivere. Restano ancora elementi che richiedono un ulteriore perfezionamento. Infatti, ancora oggi alcuni corpora di interpretazione si basano su dati sperimentali creati appositamente per un determinato studio, a causa delle difficoltà riscontrate nella raccolta di materiali autentici. Inoltre, le trascrizioni continuano a rendere conto unicamente di determinati fenomeni a seconda di ciò che si decide di analizzare in un particolare studio (Shlesinger e Ordan, 2012). Ad oggi comunque si può affermare che “[p]er quanto vi sia ancora tanta strada da percorrere, il panorama generale presenta un certo numero di esperienze che potrebbero ispirare futuri progetti di ricerca” (Bendazzoli, 2010:105).

Nella prossima sezione si proporrà una definizione delle varie tipologie di corpus che si sono sviluppate con l’evoluzione della linguistica dei corpora e che sono rilevanti per CTS e CIS.

2.2.3 Tipologie di corpus

Le tipologie di corpus si sono sviluppate progressivamente con l'evoluzione della linguistica dei corpora e con la nascita dei CTS e dei CIS, dando vita a varie classificazioni basate su diversi fattori quali tipo di lingua, numero di lingue coinvolte, tipologia di testi impiegati, ecc..

Mona Baker, auspicando l'applicazione dei corpora allo studio della traduzione, fornisce una sua definizione di corpus paragonabile (*comparable corpus*), che secondo la stessa autrice costituisce lo strumento migliore per esplorare le caratteristiche proprie della traduzione. Baker definisce un corpus paragonabile come

two separate collections of texts in the same language: one corpus consists of original texts in the language in question and the other consists of translations in that language from a given source language or languages

(Baker, 1995:234)

specificando che i testi costituenti il corpus dovrebbero essere rappresentativi dello stesso argomento, della stessa varietà di lingua e dello stesso periodo. Inoltre, questi dovrebbero essere paragonabili anche in termini di lunghezza. Anche Laviosa fornisce una definizione simile di ciò che riunisce sotto il nome di *comparable monolingual corpus*, ossia un corpus

made up of two single monolingual corpora: one translational, the other non-translational. The two corpora are set up according to similar design criteria, e.g. according to text genre, topic, time span, distribution of male and female authors, readership, average number of words in each text.

(Laviosa, 2002:36)

Un altro tipo di corpus che Baker ritiene importante al fine di studiare il processo traduttivo e il relativo prodotto è il corpus parallelo, ossia “testi originali in una lingua source A e le relative traduzioni in una lingua B” (Baker, 1995:230, mia traduzione). Laviosa ritiene che i due tipi di corpus finora descritti abbiano un’importanza complementare nello studio di aspetti diversi del processo traduttivo: mentre i corpora paralleli sono utili al fine di scoprire i cosiddetti *translation equivalents*, i corpora monolingue paragonabili mirano maggiormente a uno studio del contesto in cui tali equivalenti compaiono (Laviosa, 2002). Allo stesso modo Teubert avverte che per quanto affidabili, i corpora paralleli sono comunque formati da traduzioni che “cannot but give a distorted picture of the language they represent” (Teubert, 1996:247) e che quindi devono essere utilizzati solo come risorsa complementare ai corpora paragonabili, che nel suo caso, a differenza di quelli utilizzati da Baker e Laviosa, sono bilingue paragonabili.

Si può affermare che, nella pratica della traduzione, gli equivalenti traduttivi trovati in un corpus parallelo andrebbero sempre riscontrati in un corpus generale di riferimento, ossia un “[c]orpus di grandi dimensioni, rappresentativo dell’uso generale di una lingua e comprensivo di una gamma di tipologie testuali di varia natura” (Bendazzoli, 2010:19). Sempre per quanto riguarda i corpora paralleli, Stefania Gandin opera un’ulteriore classificazione riguardante la direzionalità, riprendendo quella sviluppata da Johansson nel 2003: infatti un corpus parallelo può essere uni-direzionale, contenente una sola lingua source e una sola lingua target; bi-direzionale, ossia testi originali in due lingue e relative traduzioni nelle stesse due lingue; a stella, formato quindi da una sola lingua source corredata da traduzioni in due o

più lingue; a diamante, consistente in testi originali in tre o più lingue e relative traduzioni combinate (Gandin, 2009).

Più recente è la definizione di corpus intermodale, altra risorsa fondamentale per lo studio della traduzione intesa non solo nella sua accezione tradizionale di traduzione scritta, ma comprendente anche l'interpretazione come modalità orale di traduzione. Miriam Shlesinger, ideatrice dei Corpus-based Interpreting Studies, parte da una ridefinizione dei concetti di *comparable* e *parallel corpus* esposti da Baker, per affiancare alla traduzione scritta l'interpretazione. Infatti, i corpora monolingue paragonabili dovrebbero comprendere:

three separate collections of texts in the same language: interpreted speeches from a variety of source languages, original spoken texts produced in similar settings, and written translations of source oral texts delivered in analogous circumstances.

(Laviosa, 2002:29)

Sempre secondo Shlesinger, i corpora paralleli applicabili anche allo studio dell'interpretazione dovrebbero essere così costruiti: come prima, tre sottocorpora, ossia testi in una data lingua source, le relative interpretazioni e le relative traduzioni (Shlesinger, 1998). Nel 2009, l'autrice utilizza per la prima volta la parola *intermodal* e definisce i *comparable intermodal corpora* come "those consisting solely of translations, in different modalities or in different modes [...] based on the *same* source text" (Shlesinger, 2009:240), sottolineando che il termine *translations* in questo caso comprende sia traduzione scritta che traduzione orale, ossia interpretazione. Bendazzoli riprende questa definizione e scrive che un corpus intermodale

[c]ontiene più TA [testi di arrivo] di uno stesso TP [testo di partenza], prodotti attraverso differenti modalità traduttive come,

ad esempio, traduzione scritta, interpretazione simultanea, interpretazione consecutiva, ecc..

(Bendazzoli, 2010:19)

Analizzate le principali tipologie di corpus, l'ultima parte del presente capitolo sarà dedicata a una panoramica sul background teorico relativo alle collocazioni, oggetto di studio del presente lavoro, nonché ad alcune nozioni sull'analisi di queste ultime mediante l'utilizzo di corpora.

2.3 Collocazioni

Le collocazioni costituiscono un aspetto della lingua molto dibattuto ma ancora relativamente poco studiato. Come sottolinea Masini,

esiste una notevole confusione terminologica: talvolta il termine «collocazione» viene usato come iperonimo per qualsiasi tipo di combinazione non libera di parole, altre volte viene invece usato in maniera più specifica.

(Masini, 2012:117)

Le principali correnti che si occupano di collocazioni sono la scuola russa, a cui appartiene Mel'čuk, e la scuola cosiddetta *neo-firthiana*, i cui massimi rappresentanti sono Halliday e Sinclair. Tuttavia, molti altri autori affrontano l'argomento collocazioni in maniera più o meno approfondita.

2.3.1 Definizioni

In linea di massima le collocazioni possono essere definite come “groups of words which frequently occur in combination with each other” (Fontenelle, 1998:191) che, nel cosiddetto *phraseological cline*, o gradiente fraseologico, si trovano a metà strada tra le *free combinations* e gli *idioms*. Cowie distingue due tipi di collocazioni:

le *free collocations* sono costituite da una parola che accetta diversi collocati, mentre le *restricted collocations* sono

word-combinations in which one element (usually the verb) has a technical sense, or a long-established figurative sense which has since lost most of its analogical force.

(Cowie, 1991:102)

Le collocazioni costituiscono un esempio di “linguaggio prefabbricato [che] fa parte della conoscenza dei parlanti ed è pertanto *convenzionalizzato*” (Masini, 2012:61). Dunque, le collocazioni costituiscono combinazioni di parole preferite in una determinata lingua, nonostante l'utilizzo di sinonimi per esprimere lo stesso concetto sarebbe corretto secondo le regole sintattiche della lingua in questione.

Per la scuola russa in generale, le collocazioni sono

combinations in which either words take on meanings which they do not have in other environments [...] or there are arbitrary restrictions on what words can be substituted into a particular phrase.

(Durrant e Schmitt, 2009:159)

Mel'čuk riprende la sua idea di collocazione, da lui chiamata anche *semi-phrase*, dalla definizione di ciò che Vinogradov denominava combinazioni fraseologiche, ossia combinazioni in cui “un componente è usato nel suo significato letterale mentre l'altro è usato in senso figurato, e comunque «legato» fraseologicamente dal contesto” (Masini, 2012:44). Secondo Mel'čuk, infatti, una collocazione formata da due parole “includes 'intact' the meaning of one [of] its constituents, but [...] the other component of its meaning is expressed by an element contingent on the first” (Cowie, 1998:6).

I *firthiani* caratterizzano invece le collocazioni come “words which appear together in the language more often than their individual frequencies would predict” (Durrant e Schmitt, 2009:159), introducendo la frequenza come caratteristica che denota le collocazioni e postulando così l’utilizzo di corpora per la loro individuazione. Sinclair, in particolare, definisce una collocazione come “the occurrence of two or more words within a short space of each other in a text. The usual measure of proximity is a maximum of four words intervening” (Sinclair, 1991:170). Definire uno span di non più di quattro parole è necessario secondo l’autore in quanto “la forza di attrazione dei collocati da parte del nodo diminuisce gradualmente con la distanza” (Sinclair, 1991:106, mia traduzione). Sinclair utilizza i termini nodo e collocati per definire le parti di cui si compone una collocazione: il nodo è la parola oggetto dello studio, mentre con collocati si intendono quelle parole che gravitano intorno al nodo (Sinclair, 1991). Parlando di collocazioni, Sinclair introduce inoltre la nozione di *idiom principle*, in contrapposizione con quella di *open-choice principle*. Infatti, secondo Sinclair, il parlante ha a disposizione “a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments” (Sinclair, 1991:110). Dunque, le parole nel testo si distribuiscono in modo non casuale e il significato globale di una frase deriva non tanto dal significato di ogni sua singola parte, ma dal significato che tali parti assumono in relazione con le altre. Infatti, come fa notare Siepmann, “le parole derivano almeno in parte il loro significato dai contesti in cui solitamente compaiono e, più nello specifico, dai collocati che abitualmente le accompagnano” (Siepmann, 2005:6, mia traduzione). Per Sinclair le collocazioni sarebbero la prova della validità dell’*idiom principle*, in quanto si tratta di parole utilizzate in

coppia o in gruppo (Sinclair, 1991). Tuttavia, le collocazioni si distinguono dagli *idioms*, poiché questi ultimi hanno un significato diverso rispetto al significato delle singole parole che li compongono, mentre le collocazioni riflettono in una certa misura il significato delle singole parole.

Come già affermato, altri autori si occupano di collocazioni nei loro studi. La definizione fornita da Hausmann è un sunto dei pensieri portati avanti da Mel'čuk e da Sinclair. Infatti Hausmann stabilisce che le collocazioni consistono nella

tendency of two words, called the 'base' (or 'node') and the 'collocate' to co-occur; [...] semi-prepackaged phrases that speakers do not put together creatively. The base is semantically autonomous [...], while the meaning conveyed by the collocate is dependent on its co-occurrence with the base [...].

(Siepmann, 2005:45)

Altenberg a sua volta, pur non facendo esplicito riferimento alle collocazioni, parla di *recurrent word-combinations* ossia "qualsiasi stringa continua di parole che compaia più di una volta nella stessa forma" (Altenberg, 1998:101, mia traduzione), restringendo così il concetto di collocazione elaborato da Sinclair che stava a includere anche collocati in uno span di quattro parole dal nodo. Inoltre, Altenberg descrive le collocazioni come "seldom completely fixed but [...] 'preferred' ways of saying things" (Altenberg, 1998:121-122). L'autore attribuisce a tali combinazioni una funzione pragmatica all'interno della lingua, in quanto i parlanti hanno bisogno di alcune espressioni convenzionalizzate per esprimere all'interlocutore determinate intenzioni o reazioni (Altenberg, 1998).

Anche Bahns usa l'espressione *recurrent combinations* come sinonimo di collocazioni. L'autore le definisce come "fixed,

identifiable, non-idiomatic phrases and constructions” (Bahns, 1993:57) e ne propone una divisione in due gruppi: *grammatical collocations* e *lexical collocations*. Queste ultime non contengono le cosiddette parole funzionali, ma consistono in combinazioni di nomi, aggettivi, verbi e avverbi (Bahns, 1993). Le principali caratteristiche delle collocazioni secondo Bahns sono le seguenti:

their meanings reflect the meaning of their constituent parts (in contrast to idioms) and [...] they are used frequently, spring to mind readily, and are psychologically salient (in contrast to free combinations).

(Bahns, 1993:57).

Van Roey (1990, citato da Granger) descrive una collocazione come:

the linguistic phenomenon whereby a given vocabulary item prefers the company of another item rather than its “synonyms” because of constraints which are not on the level of syntax or conceptual meaning but in that of usage

(Granger, 1998:146-147)

e distingue vari gradi di collocabilità, da *restricted collocability* a *more open collocability*.

Anche Masini, dopo aver passato in rassegna le varie correnti di pensiero che si sono occupate del concetto di collocazione, propone una sua definizione:

combinazioni (tipicamente sequenze testa+modificatore) che sono dotate di particolare coesione interna, ma che si distinguono dalle espressioni libere per via di un meccanismo di selezione lessicale *obbligatorio* tra i costituenti che l[e] compongono.

(Masini, 2012:120)

Masini decide di non classificare le collocazioni all'interno delle espressioni multiparola di cui si occupa nella sua analisi “in

quanto non sono «costruzioni», ma il frutto di una (co-)selezione tra lessemi/costruzioni” (Masini, 2012:122).

Riassumendo, vengono considerate collocazioni quelle coppie o quei gruppi di parole che, siano esse contigue o no, occorrono con una certa frequenza in determinati contesti. Si tratta di parole che si scelgono non tanto per motivi di sintassi quanto di utilizzo e che non perdono completamente il loro significato originale ma possono assumere sfumature diverse in base all’occorrenza con determinati collocati.

Si vedrà ora come le collocazioni sono state studiate in diversi contesti di utilizzo mediante l’analisi di corpora.

2.3.2 Collocazioni e corpora

I corpora sono uno strumento utile per lo studio delle collocazioni, ma è necessario disporre di corpora di dimensioni considerevoli per avere dati sufficienti al fine di effettuare un’analisi statistica (Sinclair, 1991). Solitamente, infatti, corpora di dimensioni ridotte vengono utilizzati per studi a livello grammaticale, in quanto “la frequenza di occorrenza delle cosiddette parole grammaticali o funzionali risulta essere abbastanza elevata” (Sinclair, 1991:100, mia traduzione).

Lo studio dei corpora da parte della scuola *firthiana* ha rivelato che “between 50 and 80 per cent of all text is made up of habitual word associations” (Siepmann, 2005:6), dato a sostegno dell’*idiom principle* postulato da Sinclair. Come già accennato in precedenza, Sinclair, basando la sua ricerca sull’utilizzo di corpora, considera potenziali collocati di un determinato nodo tutte le parole che si trovano entro uno span di -4/+4 dal nodo stesso. La stessa unità di misura viene ripresa da Smadja per cui due parole possono essere considerate una collocazione se si trovano

nello stesso periodo e se tra loro intercorrono meno di cinque parole (Fontenelle, 1998). La nozione di span è importante anche per poter parlare di *multi-word collocations*, in contrapposizione con la teoria di Aisenstadt, il quale pensa che le collocazioni possano essere unicamente combinazioni di due parole e che quindi i gruppi formati da, per esempio, verbo+preposizione+nome “can be viewed as two-word collocations in which the use of the preposition may be determined by either element” (Siepmann, 2005:46).

Sono stati effettuati diversi studi sulle collocazioni tramite l’ausilio dei corpora, andando a ricercare vari fenomeni negli ambiti più disparati.

Moon e Gandin si servono di corpora per la ricerca di collocazioni. La prima constata che mentre le collocazioni di tipo grammaticale risultano essere frequenti, altre collocazioni, oltre a essere meno frequenti, si prestano a modifiche sia a livello strutturale che a livello lessicale a seconda del contesto in cui sono inserite. Moon distingue tra *anomalous collocations* (simili alle *restricted collocations* di Cowie; si veda 2.3.1), *formulae* e *metaphors* a seconda del grado di fissità che queste presentano (Moon, 1998). La seconda, invece, giunge alla conclusione che i corpora risultano utili tra le altre cose anche per “identificare le collocazioni terminologiche tipiche del genere analizzato” (Gandin, 2009:142).

Siepmann a sua volta studia l’utilizzo in varie lingue di un tipo particolare di fraseologia, i *second-level discourse markers*, ovvero

medium-frequency fixed expressions or collocations composed of two or more printed words acting as a single unit [which] facilitate the process of interpreting coherence relations.

(Siepmann, 2005:52)

Il particolare oggetto dello studio portato avanti da Siepmann lo obbliga a riformulare l'*idiom principle* di Sinclair, nonché la nozione di span difesa da quest'ultimo. Infatti, in casi come questi non si potrà utilizzare la *vicinity* come elemento discriminante nell'individuare eventuali collocazioni, in quanto "lexical dependencies [...] manifest themselves not only between immediately adjacent items, but also over greater distances" (Siepmann, 2005:101). In questo modo anche lo span di -4/+4 deve essere adattato alle caratteristiche dello studio in questione: l'autore parla quindi di uno span che può coprire anche diversi paragrafi (Siepmann, 2005).

L'analisi di corpora per l'individuazione di collocazioni ha avuto inoltre molto successo nell'evoluzione della lessicografia e della composizione dei dizionari bilingue poiché "the ultimate goal is to find the TE [translation equivalent] of a larger semantic unit, not just of a word" (Teubert, 1996:248). Nel caso della ricerca di equivalenti traduttivi, Teubert ritiene che il miglior tipo di corpus da analizzare sia quello parallelo, in quanto fornisce non solo l'equivalente traduttivo in una determinata lingua target, ma anche il contesto di utilizzo. Occorre comunque sempre tenere a mente le questioni relative all'affidabilità di tali corpora di cui si è discusso in precedenza (2.2.3).

2.3.2.1 Le collocazioni dei parlanti non nativi

Granger (1998) e Howarth (1998) si propongono di ricercare l'utilizzo della fraseologia da parte di parlanti non nativi. Granger ipotizza che i parlanti non nativi usino meno "prefabricated language (collocations and formulae)" (Cowie, 1998:13). Inoltre, riprendendo Bahns (1993), l'autrice fa notare che la fraseologia

utilizzata dai parlanti in esame è solitamente quella che ha una traduzione diretta nella lingua madre degli stessi e che i parlanti non nativi tendono a utilizzare la fraseologia a loro più familiare, ossia “cling on’ to certain fixed phrases and expressions which they feel confident in using” (Granger, 1998:156). Queste caratteristiche portano a una minore naturalezza dei testi prodotti dai parlanti non nativi. Howarth, dal canto suo, conclude che non sono gli *idioms* a creare problemi ai parlanti non nativi, poiché sono fissi e costituiscono solo una minima parte della fraseologia di una lingua. La difficoltà maggiore sarebbe invece rappresentata dalle *restricted collocations*, la parte più consistente della fraseologia. Nel caso di questo tipo di collocazioni è infatti necessario sapere “when, and how far, the elements of a collocation are able to recombine with other items” (Cowie, 1998:15).

Anche Durrant e Schmitt basano il loro studio delle collocazioni sull’analisi di testi prodotti da parlanti madrelingua e non, giungendo alle stesse conclusioni di Granger, ossia che i parlanti non nativi tendono a usare collocazioni a più alta frequenza d’uso e a ripetere sempre le stesse collocazioni con cui hanno familiarità (Durrant e Schmitt, 2009). Anche dallo studio di Siepmann citato più in alto, nel confronto tra la produzione di parlanti nativi e non, risulta che le collocazioni utilizzate dai non nativi sono poche e ripetute, solitamente coincidenti con espressioni che hanno un equivalente traduttivo nella lingua madre. A volte le collocazioni tendono invece a non essere riconosciute come tali e vengono trattate come semplici *free combinations* (Siepmann, 2005).

2.3.2.2 Le collocazioni in traduzione e interpretazione

Diverso è il caso riguardante l'utilizzo delle collocazioni da parte di traduttori e interpreti, che solitamente sono non nativi rispetto alla lingua di partenza ma nativi rispetto alla lingua di arrivo e dovrebbero quindi produrre testi caratterizzati dalle collocazioni tipiche della propria madrelingua. A questo proposito, uno studio simile a quello di Durrant e Schmitt presentato in precedenza viene effettuato da Dayrell. L'autrice giunge alla conclusione che "collocational patterns tend to be less diverse in translated texts in comparison with non-translated texts of the same language" (Dayrell, 2007:395) e che il comportamento dei traduttori sembra essere influenzato dalla lingua del testo di partenza.

Sempre riguardo alla traduzione, ma in una diversa modalità, quella interpretata, lo studio delle collocazioni ha portato alla conclusione che gli interpreti, a causa delle caratteristiche del genere (minor tempo a disposizione rispetto ai traduttori), optano per *cognate solutions*, *transcoding*, espressioni più colloquiali o parafrasi anziché cercare l'equivalente traduttivo specifico, limitando così la varietà lessicale della propria produzione (Shlesinger, 2009; Shlesinger e Ordan, 2012). Inoltre, mettendo a confronto il prodotto delle due modalità di traduzione ottenuto dalla stessa persona, emerge che esiste "a clear preference for the unmarked form when interpreting but a clear preference for a formal, marked alternative when translating" (Shlesinger, 2009:248-249).

Il prossimo capitolo si focalizzerà sulla descrizione del corpus EPTIC_01_2011 e sulla sua costruzione.

3 EPTIC_01_2011

3.1 Introduzione al capitolo

Il presente capitolo ha come oggetto la fase di preparazione del materiale da analizzare nello studio in questione. Dopo una presentazione del corpus nel suo insieme, in termini di tipologia e struttura (3.2 e 3.3), si andranno ad analizzare più in dettaglio le diverse parti che lo compongono con particolare attenzione alle convenzioni di editing, impiegate in fase di trascrizione e di modifica dei testi raccolti, e ai metadati, necessari alla ricerca di determinati testi nel corpus (3.4).

3.2 Il corpus

La fase di reperimento e preparazione dei testi, nonché la fase decisionale in termini di convenzioni da utilizzare, sono state svolte in collaborazione con Rita Micchi, di modo da ottenere un prodotto finale uniforme e standardizzato su cui svolgere l'analisi in questione. Si è voluto portare avanti ed espandere il progetto EPTIC – European Parliament Translation and Interpreting Corpus (Bernardini et al., accepted), che, a sua volta, consiste in un'evoluzione del corpus EPIC – European Parliament Interpreting Corpus realizzato dall'Università di Bologna tra il 2004 e il 2006 (Sandrelli e Bendazzoli, 2005).

Tale espansione di EPTIC, denominata EPTIC_01_2011, consiste in un corpus trilingue, bidirezionale, intermodale, paragonabile e parallelo (per la distinzione tra le varie tipologie di corpus si veda 2.2.3).

Si è deciso di tornare alla versione trilingue come era per EPIC, sostituendo però lo spagnolo con il francese, in modo da poter studiare i fenomeni delle proprie lingue di lavoro. La scelta di

riprendere il trilinguismo è stata dettata dalla volontà di arricchire il corpus EPTIC e preparare il terreno per un'ulteriore futura espansione del progetto.

Per quanto riguarda la bidirezionalità, va fatta un'osservazione: mentre in EPIC venivano sfruttate tutte le combinazioni linguistiche (inglese<>italiano, spagnolo<>italiano e inglese<>spagnolo), nel nostro caso abbiamo deciso di lavorare solamente con le combinazioni che includono l'italiano (inglese<>italiano e francese<>italiano), lasciando da parte la combinazione tra le due lingue straniere (inglese<>francese). È stata preferita questa soluzione perché l'espansione del progetto è stata incentrata sulla lingua italiana. Non è però da escludersi, ma è anzi auspicabile, una futura collaborazione per l'inserimento non solo della combinazione inglese<>francese ma anche di ulteriori lingue volte ad alimentare e rendere sempre più completo il corpus.

Inoltre, come nel caso di EPTIC, si è deciso di mettere a punto un corpus intermodale, ovvero costituito sia da interpretazioni che da traduzioni, con la possibilità quindi di operare un'analisi contrastiva delle due modalità di mediazione.

EPTIC_01_2011 è altresì un corpus paragonabile, ovvero un corpus "in cui originali in una lingua sono esaminati a fianco di traduzioni verso quella lingua di testi appartenenti allo stesso genere" (Zanettin, 2001), dove il termine "traduzione" è utilizzato in senso lato, includendo sia le interpretazioni che le traduzioni (scritte). Infine, si tratta di un corpus parallelo, formato cioè da testi originali e relative traduzioni/interpretazioni.

Andremo ora ad analizzare più nel dettaglio la struttura di EPTIC_01_2011.

3.3 Struttura del corpus

Come già affermato in precedenza (3.2), essendo intermodale e trilingue, il corpus contiene interpretazioni e traduzioni di discorsi originali in italiano, inglese e francese, per un totale di 14 sottocorpora (si veda diagramma in Figura 1) e all'incirca 280.000 parole (per maggiori informazioni sul conteggio delle parole si faccia riferimento alla Tabella 3.1 e a 3.4.4.1).

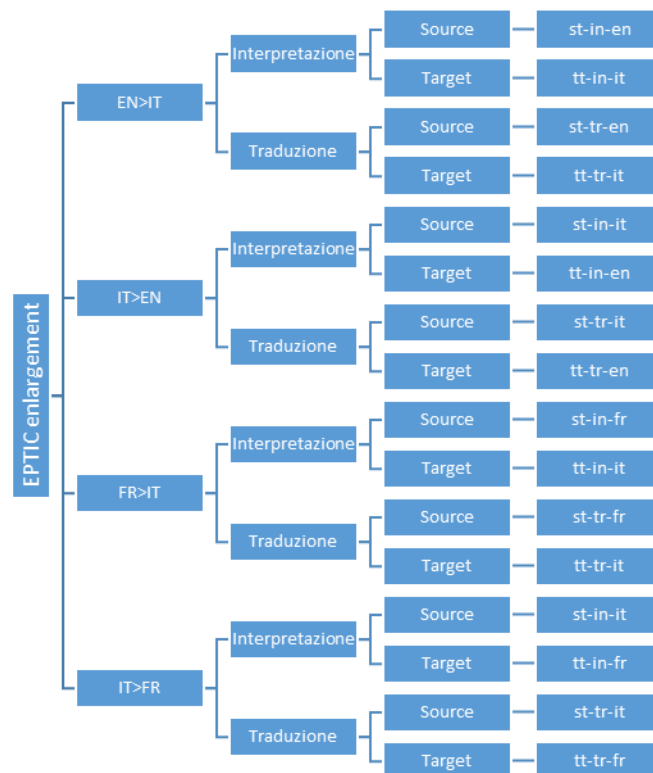


Figura 3.1 - Struttura di EPTIC

Leggendo il diagramma, i sottocorpora risultano essere 16, ma si deve tenere conto del fatto che per le direzioni italiano>inglese e italiano>francese, sia nel sottocorpus di interpretazione sia in quello di traduzione, si parte dallo stesso source. Quindi il sottocorpus st-in-it della direzione IT>EN è lo stesso del st-in-it utilizzato per la direzione IT>FR e lo stesso vale per il sottocorpus st-tr-it nelle medesime direzioni. I suddetti sottocorpora sono quindi stati inseriti due volte nel diagramma unicamente per motivi di chiarezza e completezza. Per quanto

riguarda il numero di discorsi che si è deciso di trascrivere, questo oscilla tra i 60 e i 70, a seconda della lingua di partenza. Infatti, si è notato che i discorsi originali in lingua inglese sono generalmente lunghi e oscillano quasi tutti tra le 700 e le 1.000 parole e oltre, mentre quelli in lingua francese tendono a essere di media lunghezza (300-700 parole). Quelli in lingua italiana, invece, sono quasi tutti brevi (meno di 300 parole). Per avere quindi un numero di parole bilanciato nei vari sottocorpora di interpretazione, il numero di testi raccolti per l'inglese si attesta a 62, mentre per la lingua francese e quella italiana sono stati presi in considerazione rispettivamente 65 e 68 testi. Gli stessi numeri valgono naturalmente per i sottocorpora di traduzione. Sorprendentemente, al momento del conteggio parole, è risultato che i discorsi originali in lingua francese superano per numero di parole quelli in lingua inglese. Dunque, probabilmente per il francese sarebbe stato sufficiente trascrivere un numero di interventi pari a quelli trascritti per l'inglese. L'italiano rimane la parte meno consistente del corpus se si fa riferimento ai numeri relativi ai source, mentre se si guarda all'intero sottocorpus avente come source la lingua italiana, esso risulta essere il più ampio. Quest'ultima caratteristica è tuttavia dovuta unicamente al fatto che il suddetto sottocorpus comprende i target di entrambe le altre lingue del progetto. Si riporta qui di seguito una tabella volta a riassumere i dati relativi alle dimensioni di EPTIC_01_2011, nonché dei diversi sottocorpora che lo compongono (si veda Tabella 3.1), e un grafico relativo alle percentuali coperte dalle diverse lingue source (si veda Figura 3.2).

Sottocorpus	N. testi	Parole	% EPTIC_01_2011
st-in-en	62	21.454	7,65

tt-in-it	62	17.958	6,41
st-tr-en	62	20.339	7,25
tt-tr-it	62	20.777	7,41
Subtotale	248	80.528	28,72
st-in-fr	65	24.686	8,81
tt-in-it	65	19.558	6,98
st-tr-fr	65	23.979	8,55
tt-tr-it	65	21.618	7,71
Subtotale	260	89.841	32,05
st-in-it	68	18.028	6,43
tt-in-en	68	16.800	5,99
tt-in-fr	68	19.344	6,90
st-tr-it	68	17.687	6,31
tt-tr-en	68	18.340	6,54
tt-tr-fr	68	20.544	7,33
Subtotale	408	110.743	39,50
TOTALE	916	280.346	100

Tabella 3.1 - Conteggio parole EPTIC¹

¹ I conteggi sono stati arrotondati per difetto o per eccesso al secondo decimale.

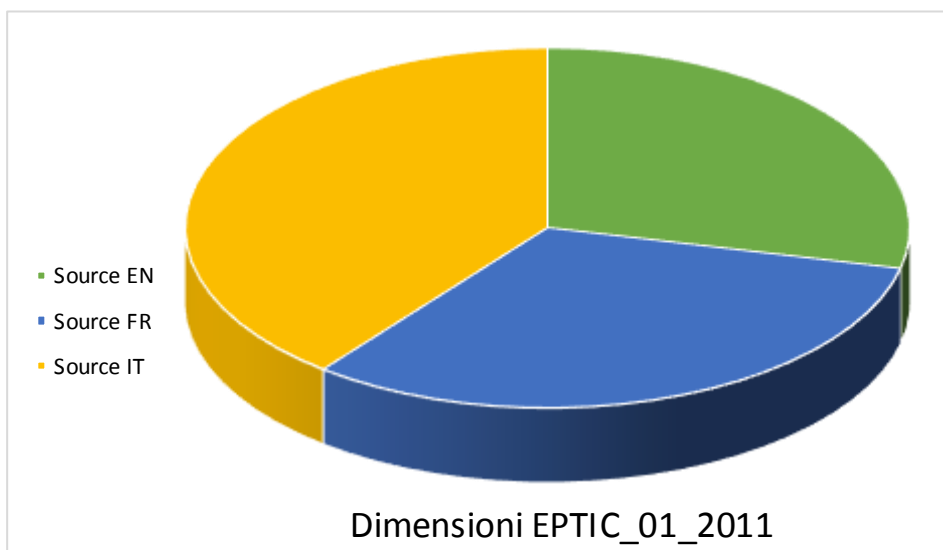


Figura 3.2 - Grafico percentuali lingue source

3.4 Preparazione del corpus

Per la preparazione del corpus, sia per quanto riguarda le interpretazioni che le traduzioni, si è deciso di raccogliere i testi necessari per le tre lingue tra gli interventi di una seduta plenaria della settima legislatura (2009-2014) del Parlamento europeo, tenutasi a Strasburgo tra il 17 e il 20 gennaio 2011. Sono stati scelti testi più recenti rispetto a EPIC ed EPTIC perché più rilevanti e rappresentativi ai fini dell'analisi oggetto della presente tesi sia in termini di lingua che di tematiche affrontate. Non è stato tuttavia possibile scegliere una seduta plenaria tenutasi in anni più recenti, poiché al momento della scelta delle giornate, ovvero novembre 2013, ci si è resi conto del fatto che il Parlamento europeo aveva reso disponibili sul suo sito i verbatim report, necessari alla costruzione dei sottocorpora di traduzione, solamente fino alla seduta plenaria del mese di giugno 2011. Allo stesso modo, anche il sito da cui riprendere le trascrizioni da modificare per costruire i sottocorpora di interpretazione viene aggiornato solo periodicamente, in base a quando i traduttori delle varie lingue rendono disponibili le traduzioni dei verbatim report. Inoltre, l'idea iniziale era quella di raccogliere un numero maggiore

di testi e dunque confrontarsi con più di una seduta plenaria, avendo così a disposizione discorsi più variati sia in termini di oratori sia in termini di tematiche affrontate in sede di dibattito. Si è però deciso di portare a termine solamente la prima seduta plenaria dell'anno 2011 consistente nelle quattro giornate di cui sopra, perché i testi raccolti sarebbero stati sufficienti per un primo studio esplorativo, sempre lasciando aperta la strada per future collaborazioni volte a implementare EPTIC_01_2011. Per coprire il numero di testi fissati (3.3), le giornate del 17 e del 18 gennaio sono state utilizzate per tutte le lingue, la giornata del 19 gennaio è stata utilizzata per il francese e l'italiano, mentre dalla giornata del 20 sono stati raccolti solamente i testi necessari per completare i sottocorpora di francese.

Nello scegliere i testi che sarebbero andati a formare il corpus si è deciso di non raccogliere gli interventi del Presidente e dei vicepresidenti quando rivestono il ruolo di moderatori in sede di dibattito, nonché i testi più corti di 60 parole, in quanto troppo brevi, ripetitivi e contenenti poche parole semanticamente piene. Inoltre, per quanto riguarda gli interventi dei moderatori del dibattito, la versione riportata nei verbatim report non corrisponde a ciò che Presidente e vicepresidenti dicono realmente in sede di dibattito. Tale decisione è volta inoltre a evitare un'eccessiva frammentazione del corpus.

Si vedrà ora più nel dettaglio come sono stati costruiti i sottocorpora di interpretazione e traduzione.

3.4.1 Sottocorpora di interpretazione

Per i sottocorpora di interpretazione, il reperimento dei testi è stato più facile rispetto alla costruzione di EPIC. Infatti, da qualche anno il Parlamento europeo mette a disposizione sul suo sito web

i video delle sedute con le relative trascrizioni ripulite.² Un potenziale problema è però costituito dal fatto che i video sono disponibili solamente in streaming e il loro utilizzo dipende dunque dalla possibilità di fare uso di una connessione internet e dal funzionamento del sito web che li ospita, mentre con gli audio registrati utilizzati per la costruzione di EPIC era possibile lavorare anche offline.

Una volta decise le giornate su cui lavorare, le trascrizioni fornite dal sito sono state manipolate in Notepad++³ secondo l'audio dell'oratore o dell'interprete e applicando alcune convenzioni di editing (3.4.3). Si riporta qui di seguito (Tabella 3.2) un esempio del testo così come reperibile sul sito del Parlamento europeo affiancato dallo stesso frammento di testo modificato secondo le convenzioni di editing di cui si discuterà in seguito.

Testo originale	Testo modificato con convenzioni
<p>Ciononostante, il "sì" ha vinto con solo il 54% dei voti. Con questo accordo, l'Italia dei diritti torna indietro di parecchi decenni. È giusto che l'Europa sappia e ne tenga conto.</p>	<p>Ciononostante, il "sì" ha vinto con solo il 50 per ce- il 54 per cento dei voti. Con questo accordo, l'Italia dei diritti torna indietro di parecchi decenni. È giusto che l'Europa /europia/ sappia e ne tenga conto.</p>

Tabella 3.2 - Testo prima e dopo la modifica

² È possibile accedere ai video e alle trascrizioni delle sedute all'indirizzo <http://www.europarl.europa.eu/plenary/it/debates-video.html> impostando la legislatura e le date di interesse, scegliendo uno dei documenti relativi alla giornata prescelta, cliccando su "Testo integrale" e cambiando lingua a seconda delle esigenze.

³ Software gratuito scaricabile al link <http://notepad-plus-plus.org/>

3.4.2 Sottocorpora di traduzione

Per quanto riguarda i sottocorpora di traduzione, il lavoro è stato ancora più semplice rispetto a quello necessario per la costruzione dei sottocorpora di interpretazione. I testi, corrispondenti agli originali del sottocorpus di interpretazione, e le relative traduzioni sono stati ripresi dai verbatim report scaricabili dal sito del Parlamento.⁴ I verbatim report consistono in trascrizioni degli interventi ripulite da tutte le caratteristiche proprie dell'oralità, quali esitazioni, ripetizioni ed errori. Le traduzioni di tali testi ripuliti vengono elaborate indipendentemente dalle interpretazioni fornite durante la seduta. Per costruire i sottocorpora di traduzione, si è proceduto a una rilettura dei testi scaricati, modificandoli in Notepad++ secondo alcune convenzioni di editing (3.4.3). Naturalmente, le convenzioni applicate al sottocorpus di traduzione non comprendono quelle strettamente legate all'oralità, quali trascrizione di disfluenze, pause, esitazioni, ecc.. In caso di errori del traduttore, non sono state apportate modifiche al testo, ma i problemi incontrati sono stati segnalati nell'attributo "comments" dei metadati (3.4.4).

Vediamo ora quali sono le convenzioni di editing utilizzate per manipolare i testi da inserire in EPTIC_01_2011.

3.4.3 Convenzioni di editing

Per modificare i testi, sono state impiegate in linea di massima le convenzioni di trascrizione utilizzate per EPIC⁵ (si veda Tabella 3.3) e per la relativa implementazione (Russo et al., 2010:4). Tali

⁴ È possibile scaricare i verbatim reports al seguente link <http://www.europarl.europa.eu/RegistreWeb/search/typedoc.htm?codeTypeDocu=PCRE>, scegliendo l'anno di interesse, la lingua del documento e filtrando i risultati in ordine cronologico.

⁵ http://sslmitdev-online.sslmit.unibo.it/corpora/additionalpages.php?path=E.P.I.C.&source=project&content=transcription_conventions.desc

convenzioni, modificate e aggiornate a seconda delle esigenze, sono state utilizzate non solo per i sottocorpora di interpretazione, ma, quando necessario, anche per la manipolazione dei testi inclusi nei sottocorpora di traduzione.

SPEECH FEATURE	EXAMPLE	TRANSCRIPTION CONVENTION
Word truncations	propo pro posal	propo- proposal /pro_posal/
Pronunciation disfluencies	Parlomento	Parlamento /Parlomento/
Pauses	(filled / empty)	ehm ...
Numbers figures dates	532 4% 1997	five hundred and thirty-two four per cent nineteen ninety-nine
Unintelligible		#
Units	based on syntax & intonation	//

Tabella 3.3 - Convenzioni trascrizione EPIC

3.4.3.1 Parole troncate

Le convenzioni fissate per indicare parole troncate sono state mantenute. Quando l'oratore comincia a pronunciare una parola ma non la termina, si scrive ciò che si sente seguito da un trattino ("propo-"), mentre nel caso in cui l'oratore inizi a pronunciare una parola, faccia una pausa e finisca di pronunciarla, si scrive la parola in modo convenzionale seguita dalla stessa parola tra barre, con un trattino basso a indicare il punto in cui l'oratore si interrompe ("proposal /pro_posal/").

3.4.3.2 Disfluenze

Per quanto riguarda le cosiddette disfluenze, sono stati assunti comportamenti diversi. In caso di parole pronunciate in modo sbagliato e per cui non esiste un corrispondente nella lingua in

questione si inserisce nella trascrizione la parola corretta seguita dalla parola con pronuncia sbagliata tra slash (“Parlamento /parlomento/”). Ciò che si trova tra slash non costituisce una vera e propria trascrizione fonetica, ma la trascrizione della parola originale con le modifiche necessarie per riprodurre la pronuncia dell’oratore o dell’interprete. Nel caso della parola inglese “Cameroon”, per esempio, l’accento cadrebbe sulla sillaba finale (“-roon”). In un intervento, tuttavia, l’oratore pronuncia la parola mettendo l’accento sulla prima sillaba: la parola è stata quindi trascritta con grafia corretta seguita da “/càmeroon/”, a indicare come l’oratore ha effettivamente pronunciato la parola in questione. Si noti che ciò che compare tra slash è sempre riportato con la lettera minuscola, dato che la funzione degli slash e del loro contenuto è quella di riprodurre un suono e non di rispecchiare fedelmente le regole di ortografia. Per la lingua francese è necessario applicare una regola della fonetica nel caso della liaison tra la “s” e una vocale per distinguere il fenomeno dal suono puro della “s”: si avrà quindi, per esempio, “/vou_zauriez/” (in grafia standard “vous auriez”) in contrapposizione con “/vou_sauriez/” (nel caso di “vous sauriez”). La decisione di non utilizzare l’alfabeto fonetico è dovuta a una volontà di maggiore chiarezza e apertura verso un pubblico di “non addetti ai lavori”, nonché alla difficoltà di lettura di un alfabeto non convenzionale da parte del computer e dei programmi utilizzati per la costruzione e lo studio del corpus. Non sono considerate errore le pronunce sbagliate degli oratori non madrelingua, in quanto da considerarsi una marca di stile dovuta alle origini del parlante. Sempre tra slash si riportano casi di suoni allungati (ad esempio, “in questo /questoo/ progetto”) e casi in cui due parole vengono per sbaglio fuse formandone una nuova non esistente nel vocabolario di una determinata lingua (“evitare eventuali /eventuare/”). Per quanto

riguarda la lingua inglese, tra slash si trovano spesso anche le pronunce non convenzionali di numeri, per esempio “2010 /twenty ten/” e “1200 /twelve hundred/”. Nel caso invece di errori relativi a parole esistenti nel vocabolario della lingua in questione, nonostante le parole effettivamente pronunciate non abbiano senso nel contesto o non siano ciò che l’oratore aveva intenzione di dire, si è optato per l’inserimento di tali parole nella trascrizione senza ulteriori aggiunte o precisazioni. Per esempio, un caso di questo tipo sarebbe quello in cui l’oratore o l’interprete pronuncia “il vario di una riforma”, mentre avrebbe dovuto dire “varo”. Si è deciso di non segnalare il fenomeno essendo facile reperire tali parole in fase di ricerca e poiché a volte la scelta del termine che l’oratore avrebbe voluto usare risulta arbitraria.

In generale, gli errori di pronuncia e le disfluenze sono stati inseriti tra slash per agevolare la ricerca e la fase di POS-tagging.

3.4.3.3 Pause

Per quanto concerne la trascrizione delle pause, siano esse piene o vuote, sono state mantenute le convenzioni fissate per EPIC. Bisogna tuttavia specificare che i puntini di sospensione indicanti le pause vuote devono essere inseriti solamente nel caso in cui un ragionamento o un periodo siano interrotti in un punto non naturale: un esempio è “abbiamo varato una... riforma”, in contrapposizione con “abbiamo varato una riforma... perché”. Inoltre, i puntini vengono inseriti solamente nel caso in cui la pausa si estenda per un lasso di tempo che colpisce l’uditore in quanto innaturale. Rispetto alle convenzioni originali, i tre puntini di sospensione vengono utilizzati inoltre alla fine di determinati interventi, stando a indicare il punto in cui l’oratore viene interrotto dal Presidente, dando così vita a una frase non finita e priva di

senso compiuto. Nelle interpretazioni corrispondenti, nella maggior parte dei casi, il discorso termina invece con un punto fermo, dato che l'interprete, grazie al *décalage*, riesce a cogliere l'interruzione del Presidente e dunque a formulare comunque una frase di senso compiuto.

3.4.3.4 Numeri

Le convenzioni riguardanti i numeri sono state modificate: tutti i numeri, ad eccezione delle frazioni, vengono mantenuti o trasformati in cifre e non trascritti in lettere come in EPIC al fine di rispecchiare non solo l'orale ma anche lo scritto, essendo il corpus formato non solo da interpretazioni e discorsi originali, ma anche da traduzioni scritte e rispettivi testi originali. Per indicare le migliaia e i decimali vengono utilizzate le convenzioni vigenti in ciascuna lingua (per i decimali punto in inglese e virgola in italiano e francese, mentre per le migliaia punto in italiano e francese e virgola in inglese). Per i milioni e i miliardi, tuttavia, si è deciso di trascrivere in parola le diciture "milione"/"milioni" e "miliardo"/"miliardi": nel caso per esempio di 12.000.000 si troverà "12 milioni". Nelle percentuali, si mantiene il numero in cifre e si scrive il simbolo "%" in parola ("per cento", "percent" e "pour cent"). Attenzione particolare meritano i numeri ordinali. Se questi fanno parte di date o si tratta di numeri di congressi, leggi o direttive, si trascrive il numero ordinale secondo le convenzioni di ciascuna lingua, ossia numero in cifre seguito dall'abbreviazione in lettere degli ordinali (per l'italiano o/a; per l'inglese st/nd/rd/th; per il francese er/ère/ème). In inglese e in francese vengono indicati nello stesso modo anche i secoli, che invece in italiano richiedono i numeri romani. I numeri romani vengono infine utilizzati in tutte e tre le lingue nel caso di nomi di personalità contenenti numeri ordinali (ad esempio, "Giovanni Paolo II").

3.4.3.5 Problemi di comprensione

Il simbolo “#” ha mantenuto la funzione che aveva nelle trascrizioni di EPIC, ossia indicare difficoltà di comprensione, dovute non solo a problemi tecnici quali la qualità dell’audio, ma anche difficoltà dovute a una più scarsa competenza in una determinata lingua, nonché alla velocità dell’oratore/interprete o al suo particolare accento. Il “#” sostituisce inoltre alcune parti di un intervento, nel caso in cui un oratore continui il proprio discorso in una lingua diversa da quella del sottocorpus in questione. Per evitare di confondere le due funzioni che lo stesso simbolo ricopre, quest’ultimo fenomeno viene spiegato nell’attributo “comments” dei metadati (3.4.4) relativi all’intervento in cui appare.

3.4.3.6 Punteggiatura

Diversamente dalle convenzioni adottate in EPIC e in EPTIC, si è deciso di mantenere la punteggiatura in fase di trascrizione, di nuovo al fine di rispecchiare sia la lingua orale che la lingua scritta, modificando tuttavia la base fornita dal sito del Parlamento europeo per trasmettere ciò che effettivamente si percepisce ascoltando l’audio. La punteggiatura mantiene le convenzioni standard proprie a ciascuna lingua: in francese, infatti, il punto interrogativo, quello esclamativo, i due punti e il punto e virgola sono preceduti e seguiti da uno spazio. La trascrizione delle virgolette, invece, costituisce un’eccezione: nonostante il francese utilizzi di norma le virgolette basse («»), per motivi di standardizzazione e per non creare problemi di *encoding* dei caratteri in fase di costruzione e consultazione del corpus, esse vengono sostituite da quelle alte utilizzate per l’italiano. Le virgolette singole (“”), solitamente utilizzate in lingua inglese, non pongono invece problemi di codifica e vengono quindi mantenute.

Le virgolette singole, così come quelle alte utilizzate in italiano e gli apostrofi, vengono tuttavia sottoposte a un processo di standardizzazione secondo il set di caratteri ANSI.

3.4.3.7 Applausi e parole inventate

Come in EPIC, l'interruzione del discorso da parte di applausi è indicata con “[applause]”.

È stato poi introdotto l'utilizzo delle parentesi graffe, per potere evidenziare altri casi particolari non previsti, sorti nel corso del lavoro di trascrizione. In un paio di occasioni, infatti, l'interprete utilizza un calco al posto della parola corretta nella lingua di arrivo: si tratta del francese “manganelle” per l'italiano “manganello” e della parola “iter” utilizzata erroneamente al posto di “processus”. La stessa convenzione vale per casi in cui l'interprete non capisce e si trova costretto a inventare una parola, riproducendo all'incirca il suono che ha sentito. In questi casi, non si è ritenuto opportuno inserire le parole corrette seguite dalla pronuncia effettiva tra slash, né tantomeno effettuare una normalizzazione sostituendo le parole errate con quelle corrette, poiché la comprensione è possibile solamente facendo riferimento all'originale e grazie alla conoscenza di una determinata lingua. Questo aspetto merita inoltre un'ulteriore riflessione. Questo tipo di errori non è probabilmente dovuto alla competenza linguistica deficitaria dell'interprete, ma ai tempi ristretti propri dell'interpretazione simultanea che costringono gli interpreti a elaborare una soluzione efficace in pochi secondi. Nel caso di EPTIC_01_2011, il fenomeno è sporadico, ma occorre tenere a mente che le dimensioni del corpus sono ancora relativamente ridotte e bisognerebbe analizzare un maggior numero di testi per avere un dato veritiero sulla sporadicità o meno del fenomeno.

Occorrerebbe inoltre effettuare uno studio più approfondito sul profilo degli interpreti per escludere con certezza la possibilità che tali errori siano commessi da interpreti con una minore competenza della lingua di arrivo.

3.4.3.8 Lettere maiuscole e appellativi

Per le convenzioni di tipo prettamente ortografico, quali lettere maiuscole e minuscole, nomi ufficiali di istituzioni e programmi, nonché sigle, si fa riferimento ai Manuali interistituzionali di convenzioni redazionali,⁶ guide di stile per la redazione di testi in seno alle istituzioni europee. Se alcuni casi non fossero presenti in tali guide di stile, si ricorre a siti ufficiali di istituzioni e organizzazioni, altri testi del Parlamento europeo o della Commissione e dizionari monolingue online.⁷ Nel caso di nomi attestati non pronunciati con la dicitura corretta, non vengono utilizzate le lettere maiuscole. Per quanto riguarda le parole “commissione” e “commissario” si è deciso di utilizzare le maiuscole e le minuscole per distinguere la Commissione europea dalle diverse commissioni parlamentari. Si scrive quindi “Commissione”/“Commissario” se si sta parlando della Commissione europea e “commissione”/“commissario” se si fa riferimento a una delle commissioni parlamentari. Infine, nel caso degli appellativi allocutivi, si è deciso di utilizzare le seguenti convenzioni: per l’italiano “signore”, “signor”, “signora”, per l’inglese “Mr”, “Mrs”, “Ms”, “madam” e “sir”, per il francese “monsieur”, “madame” e “mademoiselle”.

⁶ È possibile scaricare i manuali interistituzionali di convenzioni redazionali delle lingue di interesse al link <http://bookshop.europa.eu/it/manuale-interistituzionale-di-convenzioni-redazionali-2011-pbOA3110655/>.

⁷ Abbiamo deciso di utilizzare i seguenti dizionari monolingue: il Vocabolario Treccani (<http://www.treccani.it/vocabolario/>) per l’italiano, l’Oxford Learner’s Dictionaries (<http://www.oxfordlearnersdictionaries.com/>) per l’inglese e il CNRTL (<http://www.cnrtl.fr/definition/>) per il francese.

Si riporta qui di seguito una tabella che riassume le convenzioni di editing descritte sopra.

Aspetto del testo	Esempio	Convenzione di editing
Punteggiatura		<i>Basata su sintassi e intonazione</i>
		<i>Standardizzata in base all'encoding ANSI</i>
Parole troncate	propo	<i>propo-</i>
	pro posal	<i>proposal /pro_posal/</i>
Disfluenze	preposal	<i>proposal /preposal/</i>
	pròposal	<i>proposal /pròposal/</i>
	proposal	<i>proposal /proposal/</i>
	eventuare	<i>evitare eventuali /eventuare/</i>
Pause	piene	<i>ehm</i>
	vuote	<i>...</i>
Interventi interrotti		<i>...</i>
Numeri	numeri	<i>in cifre</i>
	326,000,000	<i>326 million / millions / milioni</i>
	326,000,000,000	<i>326 billion / milliards / miliardi</i>
	34%	<i>34 percent / pour cent / per cento</i>
	anni	<i>in cifre</i>

	frazioni	<i>in lettere</i>
	numeri ordinali	<i>numero+st / nd / rd / th er / ère / ème o / a o numeri romani</i>
Problemi di comprensione o di ascolto		#
Cambio di lingua		#
Calchi/Parole inventate	manganelle	{ <i>manganelle</i> }
Applausi		[<i>applause</i>]
Maiuscole		<i>Si vedano Manuale interistituzionale di convenzioni redazionali, siti ufficiali o dizionari monolingue</i>
	Commissione europea	<i>Commission / Commission / Commissione</i>
	commissione parlamentare	<i>commission / commission / commissione</i>
	membro della Commissione europea	<i>Commissioner / Commissaire / Commissario</i>
	membro di una commissione parlamentare	<i>commissioner / commissaire / commissario</i>
Sigle e nomi ufficiali		<i>Si vedano Manuale interistituzionale di convenzioni redazionali o siti ufficiali</i>
Appellativi allocutivi		<i>Mr / Mrs / Ms / sir / madam monsieur / madame /</i>

		<i>mademoiselle</i> <i>signor / signore / signora</i>
--	--	--

Tabella 3.2 - Convenzioni di editing

Nel prossimo paragrafo si andranno ad analizzare i metadati relativi ai testi contenuti in EPTIC_01_2011.

3.4.4 Metadati

Una discussione a parte meritano i metadati, ossia una specie di intestazione che precede ogni testo di ciascun sottocorpus e che può essere utile in fase di ricerca per filtrare i risultati da ottenere. I metadati sono composti da una serie di attributi e dai relativi valori. Gli attributi sono separati da trattini bassi, mentre i valori sono tra virgolette e le parole da cui sono composti sono separate da trattini. I valori, inoltre, non devono contenere simboli (il gruppo politico "S&D" viene riportato nei metadati come "SD"). Rispetto ai metadati utilizzati in EPIC e in EPTIC, quelli utilizzati per EPTIC_01_2011 hanno subito alcune importanti modifiche volte a migliorarne la fruibilità da parte degli utenti e a renderli il più completi possibile. I metadati sono in lingua inglese perché EPTIC_01_2011 è un corpus internazionale. Si riporta qui di seguito un esempio dell'ultima versione dei metadati.

```
#text_id="001"_date="17-01-11-
a"_length="short"_lengthw="232"_duration="short"_duration
s="76"_speed="high"_speedwm="182.7"_delivery="read"_topic
="Politics"_topicspec="Order-of-business"_type="st-in-
en"_comments="NA"

speaker_name="Tavares-
Rui"_gender="M"_country="Portugal"_native="n"_politfunc="
MEP"_politgroup="GUE-NGL"

st_length="NA"_lengthw="NA"_duration="NA"_durations="NA"_
speed="NA"_speedwm="NA"_delivery="NA"
```

`interpreter_gender="NA" _native="NA"`

Si andranno ora ad analizzare i metadati più nel dettaglio per capire quali modifiche sono state apportate e quali informazioni essi ci forniscono.

3.4.4.1 Metadati relativi al testo

L'attributo "text_id" è il numero identificativo del testo. Si è deciso di non numerare progressivamente i testi di tutti i source, ma di ricominciare la numerazione dall'inizio per ogni lingua di partenza. Quindi si avranno testi che partono da 001 sia per il sottocorpus st-in-en, sia per quello st-in-fr che per il st-in-it. Lo stesso id viene poi ripreso per le relative interpretazioni e per i testi source (da questo punto in poi "st") delle traduzioni e i relativi testi target (da questo punto in poi "tt").

Nell'attributo "date", si inserisce la data dell'intervento corredata dall'indicazione del momento della giornata in cui l'oratore è intervenuto ("m" per morning, se l'intervento si è svolto prima delle 13, e "a" per afternoon, dalle 13 in poi).

Gli attributi "lengthw", "durations" e "speedwm" contengono i valori numerici relativi al numero di parole del testo in questione, alla durata dell'intervento in secondi e alla velocità dell'oratore, ovvero il numero di parole che pronuncia al minuto. I valori di tali attributi determinano i valori degli attributi "length", "duration" e "speed", secondo intervalli preesistenti, messi a punto per l'implementazione di EPIC e utilizzati successivamente per la prima versione di EPTIC. La Tabella 3.5 (Sandrelli e Bendazzoli, 2005:4) riportata qui di seguito riassume i parametri di riferimento.

duration	short < 2 minutes medium 2-6 minutes long > 6 minutes
text length	short < 300 words

	medium 301-1000 words long > 1000 words
speed of delivery	low < 130 words per minute (w/m) medium 131-160 w/m high > 160

Tabella 3.5 - Parametri calcolo durata, lunghezza e velocità

Si è deciso di adottare questi parametri, che differiscono da altri utilizzati per gli intervalli che determinano i valori dell'attributo "speed", poiché utilizzando gli altri, in cui il valore massimo è fissato a 120 (Sandrelli et al., 2010:175), la velocità di esposizione degli interventi risulterebbe "high" nella maggior parte dei casi, rendendo dunque l'attributo "speed" superfluo e non rappresentativo. Per quanto concerne il conteggio del numero di parole necessario a riempire gli attributi "length" e "lengthw", è stato necessario elaborare l'espressione regolare riportata di seguito da utilizzare su Notepad++ con il plugin TextFX Quick, in modo da calcolare con la maggiore precisione possibile il numero di parole di ciascun testo.

`[a-zA-Z0-9\u00-\uff]+([-,.:][a-zA-Z0-9\u00-\uff]+)*`

Non è stato tuttavia possibile evitare alcune imprecisioni: per esempio, le parole unite da trattino vengono conteggiate come un'unica parola. Il problema non sussiste nel caso dell'italiano e dell'inglese, dove le parole unite da trattino sono effettivamente da considerare un'unica parola. In francese, invece, le parole con trattino a volte sono considerate un'unica parola, mentre a volte restano due parole separate, come nel caso delle domande con inversione. Non si è comunque ritenuto necessario trovare un'altra soluzione, perché i casi di questo tipo corrispondono a una parte non sostanziale del corpus. Infatti, prendendo in considerazione il sottocorpus st-in-fr (24.686 parole), le parole con trattino sono 161 (0,65%) e di queste solo 84 sono da considerarsi due parole

(0,34%). In caso di lassi di tempo separati da trattino, invece, si è deciso di separare manualmente gli elementi inserendo uno spazio sia prima che dopo il trattino (es. “2009 – 2014”). Dal conteggio parole sono esclusi gli slash con il loro contenuto e la punteggiatura, mentre vengono conteggiate le pause piene e le parole troncate. Quando i valori numerici contenuti negli attributi di cui sopra risultano essere numeri non finiti, si riporta solo il primo decimale arrotondando per eccesso nel caso in cui il secondo decimale sia ≥ 5 o per difetto in caso contrario. Non si riportano mai numeri seguiti da “.0”, né in caso di numeri finiti né in caso di decimale da arrotondare per difetto (quindi, nel caso di 132.02, si troverà “132”). Nei sottocorpora di traduzione, gli attributi “duration”, “durations”, “speed” e “speedwm” vengono riempiti con “NA”, in quanto non applicabili alla modalità scritta.

L’attributo “delivery” ci informa sul modo in cui si esprime l’oratore, ossia se parla senza una traccia scritta (“impromptu”), se segue una traccia scritta ma non legge oppure se alterna le due modalità (“mixed”) o se legge (“read”). Nel caso delle interpretazioni, si riempie l’attributo con il valore “interpreted”, dato che nessuno dei tre valori precedenti può essere applicato al caso specifico. Nei sottocorpora di traduzione, questo attributo è riempito con “NA”.

Gli attributi “topic” e “topicspec” forniscono informazioni riguardo all’argomento di cui si parla nel relativo testo. Più nello specifico, nel “topicspec” si riprende esattamente il titolo della sezione in cui si trova il testo sul sito del Parlamento europeo o nei verbatim report, mentre il “topic” viene fatto derivare dal titolo di cui sopra o, in caso di titoli troppo generici (per esempio, “Explanations of vote”), da ciò di cui si parla nell’intervento. I valori

dell'attributo "topic" sono stati ripresi dalla lista elaborata per EPIC (Sandrelli et al., 2010:175).

Agriculture-and-Fisheries
Economics-and-Finance
Employment
Environment
Health
Justice
Politics
Procedure-and-Formalities
Society-and-Culture
Science-and-Technology
Transport

L'attributo "type" contiene informazioni relative al sottocorpus in cui si trova il testo. Al primo posto si trovano le sigle "st" o "tt", a indicare se si tratta di un st o di un tt. Poi si trovano le diciture "in" o "tr", per differenziare tra interpretazioni e traduzioni. Infine, compare una tra le sigle "en", "fr" e "it", a indicare la lingua dei testi contenuti nel sottocorpus in questione.

L'attributo "comments" è riservato invece a commenti di vario tipo, per esempio riguardanti problemi tecnici riscontrati durante la trascrizione, spiegazione di simboli non convenzionali, oppure convenzionali ma utilizzati con funzione diversa da quella abituale, oppure segnalazioni di errori nelle traduzioni. Nel caso non ci siano precisazioni da fare sul testo in questione, l'attributo conterrà il valore "NA".

3.4.4.2 Metadati relativi all'oratore

Ci sono poi alcuni attributi riguardanti l'oratore del testo originale, ripresi anche nella relativa interpretazione, nonché nel st della traduzione e nella traduzione stessa. L'attributo "name" deve essere riempito nel seguente formato: "Nome-Cognome". L'attributo "gender" può essere riempito con "M" o "F". Per riempire gli attributi "country", "native" (valori: "y/n") e "politfunc" si

fa riferimento alla scheda personale dell'oratore, nel caso in cui questo sia un MEP, o ad altri documenti ufficiali delle istituzioni europee. Se un MEP ricopre anche una carica più importante, come "MEP Chairman of the session" o "Vice-President of the European Parliament",⁸ l'attributo viene riempito con la carica più importante. I membri della Commissione europea vengono segnalati tra parentesi sul sito del Parlamento prima del testo dell'intervento. I valori necessari al riempimento di tale attributo sono stati ripresi da EPIC (Sandrelli et al., 2010:174), come riportato qui di seguito.

MEP
MEP-Chairman-of-the-session
President-of-the-European-Parliament
Vice-President-of-the-European-Parliament
European-Commission
European-Council
Guest

Per l'attributo "politgroup", non è stato invece possibile riprendere i valori validi per EPIC, in quanto i gruppi politici facenti parte del Parlamento europeo sono cambiati con il passare degli anni. Si è quindi inserito il gruppo politico riportato tra parentesi prima dei vari interventi sul sito, controllando comunque sempre sulla scheda personale di ogni parlamentare. Per i membri della Commissione europea e gli eventuali "Guest" non è possibile definire un gruppo politico, quindi l'attributo è riempito con "NA".

3.4.4.3 Metadati relativi al testo source

La successiva stringa di metadati è stata aggiunta per EPTIC_01_2011, al fine di rendere ancora più complete le informazioni fornite e affinare maggiormente le ricerche. Si tratta

⁸ È possibile trovare la lista dei 14 vicepresidenti del Parlamento europeo per la legislatura 2009-2014 al seguente link
<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+IM-PRESS+20090714IPR58147+0+DOC+XML+V0//IT>

infatti di dati relativi al st corrispondente, che potrebbero essere utili al fine di comprendere alcuni fenomeni riscontrati nel tt in analisi. Ad esempio, alcuni errori da parte dell'interprete potrebbero essere dovuti alla velocità con cui l'oratore parla in aula. Nei st, gli attributi di questa stringa vengono completamente riempiti con "NA", essendo i dati relativi al st già presenti nella prima stringa, ossia quella del testo in questione.

3.4.4.4 Metadati relativi all'interprete

Troviamo infine una stringa di metadati relativi all'interprete, "gender" ("M/F") e "native" ("y/n"). Il dato relativo alla *nativeness* è deducibile solamente in fase di ascolto della performance dell'interprete, in quanto il Parlamento non fornisce alcuna informazione in tale direzione. In caso di dubbi sulla *nativeness* o meno dell'interprete l'attributo "native" può essere riempito con il valore "unknown". Nel caso in cui invece l'interprete cambi durante uno stesso discorso, passando tra l'altro da uomo a donna o viceversa, questa stringa di metadati verrà ripresa e modificata a seconda del caso prima di ogni porzione di testo in cui l'interprete è diverso da quello della porzione di testo precedente. Inutile dire che questi attributi si applicano solamente ai sottocorpora "tt-in" delle tre lingue, essendo gli unici in cui interviene l'interprete. Negli altri sottocorpora tali attributi vengono quindi riempiti con "NA".

Di seguito è riportata una tabella volta a riassumere gli attributi dei metadati e i relativi valori.

	Attributo	Valore
Dati testo	text_id	<i>numero che identifica il testo</i>
	date	<i>gg-mm-aa-m/a</i>

	length	<i>short/medium/long</i>
	lengthw	<i>numero di parole</i>
	duration	<i>short/medium/long/NA</i>
	durations	<i>durata dell'intervento in secondi/NA</i>
	speed	<i>low/medium/high/NA</i>
	speedwm	<i>numero di parole al minuto/NA</i>
	delivery	<i>read/mixed/impromptu/interpreted/NA</i>
	topic	<i>Agriculture-and-Fisheries</i> <i>Economics-and-Finance</i> <i>Employment</i> <i>Environment</i> <i>Health</i> <i>Justice</i> <i>Politics</i> <i>Procedure-and-Formalities</i> <i>Society-and-Culture</i> <i>Science-and-Technology</i> <i>Transport</i>
	topicspec	<i>titolo della sezione sul sito</i>
	type	<i>st/tt-in/tr-sigla lingua</i>
	comments	<i>commento sul testo/NA</i>
Dati oratore	speaker_name	<i>Cognome-Nome</i>
	gender	<i>M/F</i>
	country	<i>nome del paese</i>
	native	<i>y/n</i>
	politfunc	<i>MEP</i>

		<i>MEP-Chairman-of-the-session</i> <i>President-of-the-European-Parliament</i> <i>Vice-President-of-the-European-Parliament</i> <i>European-Commission</i> <i>European-Council</i> <i>Guest</i>
	politgroup	<i>abbreviazione del gruppo politico come indicato sul sito/NA</i>
Dati testo source	st_length	<i>short/medium/long</i>
	lengthw	<i>numero di parole</i>
	duration	<i>short/medium/long/NA</i>
	durations	<i>durata dell'intervento in secondi/NA</i>
	speed	<i>low/medium/high/NA</i>
	speedwm	<i>numero di parole al minuto/NA</i>
	delivery	<i>read/mixed/impromptu/interpreted/NA</i>
Dati interprete	interpreter_gender	<i>M/F/NA</i>
	native	<i>y/n/NA</i>

Tabella 3.6 - Attributi e valori dei metadati

Al termine del processo di preparazione del corpus EPTIC_01_2011, i testi sono stati taggati e lemmatizzati utilizzando TreeTagger⁹ e indicizzati con Corpus WorkBench.¹⁰

Nel prossimo capitolo verrà descritta la metodologia di ricerca e di analisi delle collocazioni in EPTIC_01_2011, seguita dai risultati della ricerca stessa (si veda capitolo 5) e da una discussione di tali risultati (si veda capitolo 6).

⁹ <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

¹⁰ <http://cwb.sourceforge.net/>

4 Metodo

4.1 Introduzione al capitolo

Il presente capitolo ha come scopo la presentazione della metodologia di ricerca applicata allo studio delle collocazioni in EPTIC_01_2011, che ha come obiettivo l'identificazione di differenze e analogie nell'utilizzo di pattern fraseologici da parte di interpreti e traduttori. Nella prima parte del capitolo verrà descritta la fase di estrazione e classificazione dei bigrammi, mentre le restanti sezioni saranno focalizzate in modo più dettagliato sul metodo impiegato per le due analisi effettuate, ossia quella quantitativa e quella qualitativa.

4.2 Estrazione e classificazione dei bigrammi

Per l'analisi si è deciso di utilizzare il metodo sviluppato e descritto da Durrant e Schmitt (Durrant e Schmitt, 2009), poi ripreso da Ferraresi e Miličević in una prima esplorazione di EPTIC (Ferraresi e Miličević, submitted). Il metodo, che si basa su una definizione del concetto di collocazione/fraseologia ispirata alla tradizione della scuola *neo-firthiana* (Sinclair, 1991; Stubbs, 1995), ovvero basata sul principio della frequenza, è stato scelto per replicare un'analisi che si è rivelata vincente sotto diversi aspetti. Innanzitutto, permette una precisa classificazione delle collocazioni al fine di effettuare un'analisi quantitativa e qualitativa tra sottocorpora; inoltre, grazie a tale metodo è possibile lavorare anche con corpora di dimensioni relativamente piccole, in quanto i dati raccolti nei corpora analizzati vengono confrontati con le frequenze per i medesimi fenomeni osservate in corpora di riferimento, il che permette di superare il problema della "data sparseness" riscontrato da studi precedenti (si veda 2.2.3). Infine, trattandosi di un'analisi basata su singoli testi e non su interi

sottocorpora, permette uno studio più approfondito sui diversi fenomeni incontrati evitando generalizzazioni ritenute valide indiscriminatamente per tutti i testi. Infatti, essendo i testi per la maggior parte prodotti da persone diverse, le strategie di resa possono risultare varie e necessitano di un esame specifico e individuale. Si è inoltre deciso di riprendere tale modello di analisi in modo da poter eventualmente confrontare i dati ottenuti grazie al presente studio con quelli di studi precedenti.

Come nel caso dello studio di Ferraresi e Miličević (Ferraresi e Miličević, submitted), anche quella presentata in questo lavoro è un'analisi intermodale, il cui focus consiste nel confronto tra il prodotto dei due tipi di mediazione, ossia il tt interpretato e quello tradotto di uno stesso st. Diversamente dallo studio effettuato da Ferraresi e Miličević (Ferraresi e Miličević, submitted), invece, la presente analisi vuole rendere conto non solo dei fenomeni presenti nei testi in lingua italiana, ma anche di quelli riscontrati nella lingua straniera oggetto di studio, ossia l'inglese.

Anche i pattern sintattici su cui si è scelto di focalizzarsi sono stati in linea di massima ripresi dai due paper menzionati sopra. Infatti, sono stati estratti bigrammi contigui formati da aggettivo+nome o da nome+aggettivo. Grazie al tagging, è stato possibile evitare l'estrazione di bigrammi contenenti, ad esempio, nomi propri, numeri, pronomi, possessivi e dimostrativi. Naturalmente, essendo le procedure di tagging e di estrazione completamente automatizzate e non ritenendo fattibile una verifica manuale dell'output di tali procedure, inevitabilmente risultano alcuni errori di estrazione: si trovano infatti casi di bigrammi quali "signor Haerynck" (nome+nome proprio, che tuttavia risultano rari nei diversi sottocorpora, e quindi non vanno a inficiare i risultati

globali della ricerca. Sempre per quanto riguarda il pattern, la scelta di questa specifica struttura rispetto ad altre che tipicamente formano collocazioni, come nome+verbo o avverbio+aggettivo, è stata dettata dall'alto livello di produttività collocazionale tipica del pattern stesso e, ancora una volta, dalla volontà di replicare gli studi dei due paper menzionati, con la conseguente possibilità di confrontare i risultati. Tuttavia, ciò non esclude la possibilità di analizzare altri patterns in futuro, al fine di mettere in luce ulteriori fenomeni e poter confutare o avvalorare le ipotesi proposte nel presente lavoro.

La frequenza dei bigrammi estratti è stata controllata nei corpora di riferimento, ovvero due subset casuali del corpus ukWaC e due subset casuali del corpus itWaC (Baroni et al., 2009), contenenti circa 90 milioni di parole ciascuno. A seguito di tale confronto, ai bigrammi sono poi stati assegnati dei valori calcolati in base a tre misure di associazione lessicale: Mutual Information (da questo punto in poi MI), *t*-score e frequenza congiunta. Il parametro della frequenza consente di individuare i bigrammi che risultano non frequenti o non attestati nel corpus di riferimento, e che quindi non possono essere considerati collocazioni (Ferraresi e Miličević, submitted:11). Le misure di associazione MI e *t*-score, invece, permettono di distinguere due tipi diversi di possibili collocazioni: mentre il *t*-score denota collocazioni molto frequenti, la MI mette in luce combinazioni meno frequenti e i cui componenti sono raramente riscontrabili singolarmente, al di fuori della collocazione stessa (Durrant e Schmitt, 2009; Ferraresi e Miličević, submitted).

Una volta assegnati i valori di frequenza ai bigrammi estratti, sono stati presi in considerazione solamente i bigrammi che potevano soddisfare le seguenti condizioni: per la prima

analisi relativa alla frequenza e volta a determinare l'assenza probabile di un legame collocazionale, i bigrammi dovevano avere una frequenza < 2 (ossia comparire nel corpus di riferimento 0 o 1 volta); per determinare invece quali fossero i probabili candidati a collocazione in termini di t -score e MI, i parametri utilizzati sono stati rispettivamente: a) t -score ≥ 10 e frequenza congiunta ≥ 5 ; b) MI ≥ 7 e frequenza congiunta ≥ 5 .

Rispetto alla metodologia di Durrant e Schmitt (2009), però, emerge una differenza riguardante tali soglie, per cui è necessaria una precisazione. La soglia fissata per l'analisi relativa ai bigrammi non frequenti/non attestati era infatti 5. Tuttavia, come in Ferraresi e Miličević (submitted), tale soglia è stata considerata molto conservativa, in quanto secondo la scuola *neo-firthiana* una collocazione può essere considerata tale se la sua frequenza minima riscontrata nel corpus di riferimento è 2 (Jones e Sinclair, 1974/1996).

Per i testi contenenti i bigrammi che soddisfacevano le condizioni descritte sopra si è deciso di effettuare un'analisi non solo quantitativa, ma anche qualitativa. I metodi seguiti per ciascuna di queste analisi sono descritti nelle sezioni che seguono.

4.3 Analisi quantitativa

Per l'analisi di tipo quantitativo, una volta estratti i bigrammi si è proceduto al calcolo delle percentuali dei bigrammi (token) non frequenti/non attestati, ad alto MI e ad alto t -score per ogni testo di ciascun sottocorpus (dove per token si intendono le singole occorrenze dei bigrammi estratti per un determinato testo). Tali percentuali sono state calcolate rispetto al numero totale dei bigrammi estratti per un determinato testo.

La distribuzione delle percentuali (corrispondenti ai singoli testi) relative ai sottocorpora è poi stata sottoposta ad analisi statistica. L'obiettivo di tale analisi è il calcolo della significatività, per cui un risultato è significativo se il valore $p < 0.05$. Innanzitutto è stato necessario accertare (Shapiro test) se la distribuzione in ognuno dei sottocorpora fosse normale. Nel caso in cui questa risultasse non normale si è fatto ricorso ad un test non parametrico (test di Wilcoxon). Quest'ultimo si basa sulla differenza tra le mediane. Nel caso in cui tutti i sottocorpora presi in analisi risultassero avere distribuzione normale, è stato invece utilizzato il t-test, che guarda la differenza tra le medie.

La metodologia si applica a entrambe le lingue di lavoro (inglese e italiano) e a tutti e tre i parametri (bigrammi non frequenti/non attestati, MI e t -score).

Nel capitolo 5 sono riportati i risultati dell'analisi quantitativa, corredati da tabelle e grafici contenenti i dati ottenuti. Si passerà ora alla descrizione della metodologia utilizzata per procedere all'analisi qualitativa dei bigrammi estratti.

4.4 Analisi qualitativa

Per procedere all'analisi qualitativa dei bigrammi estratti, è stata fatta la differenza tra le percentuali dei token dei bigrammi non frequenti/non attestati, ad alto MI e ad alto t -score ottenute nell'analisi quantitativa dei testi tradotti e dei rispettivi testi interpretati (e viceversa). Come accennato in precedenza, infatti, si tratta di un'analisi intermodale, ossia atta a studiare i due tipi di mediazione. Come per l'analisi quantitativa, la metodologia si applica a ciascuna delle due lingue target per tutti e tre i parametri.

Una volta calcolate tali differenze, i testi risultano classificati secondo un gradiente di collocazionalità: nella fascia alta si trovano i testi che sono più collocazionali nella modalità di traduzione rispetto a quella di interpretazione, mentre in quella bassa si trovano i testi più collocazionali nella modalità di interpretazione rispetto a quella di traduzione. Questo per quanto concerne i parametri MI e *t*-score. Per quanto riguarda l'analisi dei bigrammi non frequenti/non attestati, che come già detto non indica collocazionalità ma l'esatto contrario, si trovano in alto i testi contenenti un numero di bigrammi non attestati o non frequenti maggiore per la traduzione che per l'interpretazione e in basso i testi con un numero di bigrammi non attestati o non frequenti maggiore per l'interpretazione.

A questo punto sono stati considerati i testi la cui differenza tra interpretato e tradotto fosse maggiore: è stato analizzato un testo preso dalla fascia alta e uno preso dalla fascia bassa per ogni parametro (bigrammi non frequenti/non attestati, alto *t*-score e alto MI) e per le due lingue target. La scelta dei testi non è stata arbitraria ma mirata, prendendo in considerazione coppie di testi in cui i bigrammi estratti fossero interessanti e ci permettessero un'analisi qualitativa approfondita.

L'analisi qualitativa mira dunque a far luce sulle diverse strategie di resa adottate da interpreti e traduttori, andando a evidenziare le differenze e le analogie riscontrate nell'esame concreto dei testi e dei bigrammi selezionati come possibili collocazioni. In fase di analisi, i corpora di riferimento e i *st* vengono utilizzati come supporto alle ipotesi avanzate nella spiegazione dei diversi fenomeni riscontrati.

Nel prossimo capitolo verranno presentati i risultati ottenuti sia per l'analisi quantitativa che per l'analisi qualitativa, per poi

passare, nel capitolo successivo, a una discussione più generale degli esiti ottenuti grazie al presente lavoro, nonché dei possibili sviluppi futuri.

5 Risultati

5.1 Introduzione al capitolo

Il presente capitolo ha come obiettivo la presentazione dell'analisi relativa ai bigrammi estratti. Innanzitutto verranno presentati i risultati dell'analisi quantitativa sui bigrammi caratterizzati da bassa frequenza e da alti valori di t -score e di MI sia per l'inglese che per l'italiano. La seconda parte del capitolo sarà incentrata sui risultati dell'analisi qualitativa effettuata sui testi.

5.2 Analisi quantitativa

Una volta calcolate le percentuali dei token dei bigrammi non frequenti/non attestati, ad alto t -score e ad alto MI per ogni testo di ciascun sottocorpus, si è proceduto a effettuare una verifica della normalità delle distribuzioni (tramite Shapiro test), al fine di decidere il test statistico appropriato (test di Wilcoxon o t -test) per calcolare la significatività delle differenze tra sottocorpora.

5.2.1 Bigrammi non frequenti/non attestati

5.2.1.1 Inglese

Lo Shapiro test applicato ai sottocorpora tt-in-en e tt-tr-en ha rivelato una distribuzione non normale per entrambi i sottocorpora. È stato quindi eseguito il test di Wilcoxon per il calcolo della significatività. Il risultato ottenuto non è tuttavia significativo. La Figura 5.1 mostra le percentuali dei bigrammi non frequenti/non attestati nei due sottocorpora, mentre la Tabella 5.1 contiene le mediane (M) e il risultato del test di Wilcoxon. Come già ricordato sopra, quest'ultimo, per essere considerato significativo, deve avere un valore di $p < 0.05$.

Intermodale	
tt-in-en	tt-tr-en
M=22.22	M=18.18
W=2545.5, $p=0.31$	

Tabella 5.1 - Significatività bigrammi non frequenti/non attestati EN

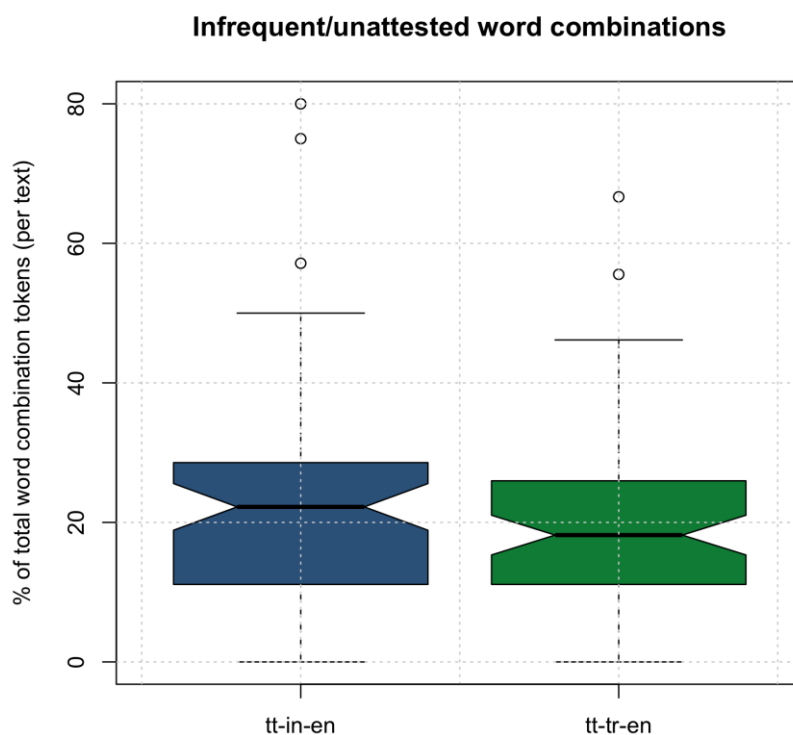


Figura 5.1 - Percentuali bigrammi non frequenti/non attestati EN

5.2.1.2 Italiano

Anche per l'italiano la distribuzione risulta non normale per entrambi i sottocorpora (tt-in-it e tt-tr-it). Di nuovo è stato applicato il test di Wilcoxon, che tuttavia ha prodotto anche in questo caso risultati non significativi. I dati relativi a tale analisi sono riportati in Tabella 5.2 e Figura 5.2.

Intermodale	
tt-in-it	tt-tr-it

M=15.27	M=14.29
W=1995, $p=0.72$	

Tabella 5.2 - Significatività bigrammi non frequenti/non attestati IT

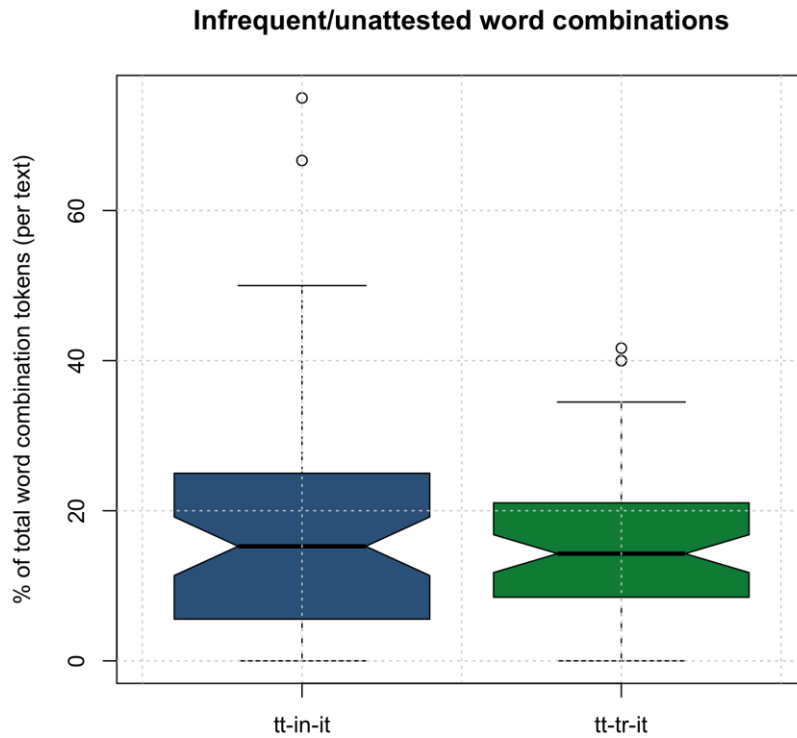


Figura 5.2 - Percentuali bigrammi non frequenti/non attestati IT

5.2.2 Alto *t*-score

5.2.2.1 Inglese

Anche nel caso dell'alto *t*-score per i sottocorpora mediati in lingua inglese la distribuzione non è normale. Ancora una volta è stato applicato il test di Wilcoxon e di nuovo il risultato non è significativo in quanto il valore p è > 0.05 (si vedano Tabella 5.3 e Figura 5.3).

Intermodale	
tt-in-en	tt-tr-en
M=27.92	M=27.27

W=2407, $p=0.68$

Tabella 5.3 - Significatività alto t-score EN

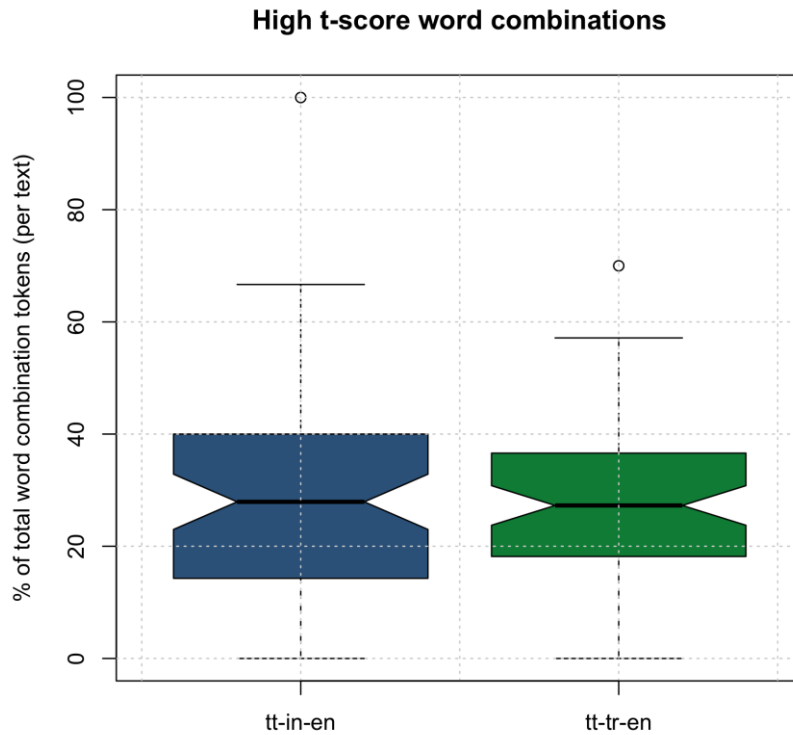


Figura 5.3 - Percentuali bigrammi alto t-score EN

5.2.2.2 Italiano

Nel caso dei sottocorpora mediati in lingua italiana la distribuzione risulta invece normale, quindi è stata considerata la media ed è stato applicato il t-test per il calcolo della significatività. Tuttavia anche questa volta il risultato ottenuto non è significativo (si vedano Tabella 5.4 e Figura 5.4).

Intermodale	
tt-in-it	tt-tr-it
M=43.8	M=43.21
$t=0.18, p=0.86$	

Tabella 5.4 - Significatività alto t-score IT

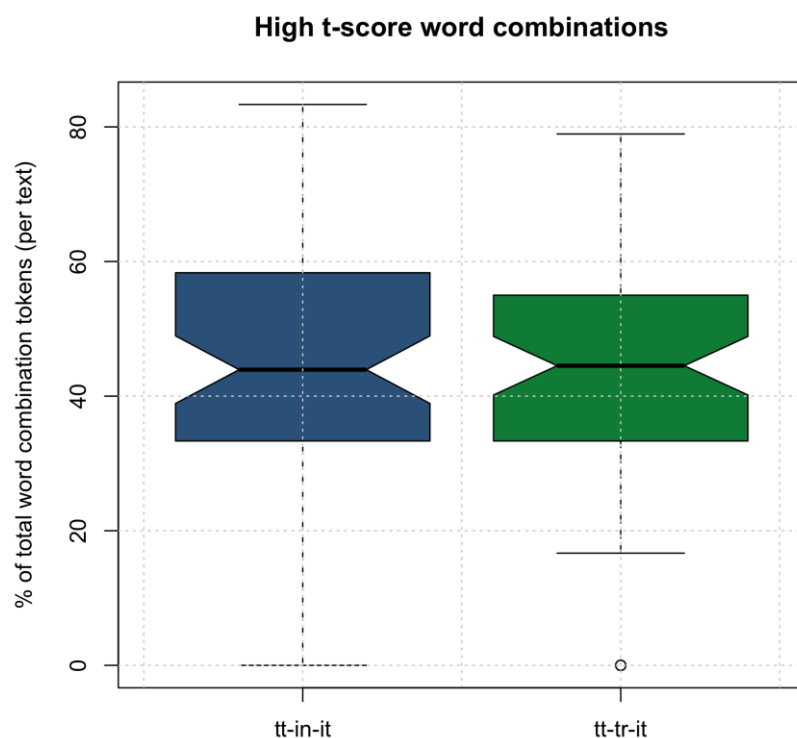


Figura 5.4 - Percentuali bigrammi alto t-score IT

5.2.3 Alto MI

5.2.3.1 Inglese

Lo Shapiro test applicato ai sottocorpora tt-in-en e tt-tr-en per tale parametro ha rivelato che anche in questo caso la distribuzione non è normale. È stato quindi effettuato il test di Wilcoxon che tuttavia ha nuovamente restituito un risultato non significativo, come si può vedere dalla Tabella 5.5 (per le percentuali dei bigrammi si veda Figura 5.5).

Intermodale	
tt-in-en	tt-tr-en
M=12.92	M=14.29
W=2091, $p=0.34$	

Tabella 5.5 - Significatività alto MI EN

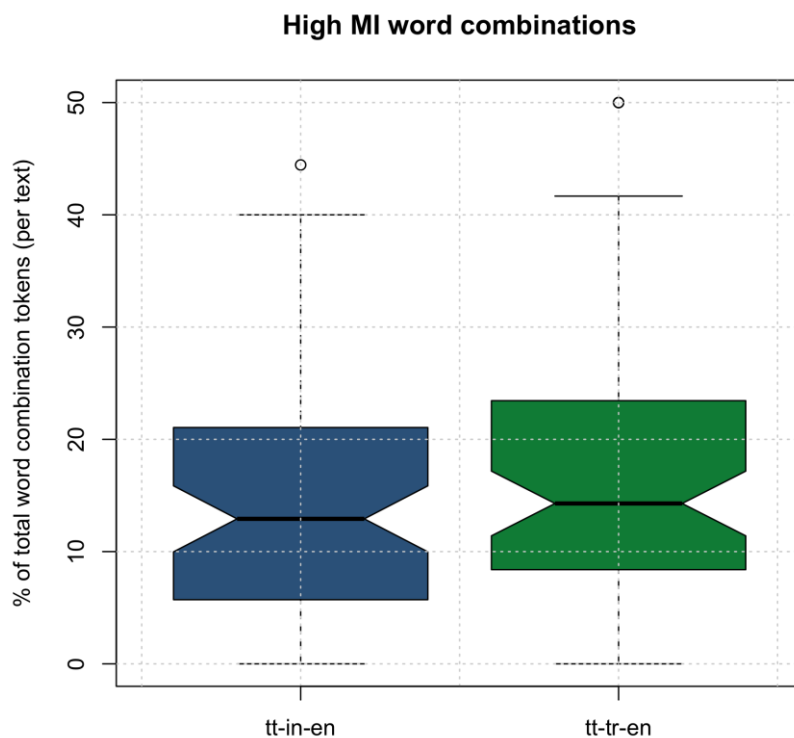


Figura 5.5 - Percentuali bigrammi alto MI EN

5.2.3.2 Italiano

In quest'ultima analisi, la distribuzione risulta normale per entrambi i sottocorpora (tt-in-it e tt-tr-it). Sono quindi state considerate le medie ed è stato applicato il t-test per il calcolo della significatività. Anche in questo caso, però, il risultato ottenuto non è significativo (si veda Tabella 5.6 e, per le percentuali dei bigrammi, Figura 5.6).

Intermodale	
tt-in-it	tt-tr-it
M=43.11	M=41.63
t=0.52, p=0.6	

Tabella 5.6 - Significatività alto MI IT

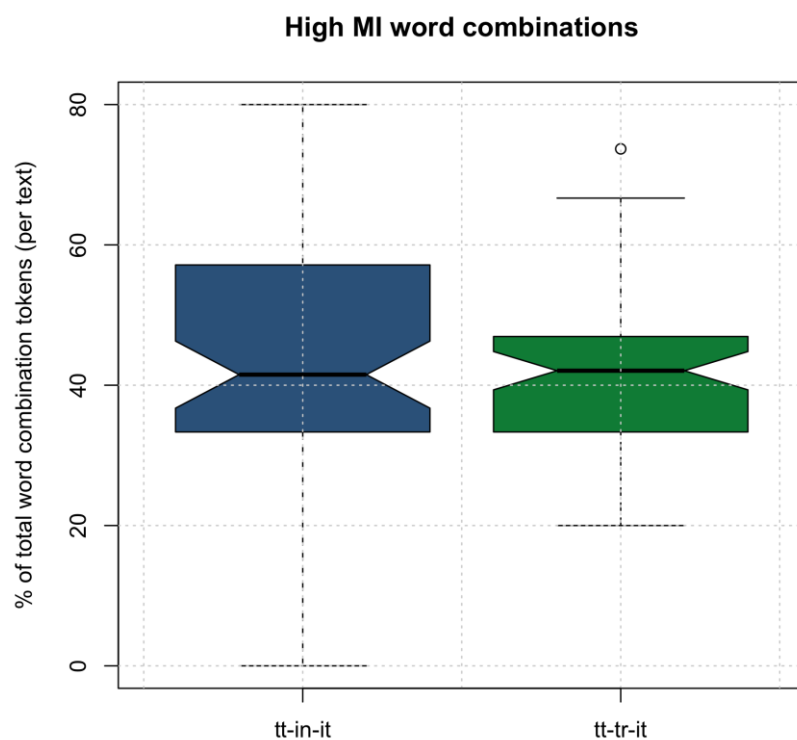


Figura 5.6 - Percentuali bigrammi alto MI IT

Riassumendo, nessuna delle analisi quantitative effettuate sui sottocorpora di lingua mediata (inglese e italiano) ha prodotto risultati significativi. Le possibili motivazioni dei risultati appena presentati e le loro implicazioni alla luce degli studi precedenti verranno presentate nel capitolo 6. Nella prossima sezione si procederà invece a un'analisi qualitativa più approfondita dei singoli testi in cui è stata osservata una divergenza significativa dal punto di vista fraseologico nell'output di traduttori e interpreti.

5.3 Analisi qualitativa

Come spiegato in 4.4, una volta terminata l'analisi quantitativa si è proceduto all'analisi qualitativa dei bigrammi estratti per determinati testi. I testi sono stati scelti in base alla maggiore o minore collocazionalità dell'interpretato rispetto al tradotto (e viceversa) per i parametri alto *t*-score e alto MI e in base al maggior o minor numero di bigrammi non frequenti/non attestati.

Le tabelle 5.7, 5.8 e 5.9 riportate di seguito riassumono i testi e i bigrammi che verranno analizzati in dettaglio nella parte restante del capitolo.

Lingua	Testo	Bigrammi tt-in	Bigrammi tt-tr	Status
Inglese	048	mere politicking Italian response communal plant	underused areas municipal separate- collection Italian response European veto dry separation	Più bigrammi non frequenti/non attestati nel tradotto
	034	favourable context (x2) young businessmen quality tourist one-man companies longer regard consolidated framework broad palette	same enterprises repayable aid other cross-cutting electronic observatory consolidated framework	Più bigrammi non frequenti/non attestati nell'interpretato
Italiano	017	legname illegale	legname tropicale (x2) tropicale aumenti possibile accordi legname illegale gestione accorta base vitale	Più bigrammi non frequenti/non attestati nel tradotto
	006	resistenze corporativistiche regolamentazione migliore piccole corporazioni lenta lento condizioni separate buoni steccati	signor Haerynck resistenze corporativistiche normativa migliore diverse correlazioni buone recinzioni	Più bigrammi non frequenti/non attestati nell'interpretato

Tabella 5.7 - Testi e bigrammi non frequenti/non attestati per analisi qualitativa

Lingua	Testo	Bigrammi tt-in	Bigrammi tt-tr	Status
Inglese	041	other options few years	European legislation (x2) renewable energy public health few years environmental impact	Tradotto maggiormente collocazionale
	022	natural disasters (x2) same time international level humanitarian aid human rights	natural disasters (x2) humanitarian aid dramatic increase	Interpretato maggiormente collocazionale
Italiano	061	Unione europea strada giusta segnale importante grande attenzione diritti umani Consiglio nazionale caso concreto attuale fase	Unione europea (x2) sviluppo tecnologico strada giusta sistema giudiziario segnale importante prossime elezioni Parlamento europeo fondamentalismo islamico diritti umani Consiglio nazionale caso specifico buon punto	Tradotto maggiormente collocazionale
	034	maggior parte (x2) genitori adottivi (x2) stragrande maggioranza giusto equilibrio certi casi	buona fede	Interpretato maggiormente collocazionale

Tabella 5.8 - Testi e bigrammi ad alto t-score per analisi qualitativa

Lingua	Testo	Bigrammi tt-in	Bigrammi tt-tr	Status
Inglese	023	long periods economic crisis	small-scale farming organic farming long periods geographical regions economic crisis	Tradotto maggiormente collocazionale
	004	various forms successive phases red tape legislative proposals exigent circumstances excessive bureaucracy dangerous substances	sincere thanks legislative proposals hazardous substances exact opposite bureaucratic procedures	Interpretato maggiormente collocazionale
Italiano	007	Unione europea oneri burocratici medie imprese libera circolazione crisi economica	Unione europea sostanze pericolose settore edile risultati concreti oneri burocratici medie imprese libera circolazione imprese produttrici importanza vitale crisi economica cittadini europei	Tradotto maggiormente collocazionale
	026	Unione europea (x3) prodotti ittici (x2) situazione attuale scambi commerciali industria ittica	Unione europea (x4) grandi quantità	Interpretato maggiormente collocazionale

Tabella 5.9 - Testi e bigrammi ad alto MI per analisi qualitativa

Si passerà ora all'analisi dettagliata dei bigrammi non frequenti/non attestati o candidati a collocazione divisi per parametro, testo e lingua.

5.3.1 Bigrammi non frequenti/non attestati

Per questo parametro, l'analisi è basata su bigrammi che non possono essere considerati collocazioni, in quanto non frequenti o non attestati nel corpus di riferimento. Tale sezione illustrerà i fenomeni incontrati suddividendoli in categorie quali bigrammi creativi, bigrammi altamente specifici, calchi, ecc..

5.3.1.1 Inglese

Testo 048

Il testo contiene una percentuale di bigrammi non frequenti/non attestati maggiore nel tradotto che nell'interpretato.

Dall'analisi sono stati esclusi due bigrammi ("municipal separate-collection" e "dry separation") non ben formati dal punto di vista morfosintattico: infatti tali bigrammi funzionano da pre-modificatore di un altro nome ("facilities"), che dovrebbe quindi essere la testa del bigramma da analizzare. Tali casi sono il risultato di imprecisioni nella fase di estrazione automatica dei bigrammi (si veda 4.2).

Inoltre, si è deciso di non analizzare il bigramma "Italian response", estratto sia per il testo interpretato che per quello tradotto e resa dello stesso segmento del st ("risposta italiana").

Nell'esempio 1, il bigramma "mere politicking" traduce il bigramma utilizzato nel st ("mero campanilismo").

1 TT-IN-EN: I'm not going to speak just for
mere politicking.

TT-TR-EN: I am not taking the floor **simply to stand up for my region.**

ST-IN-IT: non intervengo per **mero campanilismo**

ST-TR-IT: non intervengo per **mero campanilismo**

In entrambi i casi si tratta di bigrammi poco comuni, non attestati nei corpora di riferimento e poco attestati sul web (462 occorrenze per l'italiano e 529 per l'inglese). Tuttavia, mentre il traduttore si allontana dal rispettivo st rendendo il bigramma attraverso una parafrasi che ne trasmette tutto il senso, l'interprete mantiene la connotazione negativa, ma restringe il campo a cui fa riferimento il nome. Infatti, "campanilismo" si adatta a un popolo o a una regione geografica, mentre "politicking" denota interesse unicamente per il proprio partito politico. Si riportano qui di seguito le definizioni fornite da due dizionari monolingue e gli esempi relativi al bigramma analizzato.

campanilismo s. m. [der. di campanile]. – Attaccamento esagerato e gretto alle tradizioni e agli usi della propria città.

politicking noun [uncountable] (often disapproving). Political activity, especially to win support for yourself

Il secondo bigramma estratto per l'interpretazione è "communal plant" (2).

2 TT-IN-EN: 50 millions to fund **communal plant** for differentiated waste.

TT-TR-EN: and EUR 50 million for the financing of **municipal** separate-collection **facilities.**

ST-IN-IT: 50 milioni di euro per il finanziamento di **impianti comunali** per la raccolta differenziata.

ST-TR-IT: 50 milioni di euro per il finanziamento di **impianti comunali** per la raccolta differenziata.

Esso sembra costituire un calco dell'espressione del st "impianti comunali", in quanto l'aggettivo "communal" si riferisce a una cosa in comune, proprietà di una comunità, mentre l'aggettivo corretto da utilizzare, riferito al Comune come istituzione, sarebbe quello impiegato in traduzione, ossia "municipal". Si riportano di seguito le definizioni del dizionario monolingue relative ai due aggettivi.

communal adjective. 1 Shared by, or for the use of, a number of people, especially people who live together

Synonym **shared**

a communal kitchen/garden, etc.

As a student he tried communal living for a few years.

2 Involving different groups of people in a community

communal violence between religious groups

municipal adjective. Connected with or belonging to a town, city or district that has its own local government

municipal elections/councils

municipal workers

the Los Angeles Municipal Art Gallery

L'ipotesi di calco è avvalorata dal fatto di non aver reperito occorrenze nel corpus di riferimento. Ricercando sul web e mettendo a confronto "communal plant" e "municipal plant", entrambi i bigrammi sembrano poco diffusi; tuttavia per "municipal plant" si trovano più occorrenze (460 contro 420 di "communal plant") e la distinzione di significato sembra tenere. Si è deciso di effettuare anche una ricerca relativamente al bigramma utilizzato dal traduttore, ossia "municipal [...] facilities". Quest'ultimo risulta molto meno diffuso rispetto al bigramma "communal facilities", sia nel corpus di riferimento (2 vs. 39) che sul web (299.000 vs.

373.000), ma bisogna comunque tener conto del fatto che il nome “facilities” è molto più generale rispetto a “plant” e che “communal” è sempre inteso come servizio in comune o alla comunità.

Per quanto riguarda i bigrammi estratti per il testo tradotto, il bigramma “underused areas” (3) risulta non attestato/poco frequente in quanto fa parte dell’esplicitazione della sigla FAS (“Fund for Underused Areas”).

3 TT-TR-EN: Consequently, a sum from the Fund for **Underused Areas** (FAS) has been set aside thus far to ensure the continuation of the activities subject to the European veto

TT-IN-EN: working according to proper application of **the funding** as earmarked. So, this has continued despite the Community ban

ST-TR-IT: quindi, ad oggi si è accantonata una quota dei **FAS** per consentire il prosieguo delle attività in vigenza del veto comunitario

ST-IN-IT: quindi, ad oggi si è accantonata una quota dei **FAS** per consentire il proseguo delle attività in vigenza del veto comunitario

Una maggiore esplicitezza è in questo caso necessaria dato che la sigla denota una realtà prettamente italiana, ma la strategia non viene utilizzata nell’interpretato. Il fenomeno dell’esplicitazione è caratteristico della traduzione piuttosto che dell’interpretazione, ad eccezione di casi rari, in quanto l’interprete non ha a disposizione il tempo necessario per spiegare gli elementi più opachi, a meno che non sia l’oratore stesso a farlo nel suo discorso (in questo caso l’oratore non esplicita la sigla nel st). In questo esempio l’interprete, oltre a non esplicitare, non pronuncia nemmeno la

sigla poiché si trova molto probabilmente in un momento di difficoltà e si vede costretto a omettere diverse parti del discorso.

L'ultimo bigramma (4) estratto per la traduzione ("European veto") è stato reso con un bigramma anche in interpretazione ("Community ban").

4 TT-TR-EN: the activities subject to the **European veto** on the granting of ROP funds.

TT-IN-EN: So, this has continued despite the **Community ban** b- using previous funds

ST-TR-IT: delle attività in vigore del **veto comunitario** al riconoscimento dei fondi POR.

ST-IN-IT: delle attività in vigore del **veto comunitario** al riconoscimento dei fondi POR.

Entrambi i bigrammi non sono attestati nel corpus di riferimento e sono poco diffusi sul web (rispettivamente 409 e 373 occorrenze). Per accertarsi dell'eventuale differenza di uso e/o significato tra l'opzione scelta dal traduttore e dall'interprete si è deciso di controllare la definizione di "Community" sul dizionario.

the European Community. The former name (1967–1993) of the European Union.

Secondo la definizione riportata sopra, quindi, il nome dovrebbe essere diventato obsoleto a partire dal 1993 e sostituito dal più attuale "European" (utilizzato dal traduttore). Controllando su EPTIC_01_2011 e tenendo comunque conto delle dimensioni relativamente modeste del corpus, l'utilizzo del modificatore "Community" sembra essere abbastanza diffuso tra gli interpreti (12 occorrenze), meno tra i traduttori (1 occorrenza), ma non se ne trova nessuna occorrenza nei st inglesi. Si può dunque affermare che l'utilizzo di tale modificatore come sinonimo di

“European” sembra essere accettato tra gli interpreti delle istituzioni europee, ma non necessariamente la soluzione per cui opta l’interprete è corretta.

Testo 034

Il testo contiene una percentuale di bigrammi non frequenti/non attestati maggiore nell’interpretato che nel tradotto.

Anche in questo caso dall’analisi sono stati esclusi alcuni bigrammi (“quality tourist”, “longer regard” e “other cross-cutting”) che risultano essere errori di estrazione. Il caso di “quality tourist” corrisponde a quello visto nell’analisi del testo precedente, con l’estrazione di una sequenza modificatore+nome che tuttavia funge a sua volta da modificatore per la testa “destinations” andando a formare il gruppo nominale “quality tourist destinations”. Per gli altri due casi, invece, sembra esserci stato un problema di tagging che ha quindi causato problemi di estrazione. In “longer regard”, “regard” è stato trattato come nome. In realtà “regard” è un verbo inserito in un’espressione più ampia, ossia “cannot any longer regard tourism as”. Nel bigramma “other cross-cutting”, facente parte del gruppo nominale più esteso “other cross-cutting initiatives”, l’aggettivo “cross-cutting”, è stato considerato un nome: essendo il termine poco frequente, il tagger non è riuscito a riconoscere la part of speech da attribuire e di default ha inserito il tag nome.

Anche per questo testo avremmo dovuto escludere dall’analisi il bigramma “consolidated framework”, estratto sia per l’interpretazione che per la traduzione. In realtà il bigramma si trova in due frammenti diversi di testo nelle due modalità di mediazione.

Sempre partendo dall'analisi dei bigrammi estratti per il testo interpretato, si può notare che i due casi di "favourable context" (5-6) vengono resi in traduzione con due espressioni diverse, entrambe contenenti lo stesso sostantivo ("environment"), meno legato al st ("contesto") rispetto al nome utilizzato dall'interprete.

5 TT-IN-EN: A ehm **favourable context** for SMEs and ehm one-man companies

TT-TR-EN: The creation of an **environment that is conducive** to the development of micro, small and medium-sized enterprises

ST-IN-IT: La creazione di un **contesto favorevole** allo sviluppo delle medie, piccole e piccolissime imprese

ST-TR-IT: La creazione di un **contesto favorevole** allo sviluppo delle medie, piccole e piccolissime imprese

6 TT-IN-EN: we will be trying to c- create in other ways a **favourable context**.

TT-TR-EN: will likewise help to create a **favourable environment** for the sector's operators.

ST-IN-IT: permetteranno di contribuire in uguale maniera alla creazione di un **contesto favorevole** per gli attori del settore.

ST-TR-IT: permetteranno di contribuire in eguale maniera alla creazione di un **contesto favorevole** per gli attori del settore.

In entrambi i casi, sembra che il traduttore intenda evitare espressioni vicine a quelle del st, anche quando non presentano problemi. Infatti, il bigramma "favourable context", seppur poco

attestato nel corpus di riferimento, risulta abbastanza diffuso sul web (circa 50.000 occorrenze). Nonostante ciò, il bigramma “favourable environment” preferito dal traduttore in 6 risulta più frequente sia nel corpus di riferimento (16 occorrenze) sia sul web (più di 700.000 occorrenze) e l’espressione impiegata nel testo tradotto dell’esempio 5 sembra conferire un tono più formale al discorso.

In due casi (“young businessmen” e “one-man companies”, 7-8), i bigrammi che si trovano nell’interpretazione vengono resi in traduzione con espressioni più neutrali dal punto di vista del genere (“young entrepreneurs” e “micro [...] enterprises”). Inoltre, nel secondo caso, la scelta effettuata nella traduzione non fornisce l’informazione riguardante il numero di persone facenti parte dell’impresa, più in linea con il testo prodotto dall’oratore (“piccolissime imprese”).

- 7 TT-IN-EN: Erasmus for **young businessmen**
TT-TR-EN: Erasmus for **young entrepreneurs**
ST-IN-IT: l’Erasmus per i **giovani imprenditori**
ST-TR-IT: l’Erasmus per i **giovani imprenditori**
- 8 TT-IN-EN: for SMEs and ehm **one-man companies**
one-person companies
TT-TR-EN: to the development of **micro**, small
and medium-sized **enterprises**
ST-IN-IT: allo sviluppo delle medie, piccole
e **piccolissime imprese**
ST-TR-IT: allo sviluppo delle medie, piccole
e **piccolissime imprese**

In un caso (9), l'interprete sembra optare per una soluzione di registro più basso ("broad palette") rispetto al traduttore, che impiega un termine unico, una nominalizzazione, per trasmettere il messaggio del st ("diversification"). Spesso l'abbassamento di registro da parte dell'interprete può essere dovuto al diverso contesto di produzione in cui operano l'interprete e il traduttore: infatti, è come se il primo avesse un dialogo diretto con il pubblico che ascolta, mentre il secondo traspone la distanza che lo separa dal pubblico nel testo che produce.

9 TT-IN-EN: to promote the offer ehm the offer of a broad spectrum of a **broad palette** of the ehm forms of tourism

TT-TR-EN: to promote the **diversification** of tourist services

ST-IN-IT: per la promozione di un'**offerta** turistica **diversificata**

ST-TR-IT: per la promozione di un'**offerta** turistica **diversificata**

Come già accennato in precedenza, il bigramma "consolidated framework", che sembra essere comune sia all'interpretazione che alla traduzione, si trova in realtà in due frammenti diversi di testo nelle due modalità di mediazione (10-11).

10 TT-IN-EN: a new **consolidated framework** to strengthen

TT-TR-EN: a **consolidated political framework** to strengthen

ST-IN-IT: un **quadro politico consolidato** per rafforzare

ST-TR-IT: un **quadro politico consolidato** per rafforzare

11 TT-TR-EN: The creation of an environment that is conducive [...] is the principle on which this **consolidated framework** is based.

TT-IN-EN: A ehm favourable context [...] is also part of this **framework**.

ST-TR-IT: La creazione di un contesto favorevole [...] rappresenta il principio fondamentale di questo **quadro consolidato**.

ST-IN-IT: La creazione di un contesto favorevole [...] rappresenta il principio fondamentale di questo **quadro consolidato**.

In entrambi i casi l'interprete, che utilizza le espressioni "consolidated framework" e "framework", opera una riduzione sia rispetto al st ("quadro politico consolidato" e "quadro consolidato") che rispetto alla resa traduttiva ("consolidated political framework" e "consolidated framework"). Nell'esempio 11, l'interprete molto probabilmente decide di ridurre al minimo l'espressione dato che si è già parlato del referente nel corso del testo e ha dunque la possibilità di risparmiare tempo prezioso. Nell'esempio 10, invece, la condensazione è dovuta forse alla mancanza di tempo che porta l'interprete a omettere l'aggettivo meno importante, ossia "political". In questi due casi il bigramma "consolidated framework" risulta non attestato non tanto perché creativo, ma perché specifico del dominio istituzionale.

Anche i restanti bigrammi estratti per la traduzione sembrano subire strategie di condensazione o omissione da parte dell'interprete per ragioni diverse. Nel caso di "same enterprises" (12), confrontando meglio i tt con i st risulta che in realtà il traduttore utilizza il bigramma non attestato/non frequente per evitare la ripetizione del termine tecnico "SMEs".

12 TT-TR-EN: small and medium-sized enterprises [...] are fully covered by [...] the Small Business Act, which is intended to promote the business environment of those very **same enterprises**.

TT-IN-EN: SMEs [...] are fully entitled to benefit from the ehm measures that intend to ehm improve the operational context for **SMEs**.

ST-TR-IT: le piccole e medie imprese [...] beneficiano a pieno titolo [...] dello Small Business Act che mira a facilitare il contesto operativo proprio delle **piccole e medie imprese**

ST-IN-IT: le piccole e medie imprese [...] beneficiano ehm a pieno titolo [...] dello Small Business Act che mira a facilitare il contesto operativo proprio delle **piccole e medie imprese**

Negli altri due casi (“repayable aid” ed “electronic observatory”, 13-14), si tratta di bigrammi altamente specializzati e tecnici. Nel primo caso il bigramma rientra in una serie che l'interprete decide di omettere.

13 TT-TR-EN: risk capital, loans, securities, micro-credit and other forms of **repayable aid**.

TT-IN-EN: venture capital, loans ehm and loan guarantees.

ST-TR-IT: del capitale di rischio, dei prestiti, delle garanzie, del microcredito e di altre forme di **aiuti rimborsabili**.

ST-IN-IT: del capitale di rischio, dei prestiti, delle garanzie, del microcredito e di altre forme di **aiuti rimborsabili**.

Nell'altro caso, invece, l'interprete in difficoltà riduce la frase e la generalizza ("other ways"), perdendo però la sfumatura di senso che l'oratore del st voleva trasmettere.

14 TT-TR-EN: and the creation of a virtual **electronic observatory** for European tourism will likewise help to create a favourable environment.

TT-IN-EN: and equally we will be trying to c- create in **other ways** a favourable context.

ST-TR-IT: e la creazione di un **osservatorio** virtuale **con supporto elettronico** per il turismo europeo permetteranno di contribuire in eguale maniera alla creazione di un contesto favorevole

ST-IN-IT: e la creazione di un'osservazio- di un **osservatorio** virtuale **con supporto elettronico** per il turismo europeo permetteranno di contribuire in uguale maniera alla creazione di un contesto favorevole

5.3.1.2 Italiano

Testo 017

Il testo contiene una percentuale di bigrammi non frequenti/non attestati maggiore nel tradotto che nell'interpretato.

Anche per questo testo sono stati esclusi dall'analisi due dei bigrammi estratti ("tropicale aumenti" e "possibile accordi"). Nel caso di "tropicale aumenti" si è riscontrato un errore di tagging, in

quanto “aumenti” non è un nome ma un verbo e “tropicale” si riferisce a “legname”, testa del bigramma “legname tropicale”. Nel caso di “possibile accordi”, invece, “accordi” è effettivamente un nome, ma “possibile” fa parte dell’espressione “prima possibile” e non è quindi un aggettivo riferito al nome che segue.

Anche per questo testo un bigramma (“legname illegale”) è stato estratto sia per l’interpretazione che per la traduzione, ma, come nel caso di “consolidated framework” visto in precedenza, il bigramma si trova in due frammenti diversi di testo nelle due modalità di mediazione ed è stato quindi preso in considerazione per l’analisi.

Il bigramma “legname illegale” estratto per l’interpretazione viene reso diversamente rispetto al testo tradotto (1). Il traduttore, infatti, così come l’oratore del st usano rispettivamente “legno” e “timber” in riferimento a un altro sostantivo, “prodotti”/“products”.

1 TT-IN-IT: Dovremmo quindi ehm evitare che non sia più un mercato di **legname illegale** in tutto il mondo.

TT-TR-IT: Il fatto che siamo il maggiore mercato mondiale di **prodotti di legno illegale** non deve mettere a repentaglio il nostro impegno

ST-IN-EN: We must take care not to undermine our own efforts [...] while simultaneously providing the largest market for illegally.. **illegal timber products** in the world.

ST-TR-EN: We must take care not to undermine our own efforts [...] by simultaneously providing the largest market for **illegal timber products** in the world.

Mentre in inglese sia “timber” che “wood” possono essere utilizzati intercambiabilmente per indicare sia il materiale da costruzione che quello con cui fabbricare prodotti, in italiano si preferisce utilizzare “legname” esclusivamente per il materiale da costruzione e “legno” sia per la costruzione che per la fabbricazione di oggetti.

timber noun 1 [uncountable]. Trees that are grown to be used in building or for making things

the timber industry

2 [uncountable] (especially British English) (North American English usually lumber). Wood that is prepared for use in building, etc.

houses built of timber

wood noun1 [uncountable, countable]. The hard material that the trunk and branches of a tree are made of; this material when it is used to build or make things with, or as a fuel

He chopped some wood for the fire.

a plank of wood

All the furniture was made of wood.

a wood floor

furniture made of a variety of different woods

In the spring, cut out the old wood and shorten the young stems (= of bushes, etc.).

legname s. m. [lat. lignamen «armatura di legno»]. – Legno, o più spesso (con valore collettivo) insieme di legni che, interi, segati o spaccati, s’adoperano come materia prima per costruzione e per lavoro: l. fresco, asciutto, stagionato, giovane, forte, dolce, ecc. (v. legno); l. da opera, quello destinato a costruzioni o alla lavorazione, contrapp. al l. da ardere.

legno s. m. [lat. lignum]. – b. Sotto l’aspetto merceologico, il materiale fornito dai tronchi delle gimnosperme delle dicotiledoni e che viene usato per lavori di falegnameria o per altri usi: legno da

ardere (detto più com. legna), legno da costruzione e da lavoro
(detto anche legname)

Nell'esempio in 1 si può inferire che siano in atto gli effetti della produzione in tempo reale, in cui l'interprete segue l'oratore: forse perché sente tardi la parola "products", decide di non riformulare ciò che ha appena detto così da non perdere il filo del resto del discorso.

Nell'esempio in 2, mentre la traduzione si discosta maggiormente dal st, l'interprete ricalca la forma della frase pronunciata dall'oratore.

2 TT-TR-IT: Attualmente la quantità di **legname illegale** importato nell'Unione europea rappresenta quasi un quinto

TT-IN-IT: Al momento, il **legname che è abbattuto** e ehm im- esportato **illegalmente** in Europa ammonta ad un quinto

ST-TR-EN: At the moment, the amount of **timber that is illegally harvested** and imported into the EU constitutes almost one fifth

ST-IN-EN: At the moment... th- the amount of **timber that is illegally harvested** and imported into the EU constitutes almost one fifth

Le due occorrenze di "legname tropicale" reperite per la traduzione non sono presenti nell'interpretazione: nell'esempio 3, l'interprete omette il bigramma riformulando la frase del st, mentre nell'esempio 4, l'interprete non utilizza l'aggettivo "tropicale" ma lo sostituisce con un deittico ("questo") e generalizza, facendo riferimento a "legname illegale", utilizzato in precedenza nel testo.

3 TT-TR-IT: La pratica diffusa dello sfruttamento di **legname tropicale** è alimentata da un insaziabile appetito per prodotti destinati ad abbellire le nostre case.

TT-IN-IT: E vi sono però taluni che sono ehm han- ehm un appetito insaziabile di queste risorse.

ST-TR-EN: The widespread removal of **tropical timber** is driven by an insatiable appetite for products to grace our homes.

ST-IN-EN: The widespread removal of **tropical timber** is driven by an insatiable appetite for products to grace our homes.

4 TT-TR-IT: Potrebbe darsi che, per effetto degli accordi, il prezzo del **legname tropicale** aumenti, ma noi non dovremmo opporci

TT-IN-IT: Dovremmo quindi ehm evitare che non sia più un mercato di legname illegale in tutto il mondo. Sappiamo che il **questo tipo di legname** è estremamente costoso e non ci ob- ehm non ci opponiamo

ST-TR-EN: The result of these agreements may mean that **tropical timber** reaches a higher price, and that is something that we should not object to.

ST-IN-EN: The result of these agreements may mean that **tropical timber** reaches a higher price, and that is something that we should not object to.

Sia nel caso di “legname illegale” che di “legname tropicale” è stata effettuata una ricerca nel corpus di riferimento e sul web

per verificare la diffusione dei due bigrammi e metterli a confronto con le possibili alternative (“legno illegale” e “legno tropicale”). Nel corpus di riferimento il bigramma “legno illegale” risulta più collocazionale del bigramma “legname illegale” (rispettivamente 12 e 0 occorrenze). La tesi è avvalorata dalle occorrenze trovate sul web: circa 36.000 per “legno illegale” contro le circa 5.000 per “legname illegale”. Anche il bigramma “legno tropicale” risulta più diffuso rispetto a “legname tropicale”: il corpus di riferimento fornisce rispettivamente 4 occorrenze e 1 occorrenza, mentre sul web si trovano rispettivamente circa 80.000 e circa 7.000 occorrenze.

Il bigramma “base vitale” viene reso nell’interpretato con una parafrasi (5), attraverso la quale l’interprete, semplificando ed eliminando, perde però la sfumatura di senso che l’oratore aveva voluto trasmettere, ossia una foresta che non solo ospita ma offre anche sostentamento.

5 TT-TR-IT: perché essa costituisce la **base vitale** di un gran numero di specie animali e vegetali

TT-IN-IT: oltre al fatto che **ospitano** milioni e migliaia di piante e di animali diversi.

ST-TR-EN: that not only **sustains a whole host** of plant and animal species

ST-IN-EN: that not only **sustains a whole host** of planim- plant and animal species

Oltre a non rendere il verbo “sustains”, carico del significato originale voluto dall’oratore del st, infatti, l’interprete riporta il termine “host” in italiano come verbo “ospitare”. In questo caso l’interprete è conscio del significato che il termine “host” assume se utilizzato come sostantivo (si veda definizione riportata sotto) e

lo rende nel modo giusto (“milioni e migliaia”). Tuttavia, esso utilizza lo stesso termine del *st* come base per la resa del verbo (“ospitare”) che però perde il significato che voleva trasmettere il *st*.

host noun 5 [countable]. Host of somebody/something. A large number of people or things
a host of possibilities

Il bigramma “base vitale” sembra inoltre abbastanza creativo, in quanto non attestato nel corpus di riferimento e poco attestato sul web (circa 460 risultati).

L'ultimo bigramma estratto per la traduzione, “gestione accorta”, viene reso dall'interprete con un bigramma diverso, ossia “attenta gestione” (6).

6 TT-TR-IT: Grazie a una **gestione accorta** è possibile ridurre
TT-IN-IT: Comunque un'**attenta gestione** potrà anche ridurre
ST-TR-EN: **Careful management** can bring down
ST-IN-EN: **Careful management** can bring down

Si è ritenuto opportuno controllare le occorrenze dei due bigrammi per capire se a differenza del traduttore l'interprete usi un'espressione maggiormente collocazionale. Verificando sul corpus di riferimento, si può notare che il bigramma “attenta gestione” è molto più comune rispetto a “gestione accorta” (rispettivamente 24 occorrenze e 1 occorrenza). Controllando le occorrenze sul web, l'ipotesi di maggiore collocationalità di “attenta gestione” viene avvalorata, dato che le occorrenze per “gestione accorta” sono circa 39.000 al contrario di quelle per “attenta gestione” che sono circa 70.000. Si può dunque affermare

che, nonostante il fatto che l'interprete si uniformi di più al st mentre il traduttore, come già visto in un caso del tt inglese, sembri preferire allontanarsi dal relativo st, la resa dell'interprete risulta più collocazionale rispetto a quella presente nel testo tradotto.

Testo 006

Il testo contiene una percentuale di bigrammi non frequenti/non attestati maggiore nell'interpretato che nel tradotto.

Come in precedenza, anche per questo testo non è stato preso in considerazione il bigramma "lenta lento" a causa di un errore di estrazione. Infatti il bigramma è composto da due aggettivi, dovuti a un'autocorrezione dell'interprete che inizialmente sbaglia l'accordo con il sostantivo che segue.

Per il presente testo si è inoltre deciso di non analizzare il bigramma "resistenze corporativistiche", estratto non solo per l'interpretato ma anche per il tradotto e derivato dalla stessa espressione del st ("corporate resistance").

Due bigrammi sono stati estratti da una citazione (7). Si tratta di "buoni steccati" e "buone recinzioni", rispettivamente per l'interpretato e il tradotto.

7 TT-IN-IT: "buoni steccati fanno buoni vicini"

TT-TR-IT: "buone recinzioni fanno buoni vicini"

ST-IN-EN: "Good neighbours have good fences"

ST-TR-EN: "Good fences make good neighbours"

Mentre l'interprete rende con una citazione approssimativa, forse il traduttore ha la possibilità di documentarsi e riprendere

esattamente le parole impiegate dall'autore della citazione. Inoltre, confrontando i tt con i rispettivi st si può notare che anche l'oratore non utilizza le parole esatte dell'autore della citazione.

Uno dei bigrammi estratti per l'interpretazione ("piccole corporazioni") costituisce un calco dal termine impiegato nel st inglese, "corporations" (8).

8 TT-IN-IT: la possibilità a **piccole corporazioni** di mettere sul mercato i loro prodotti.

TT-TR-IT: per le **imprese più piccole** ai fini dell'immissione sul mercato interno dei loro prodotti.

ST-IN-EN: for **smaller corporations** to place their products in the internal market

ST-TR-EN: for **smaller corporations** to place their products on the internal market

Infatti, come si può osservare dalla definizione del dizionario monolingue riportata sotto, il termine "corporazione" in italiano non denota un'impresa, un'azienda.

corporazione s. f. [dal lat. tardo corporatio -onis, der. di corporare (v. corporato), sul modello dell'ingl. corporation (da cui anche il fr. corporation)]. – 1. In età romana, unione di persone (sacerdoti, funzionari, artigiani) legate tra loro con vincolo volontario per la comunanza delle funzioni o della professione. Nell'ordinamento medievale, e fino al sec. 18°, complesso di persone che, svolgendo una comune attività economica, si univano per la tutela degli interessi e per il conseguimento di fini comuni: *corporazioni di arti e mestieri; la c. dei mercanti, dei professionisti.*

Il bigramma "regolamentazione migliore" estratto per l'interpretazione, viene reso dal traduttore con il bigramma "normativa migliore" (9).

9 TT-IN-IT: Nell'Unione si deve cercare di avere una **regolamentazione migliore**..

TT-TR-IT: All'interno dell'Unione europea, dovremo puntare a una **normativa migliore**.

ST-IN-EN: In the European Union, we should always strive for **better regulations**.

ST-TR-EN: In the European Union, we should always strive for **better regulation**.

Fermo restando che l'interprete con "regolamentazione" si uniforma maggiormente al sostantivo del st "regulations", controllando le occorrenze sia sul corpus di riferimento che sul web risulta che entrambi i bigrammi sono poco attestati: "regolamentazione migliore" occorre una sola volta nel corpus di riferimento e 241 volte sul web, mentre "normativa migliore" non è attestato nel corpus di riferimento e presenta 240 occorrenze sul web. Provando a ricercare i due bigrammi nella forma aggettivo+nome ("migliore regolamentazione" e "migliore normativa"), risulta che il primo bigramma è più frequente sia nel corpus di riferimento che sul web ("migliore regolamentazione": rispettivamente 9 e circa 20.000 occorrenze; "migliore normativa": rispettivamente 0 e circa 4.000 occorrenze). È quindi chiaro che sia l'interprete che il traduttore utilizzano bigrammi poco comuni, ma nonostante ciò l'interprete si avvicina maggiormente a un bigramma più collocazionale.

Nel caso di "condizioni separate" (10), l'interprete usa un bigramma non attestato nel corpus di riferimento e poco attestato sul web (337 occorrenze) e sembra trasporre un'espressione probabilmente familiare dell'inglese ("separate conditions"), nonostante questa non sia presente nel st.

10 TT-IN-IT: N- naturalmente ci possono essere delle ehm **condizioni separate**

TT-TR-IT: Questo, assieme alle **condizioni speciali** predisposte

ST-IN-EN: in particular, the **special conditions** set for

ST-TR-EN: in particular, the **special conditions** set for

In “condizioni separate” l’interprete riflette non solo un’espressione comune in lingua inglese quale “separate conditions” (solo 5 occorrenze nel corpus di riferimento, ma circa 100.000 occorrenze sul web), ma anche il significato dell’aggettivo “separate” (si veda definizione riportata sotto).

separate adjective. 2 [usually before noun] different; not connected

It happened on three separate occasions.

For the past three years they have been leading totally separate lives.

Il traduttore, seppur più vicino al st utilizza invece il bigramma più collocazionale “condizioni speciali” (39 occorrenze nel corpus di riferimento).

Due dei bigrammi estratti per il testo tradotto, “signor Haerynck” (11) e “diverse correlazioni” (12), corrispondono a casi di omissione da parte dell’interprete. Il caso in 11, nonostante non sia effettivamente un bigramma da prendere in considerazione dato che è formato da un appellativo e un nome proprio, dimostra che i nomi propri costituiscono un problema per gli interpreti, data anche la varietà di pronunce che può caratterizzare gli oratori. L’interprete infatti omette il cognome di una delle persone di cui sta parlando l’oratore.

11 TT-TR-IT: Vorrei anche esprimere la mia gratitudine nei confronti della segreteria dell'IMCO e della presidenza belga, rivolgendomi in particolare al signor Vertessen e al **signor Haerynck** per l'attento lavoro svolto.

TT-IN-IT: Ringrazio anche il segretariato dell'IMCO e la presidenza belga, in particolare il signor Vertessen... per il lavoro che hanno svolto.

ST-TR-EN: I would also like to put on record my thanks to the IMCO Secretariat and to the Belgian Presidency, in particular, to Mr Vertessen and **Mr Haerynck** for all their dedicated work.

ST-IN-EN: I would also like to put on record my thanks to the IMCO Secretariat and to the Belgian Presidency, in particular, to Mr Vertessen and **Mr Haerynck** for all their dedicated work.

Anche nell'esempio in 12 l'interprete omette, ma in questo caso non omette semplicemente il bigramma, ma la frase intera, ricostruendo tra l'altro una parte del discorso.

12 TT-TR-IT: causate dalle condizioni lavorative e dall'inalazione di polvere. La corretta etichettatura delle sostanze pericolose può aiutare quanti soffrono di queste patologie polmonari ed illustra le **diverse correlazioni** del nostro lavoro con altri ambiti. Sono orgogliosa del fatto che questa relazione sia stata gestita e portata a compimento durante l'Anno europeo del polmone.

TT-IN-IT: Naturalmente, con buone condizioni e buone norme di sicurezza possiamo permettere a questi lavoratori di lavorare a norme migliori a condizioni migliori di sicurezza... Tra l'altro c'è l'Anno europeo del polmone

ST-TR-EN: caused by working conditions and by inhaling dust. The proper labelling of hazardous substances can help those with these lung conditions and illustrates **how interlinked** our work here is. I am proud of the fact that this report was conducted and concluded under the European Year of the Lung.

ST-IN-EN: in their working conditions and inhaling dust. The proper labelling of hazardous substances can help those with these lung conditions and illustrates **how interlinked** our work here is. I'm proud of the fact that this report was conducted and concluded under the European Year of the Lung.

Le cause di tali omissioni sono probabilmente la velocità abbastanza elevata a cui parla l'oratore (154,8 parole al secondo), il fatto che l'intervento venga letto e che quindi presenti strutture più complesse e l'accento non standard, dato che l'oratore è di origine scozzese.

Riassumendo, ciò che si può osservare maggiormente nella resa dei bigrammi non frequenti/non attestati da parte di interpreti e traduttori è una tendenza da parte degli interpreti a cambiamenti di significato e all'omissione.

Si passerà ora all'analisi dei candidati a collocazione per il parametro alto *t*-score, sempre per entrambe le lingue target.

5.3.2 Alto t-score

Per il presente parametro sono state estratte collocazioni ad alto t-score, ossia molto frequenti (si veda 4.2). Come nel caso della lowfq, vengono presi in considerazione due testi per lingua, uno la cui interpretazione è più collocazionale della traduzione e uno la cui traduzione è più collocazionale dell'interpretazione.

5.3.2.1 Inglese

Testo 041

Il testo presenta una percentuale maggiore di collocazioni nella traduzione che nell'interpretazione.

Il bigramma “few years” non è stato preso in considerazione in fase di analisi in quanto reso allo stesso modo da parte di traduttore e interprete a partire dallo stesso st (“ultimi anni”).

L'unica altra collocazione estratta per il testo interpretato, “other options”, viene resa dall'interprete in maniera più semplice sia rispetto al testo tradotto, dove si incontra il bigramma “virtuous alternatives”, che a entrambi i st, dove si legge il bigramma corrispondente “alternative virtuose” (1). Tale strategia di generalizzazione non si applica unicamente a tale espressione, ma all'intero periodo.

1 TT-IN-EN: the waste Directive says that this is the last possible form but with **other options** that won't work against citizen's health and will create dro- job.

TT-TR-EN: Pursuant to the waste Directive, incinerators shall be the last option in a hierarchy of integrated waste systems that presents **virtuous alternatives** that have no

public health or environmental impact but instead create employment.

ST-IN-IT: Gli inceneritori, ce lo dice la direttiva rifiuti, sono l'ultima opzione gerarchica di un ciclo integrato di rifiuti che presenta **alternative virtuose** senza impatto sulla salute dei cittadini e sull'ambiente e che, anzi, creano occupazione.

ST-TR-IT: Gli inceneritori, lo dice la direttiva rifiuti, sono l'ultima opzione gerarchica di un ciclo integrato di rifiuti che presenta **alternative virtuose** senza impatto sulla salute dei cittadini e sull'ambiente e che, anzi, creano occupazione.

Per quanto riguarda le collocazioni reperibili nella traduzione ma non nell'interpretazione, sia un'occorrenza di "European legislation" che la collocazione "renewable energy" vengono omesse dall'interprete che forse si trova in difficoltà nel rendere le parole del st, probabilmente anche a causa della velocità a cui parla l'oratore. L'interprete usa quindi la strategia della parafrasi e rielabora il periodo (2).

2 TT-TR-EN: The European Union must insist that a worthy solution is found, or else its intervention will have been in vain. The construction of incinerators, which, by the way, are financed as **renewable energy** sources and thereby infringe current **European legislation**, will not solve the issue.

TT-IN-EN: The European Union has to make sure that its intervention is quick or it won't work. **There should be no more**

incinerators, they're a spot of waste disposal.

ST-TR-IT: L'Unione europea deve imporre che la soluzione sia virtuosa, altrimenti il suo intervento risulterebbe vano. Non è costruendo gli inceneritori, peraltro finanziati come **fonti rinnovabili** in contrasto con la **normativa europea** vigente, che si risolve la questione.

ST-IN-IT: L'Unione europea deve imporre che la soluzione sia virtuosa, altrimenti il suo interesse - il suo intervento risulterebbe vano. Non è costruendo gli inceneritori, peraltro finanziati come **fonti rinnovabili** in contrasto alla **normativa europea** vigente, che si risolve la questione.

La seconda occorrenza di "European legislation" estratta per il testo tradotto è invece resa dall'interprete con l'espressione "Community provisions" (3).

3 TT-TR-EN: in which **European legislation** on State aid is clearly infringed?

TT-IN-EN: where clearly were against the **Community provisions** on State aid?

ST-TR-IT: nel quale viene violata palesemente la **normativa comunitaria** relativa agli aiuti di Stato?

ST-IN-IT: nel quale viene violata palesemente la **normativa comunitaria** relativa agli aiuti di Stato?

Come già descritto in 5.3.1.1, nonostante il pre-modificatore "Community" sia teoricamente caduto in disuso e sia stato sostituito dal più attuale "European", in questo caso preferito dal traduttore, la scelta applicata all'interpretato pare essere accettata

tra gli interpreti delle istituzioni europee. È stata effettuata una ricerca sul corpus di riferimento relativamente alle due collocazioni in questione, “European legislation” e “Community provisions”, e relativamente agli stessi nomi accompagnati dall’aggettivo dell’altra collocazione (“European provisions” e “Community legislation”). La collocazione “European legislation” risulta molto più diffusa rispetto a “Community provisions” (rispettivamente 143 e 8 occorrenze). Tuttavia, confrontando lo stesso nome accompagnato da aggettivi diversi, risulta che “Community provisions” è più frequente di “European provisions” (8 vs. 1 occorrenza), mentre “European legislation” è più frequente di “Community legislation” (rispettivamente 143 e 30 occorrenze). Dunque, sulla base di tali dati, si può affermare che anche se meno diffusa rispetto alla collocazione utilizzata dal traduttore, quella impiegata dall’interprete è comunque corretta anche nella lingua corrente.

Le collocazioni “public health” e “environmental impact” meritano un discorso a parte. In realtà, “public health” dovrebbe essere considerato modificatore di “impact” al pari di “environmental” (4). Abbiamo deciso di analizzare comunque la collocazione perché si tratta di un bigramma di senso compiuto, a differenza di quelli non considerati nell’analisi dei testi con bigrammi non frequenti/non attestati.

4 TT-TR-EN: that have no **public health** or **environmental impact** but instead create employment.

TT-IN-EN: that won’t work against **citizen’s health** and will create dro- job.

ST-TR-IT: senza **impatto sulla salute dei cittadini e sull’ambiente** e che, anzi, creano occupazione.

ST-IN-IT: senza **impatto sulla salute dei cittadini e sull'ambiente** e che, anzi, creano occupazione.

In questo caso, l'interprete rende l'espressione "public health impact" con una costruzione più esplicita ("work against citizen's health"), non completamente nominale come è invece quella impiegata in traduzione, ricalcando maggiormente la costruzione verbo+preposizione tipica della lingua di partenza, l'italiano. Inoltre, l'interprete omette la resa di "environmental impact", probabilmente di nuovo per motivi di tempo a disposizione, perdendo una sfumatura di significato che l'oratore voleva trasmettere.

Testo 022

Il testo presenta una percentuale maggiore di collocazioni nell'interpretazione che nella traduzione.

Due delle collocazioni estratte ("natural disasters" per l'italiano "catastrofi naturali" e "humanitarian aid" per "aiuti umanitari") corrispondono tra l'interpretazione e la traduzione e non sono quindi state analizzate.

Una collocazione specifica del dominio in discussione come "human rights" presente in interpretazione non risulta tra le collocazioni estratte per la traduzione. Questo perché, come "public health" più in alto, la collocazione "human rights" nell'interpretazione funge da modificatore per "law", mentre nella traduzione il sostantivo "law" viene utilizzato con l'aggettivo "humanitarian" a formare il bigramma "humanitarian law", che tuttavia non è stato estratto come possibile candidato a collocazione (5).

5 TT-IN-EN: infringement of international **human rights law**.

TT-TR-EN: violations of international **humanitarian law**.

ST-IN-IT: violazioni del **diritto umanitario** internazionale

ST-TR-IT: violazioni del **diritto umanitario** internazionale

Controllando le occorrenze delle espressioni “human rights law” e “humanitarian law” nel corpus di riferimento risulta che entrambe sono attestate (rispettivamente 205 e 79 occorrenze) e dunque entrambe le scelte possono essere considerate corrette.

Le restanti due collocazioni estratte per l’interpretazione ma non per la traduzione (“international level” e “same time”), sono rese in traduzione attraverso scelte più formali (6).

6 TT-IN-EN: share the burdens out better at **international level**, at the **same time** there are regional responsibilities

TT-TR-EN: **worldwide** coordination and burden sharing, **taking into account** the regional responsibilities

ST-IN-IT: di un coordinamento e di una perequazione degli oneri a **livello mondiale**, tenendo conto allo **stesso tempo** delle responsabilità regionali

ST-TR-IT: di un coordinamento e di una perequazione degli oneri a **livello mondiale**, tenendo conto allo **stesso tempo** delle responsabilità regionali

Entrambe, inoltre, vengono rese attraverso strutture diverse da quella analizzata (aggettivo+nome/nome+aggettivo): in un caso

viene utilizzato infatti un aggettivo unico, “worldwide”, mentre nel secondo caso il traduttore opta per una costruzione preposizionale, ossia “taking into account”. Le scelte sono vincolate anche dalla costruzione della frase scelta rispettivamente da interprete e traduttore.

L’unica collocazione estratta per la traduzione ma non per l’interpretazione, “dramatic increase”, presenta comunque un corrispettivo nel testo interpretato (“exponential increase”), che tuttavia risulta molto meno carico emotivamente rispetto alla resa traduttiva, oltre a essere più formale, meno adatto a trasmettere il senso voluto dall’oratore e più legato al st che alla lingua di arrivo (7).

7 TT-TR-EN: the **dramatic increase** in the number and severity of natural disasters

TT-IN-EN: The **exponential increase** of natural disasters, both in their number and their seriousness

ST-TR-IT: l’**aumento esponenziale** delle catastrofi naturali, sia nel loro numero sia nella loro intensità

ST-IN-EN: L’**aumento esponenziale** delle catastrofi naturali, sia nel loro numero che nella loro intensità

Inoltre, a seguito di un controllo di entrambi i bigrammi nel corpus di riferimento, risulta che “exponential increase”, sebbene attestato, è meno frequente e diffuso rispetto a “dramatic increase” (rispettivamente 12 e 114 occorrenze).

5.3.2.2 Italiano

Testo 061

In questo caso il testo tradotto è più collocazionale rispetto a quello interpretato.

Cinque dei bigrammi reperiti per l'interpretazione ("Unione europea", "diritti umani", "Consiglio nazionale", "strada giusta" e "segnale importante") sono stati reperiti anche per il testo tradotto e risultano rese delle stesse espressioni utilizzate nel st ("European Union", "human rights", "National Council", "right track" e "important signal"). Tali bigrammi non sono quindi stati presi in considerazione per l'analisi.

Le altre tre collocazioni reperite unicamente per l'interpretazione ("grande attenzione", "caso concreto" e "attuale fase") sono state rese dal traduttore per mezzo di riformulazioni e parafrasi ("siamo particolarmente attenti", "caso specifico" e "fase attuale"), pur conservando il significato trasmesso dall'oratore del st (1-2-3).

1 TT-IN-IT: noi diamo **grande attenzione** ai diritti umani e alla protezione delle ehm dei diritti delle minoranze.

TT-TR-IT: siamo **particolarmente attenti** ai diritti umani e alla tutela dei diritti delle minoranze.

ST-IN-EN: we pay of course **in particular the attention** to human rights and also attention to the protection ehm of the minority rights.

ST-TR-EN: we pay **particular attention** to human rights and also to the protection of minority rights.

2 TT-IN-IT: che tratta proprio un **caso concreto** a cui lei ha fatto riferimento nel suo intervento.

TT-TR-IT: a proposito del **caso specifico** da lei citato nel suo intervento.

ST-IN-EN: address one **concrete case** you referred ehm to in your intervention.

ST-TR-EN: addressing one **concrete case** she referred to in her intervention.

3 TT-IN-IT: nei confronti dell'**attuale fase** di del processo di ampliamento.

TT-TR-IT: nella **fase attuale** del processo di allargamento.

ST-IN-EN: approach ehm to the **current stage** of enlargement process.

ST-TR-EN: approach to the **current stage** of the enlargement process.

Tutti e tre i casi dell'interpretazione ricalcano la struttura tipica della lingua di partenza, mentre gli equivalenti utilizzati nella traduzione subiscono una strategia di resa più addomesticante. Negli esempi 2 e 3, il traduttore usa due bigrammi diversi rispetto all'interprete: vale dunque la pena di fare una ricerca nel corpus di riferimento per verificare la maggiore o minore collocazionalità delle scelte operate. Per quanto riguarda l'esempio 2, la collocazione utilizzata dall'interprete ("caso concreto") risulta meno collocazionale rispetto a quella utilizzata dal traduttore ("caso specifico"): sono infatti state trovate 348 occorrenze per il primo bigramma e 702 per il secondo. Al contrario, nel caso riportato in 3, è l'interprete a utilizzare una collocazione

maggiormente attestata: “attuale fase” è presente 201 volte contro le 88 volte di “fase attuale”.

Nella traduzione vengono utilizzate molte collocazioni in più rispetto all’interpretazione. Diverse sono le cause di tali discrepanze.

Si trovano tre casi in cui l’interprete opera una riduzione rispetto al traduttore (“sistema giudiziario” vs. “giudiziario”, “prossime elezioni” vs. “elezione” e “Parlamento europeo” vs. “Parlamento”), forse per risparmiare tempo, pur non togliendo nulla al senso della frase (4-5-6).

4 TT-TR-IT: ad esempio, l’importante questione della riforma del **sistema giudiziario**.

TT-IN-IT: per esempio in riferimento alla riforma del **giudiziario**.

ST-TR-EN: for example, the important issue of **judicial** reform.

ST-IN-EN: for example, such an important issue as the **judicial** reform.

5 TT-TR-IT: attendiamo con ansia le **prossime elezioni** del Consiglio nazionale bosniaco

TT-IN-IT: Ora a- aspettiamo l’**elezione** del Consiglio nazionale bosniaco

ST-TR-EN: we look forward to the **next elections** to the Bosniak National Council

ST-IN-EN: we are looking forward to the **repeated election** ehm of the Bosniak National Council

- 6 TT-TR-IT: Il sostegno del **Parlamento europeo** al processo di ratifica dell'Accordo
- TT-IN-IT: L'appoggio dato dal **Parlamento** al ehm processo di ratifica dell'Accordo
- ST-TR-EN: The support provided by **Parliament** to the ratification process of the SAA
- ST-IN-EN: The support provided ehm by the **Parliament** to the ratification process of the SAA

Per quanto riguarda l'esempio in 4, è stata effettuata una verifica dell'attestazione o meno nel corpus di riferimento del termine "giudiziario" utilizzato come sostantivo nell'abbreviazione dell'espressione "sistema/potere giudiziario". Ricercando le espressioni "riforma del sistema giudiziario" e "riforma del giudiziario" risulta che, mentre la prima occorre 16 volte, la seconda non è attestata. Tuttavia, ricercando il termine "giudiziario" come sostantivo con preposto l'articolo ("il giudiziario"), questo risulta attestato (se ne trovano infatti 18 occorrenze).

In un caso ("sviluppo tecnologico"), la resa in traduzione risulta più completa rispetto a quella dell'interprete, il quale semplifica e condensa, utilizzando una parafrasi (7).

- 7 TT-TR-IT: Stiamo, inoltre, riflettendo sul sostegno da dare allo **sviluppo tecnologico**.
- TT-IN-IT: Stiamo riflettendo s- su ehm un'assistenza dal **punto di vista tecnologico**.
- ST-TR-EN: Reflection is ongoing regarding the support for **technological development**.

ST-IN-EN: Ehm reflection is going on regarding ehm the support to **technological development**.

Una delle due occorrenze della collocazione “Unione europea” è stata estratta per la traduzione ma non per l’interpretazione. L’interprete decide di omettere probabilmente conscio del fatto che il pubblico capirà ugualmente di cosa si sta discutendo (8).

8 TT-TR-IT: che ha dimostrato il sostegno concreto di questa Assemblea al processo di integrazione della Serbia **nell’Unione europea**.

TT-IN-IT: che dimostra il sostegno concreto del Parlamento al processo di integrazione della Serbia.

ST-TR-EN: which has shown the House’s constructive support of the process of the **European Union** integration of Serbia.

ST-IN-EN: which has shown the constructive support of your House to the **European Union** integration process of Serbia.

Si trovano inoltre due casi particolari di collocazioni mancate da parte dell’interprete, che però risultano nella lista delle collocazioni estratte relativamente alla traduzione.

Nel caso in 9, l’interprete utilizza la collocazione “radicalismo islamico” più vicina a quella impiegata nei due st (“radical Islamism”) piuttosto che a quella scelta dal traduttore (“fondamentalismo islamico”).

9 TT-TR-IT: preferirei parlare di una politicizzazione della questione della

rappresentanza musulmana, piuttosto che di un ritorno del **fondamentalismo islamico**.

TT-IN-IT: io parlerei piuttosto della politicizzazione della questione del ehm rappresentanza di musulmani, piuttosto che di un esacerbarsi del **radicalismo islamico**.

ST-TR-EN: I would rather talk about a politicisation of the issue of Muslim representation than a surge of **radical Islamism**.

ST-IN-EN: I would rather talk about the politicisation ehm of the issue of Muslim representation rather ehm than of the surge of **radical Islamism**.

Verificando sul corpus di riferimento risulta che entrambe le collocazioni sono attestate, con rispettivamente 38 e 140 occorrenze. Si è quindi deciso di confrontare l'utilizzo delle due espressioni sul web. Secondo R. Valery Rover, esperto di storia e istituzioni del mondo islamico, "il termine "fondamentalismo" non riflette la sua maggiore aderenza alla realtà cui viene abbinato, e va addebitato molto più semplicemente alla preponderanza della lingua inglese nel settore delle comunicazioni" (Rover, 1995).¹¹ Ecco quello che afferma invece Rover a proposito di "radicalismo islamico":

appare abbastanza diffuso, anche fra gli specialisti; è comodo da usare, e appropriato quando designa il ripensamento che questi movimenti fanno dell'Islam: un ritorno alle "radici", rilette, ristudiate e riscoperte nella loro purezza e forza primitiva, nell'autentico e genuino messaggio del Profeta, scevro da tutte quelle interpretazioni e sovrastrutture, spesso devianti, accumulatesi nel corso dei secoli o per ignoranza o per malafede dei governanti e dei loro consiglieri

¹¹ <http://gnosis.aisi.gov.it/sito%5CRivista2.nsf/servnavig/4>

(Rover, 1995).

Nonostante la frequenza con cui è usato dai media, il termine “fondamentalismo” risulta quindi meno politicamente corretto rispetto a “radicalismo”, a causa delle connotazioni negative che è venuto ad assumere nel corso degli anni.

Nel caso in 10, la collocazione “buon punto” estratta per la traduzione sarebbe presente anche nell’interpretazione, ma non viene considerata come collocazione con alto *t*-score in quanto l’interprete, forse fuorviato dal sostantivo impiegato nel st inglese (“basis”), sbaglia l’accordo dell’aggettivo (“buona punto”) e non si corregge.

10 TT-TR-IT: Questo è un **buon punto** di partenza affinché la Serbia sfrutti positivamente le possibilità che l’anno 2011 le offre.

TT-IN-IT: Questo è una **buona punto** di ehm partenza per la Serbia per valorizzare il potenziale di integrazione che quest’anno, 2011, gli offrirà.

ST-TR-EN: This is a **good basis** for Serbia to fully use the integration potential that the year 2011 offers.

ST-IN-EN: And this is a **good basis** for Serbia to fully use integration potential year 2011 offers.

Testo 034

In questo caso è il testo interpretato a essere più collocazionale rispetto a quello tradotto.

Nella maggior parte dei casi (11-12-13), la collocazione che si trova nell’interpretato viene resa nella traduzione con una singola parola che trasmette lo stesso significato (“maggior parte”

vs. “molti”, “certi casi” vs. “molti”, “genitori adottivi” vs. “adozione”, “genitori adottivi” vs. “adottanti”).

11 TT-IN-IT: questo avviene per la **maggior parte** dei bambini per i loro con i loro genitori, ma in **certi casi** con **genitori adottivi**.

TT-TR-IT: **Molti** lo trovano a casa, con i propri genitori, ma **molti** lo ottengono invece con l'**adozione**.

ST-IN-EN: for **most children** they find that in their own home with their natural parents. For **many** though, they get it ehm with their **adoptive parents**.

ST-TR-EN: **Most children** find that in their own home with their natural parents. **Many** though, they get this with their **adoptive parents**.

12 TT-IN-IT: se vi sono dei **genitori adottivi** potenziali ve- dovremmo facilitar loro la vita

TT-TR-IT: se vi sono potenziali **adottanti** in buona fede e irreprensibili, anziché asfissiarli con procedure interminabili, costose ed estenuanti, noi dovremmo aiutarli

ST-IN-EN: where there are prospective **adoptive parents** who've bona fides, are beyond reproach, so they should be facilitated

ST-TR-EN: where there are prospective **adoptive parents** who are bona fide and beyond reproach, to facilitate them

13 TT-IN-IT: la **maggior parte**, la **stragrande maggioranza** era in buone case

TT-TR-IT: **molti** di loro vivevano in famiglie inappuntabili

ST-IN-EN: **many** adopted children and the **vastest majority** were in very good homes.

ST-TR-EN: **many** adopted children and the **vast majority** were in very good homes.

Nell'esempio 13, tra l'altro, l'interprete utilizza due collocazioni ("maggior parte" e "stragrande maggioranza"), così come nei due st ("many" e "vastest majority" per il st-in, "many" e "vast majority" per il st-tr), mentre è il traduttore a condensare e a utilizzare un'unica parola, "molti".

Nel caso in 14, invece, l'interprete utilizza un aggettivo diverso rispetto al traduttore ("giusto equilibrio" vs. "delicato equilibrio"). La resa del traduttore è di registro più alto rispetto a quella dell'interprete e sembra quasi sfociare in un'espressione tipica del "politichese", altisonante ma in fondo priva di contenuto. Inoltre, la resa dell'interprete è ancora una volta più vicina al st ("get the balance right") rispetto alla traduzione.

14 TT-IN-IT: Dobbiamo trovare il **giusto equilibrio**

TT-TR-IT: Va trovato questo **delicato equilibrio**

ST-IN-EN: We need to get the **balance right**

ST-TR-EN: We need to get the **balance right**

Per la traduzione è stata estratta solamente una collocazione con alto *t*-score ("buona fede") che l'interprete non rende (15), probabilmente perché l'oratore utilizza una collocazione insolita ("bona fides"), che venendo dal latino può

essere considerata di registro alto, oltre che poco familiare e dunque più difficile da ricordare e di conseguenza da rendere.

15 TT-TR-IT: se vi sono potenziali adottanti in **buona fede** e irreprensibili, anziché asfissiarli con procedure interminabili, costose ed estenuanti, noi dovremmo aiutarli

TT-IN-IT: se vi sono dei genitori adottivi potenziali ve- dovremmo facilitar loro la vita

ST-TR-EN: where there are prospective adoptive parents who are **bona fide** and beyond reproach, to facilitate them

ST-IN-EN: where there are prospective adoptive parents who've **bona fides**, are beyond reproach, so they should be facilitated

Forse l'interprete omette anche a causa dell'errore che l'oratore commette nel pronunciare la collocazione: la forma corretta sarebbe infatti quella utilizzata nel st della traduzione, ossia "bona fide".

Nel caso delle collocazioni ad alto t-score analizzate sopra, tra i vari fenomeni incontrati si può notare una maggiore tendenza all'esplicitzza e all'omissione, non solo da parte dell'interprete ma anche da parte del traduttore.

Nel prossimo paragrafo verranno analizzati i candidati a collocazione con alti valori di MI. Come per i due parametri visti in precedenza, la prima parte dell'analisi verterà sui tt in lingua inglese, mentre la seconda sarà incentrata sui tt in lingua italiana.

5.3.3 Alto MI

Per il parametro alto MI sono state estratte collocazioni meno frequenti rispetto a quelle con alto *t*-score e i cui componenti non si trovano spesso singolarmente, al di fuori della collocazione stessa (si veda 4.2). Come nel caso dei bigrammi non frequenti/non attestati e dei bigrammi ad alto *t*-score, vengono presi in considerazione due testi per lingua, di nuovo uno la cui interpretazione è più collocazionale rispetto alla traduzione e uno la cui traduzione è più collocazionale dell'interpretazione.

5.3.3.1 Inglese

Testo 023

Il presente testo è caratterizzato da una traduzione maggiormente collocazionale rispetto all'interpretazione.

Anche in questo caso la collocazione “economic crisis”, estratta sia per l'interpretazione che per la traduzione, è stata esclusa dall'analisi in quanto stessa resa in entrambe le modalità di mediazione per lo stesso st (“crisi economica”).

Anche la collocazione “long periods” risulta estratta per entrambi i tt, ma si è deciso di analizzarla ugualmente perché nel tt-tr fa parte dell'espressione più estesa “long periods of time”, in linea con l'espressione utilizzata nei st, ossia “lunghi periodi di tempo” (1).

1 TT-IN-EN: through methods that have been specifically developed in these specific regions over **long periods**.

TT-TR-EN: through the use of methods which have been specifically developed in distinct regions over **long periods** of time.

ST-IN-IT: attraverso l'uso di metodi che sono stati specificamente sviluppati in diverse regioni e per **lunghi periodi** di tempo.

ST-TR-IT: attraverso l'uso di metodi che sono stati specificamente sviluppati in diverse regioni e per **lunghi periodi** di tempo.

Non è quindi il traduttore a espandere la collocazione “long periods”, ma è l'interprete a tagliare l'espressione del st, molto probabilmente per motivi di tempo, pur mantenendone il senso.

Le altre tre collocazioni che risultano rilevanti per l'analisi qualitativa del presente testo (“small-scale farming”, “organic farming” e “geographical regions”) sono invece state estratte unicamente per il testo tradotto, indice di una maggiore collocationalità del tradotto rispetto all'interpretato, come già anticipato all'inizio del paragrafo.

Come si può evincere dall'esempio 2, “small-scale farming” e “organic farming” appartengono allo stesso periodo.

2 TT-TR-EN: We need to promote not only the competitiveness of our products, but also traditional agriculture, **small-scale farming, organic farming** and local distribution.

TT-IN-EN: We need to encourage not just the competitiveness of our products, but also traditional agriculture, **biological agriculture** and local distribution networks.

ST-TR-IT: Abbiamo bisogno di promuovere non soltanto la competitività dei nostri prodotti, ma anche l'agricoltura tradizionale, i **piccoli allevamenti,**

l'agricoltura biologica e la distribuzione locale.

ST-IN-IT: Abbiamo bisogno di promuovere non soltanto la competitività dei nostri prodotti, ma anche l'agricoltura tradizionale, i **piccoli allevamenti**, **l'agricoltura biologica** e la distribuzione locale.

Confrontando il passaggio del testo tradotto con quello dell'interpretato si può avanzare l'ipotesi che l'interprete abbia tagliato la collocazione "small-scale farming" sia perché altamente settoriale sia a causa della velocità del discorso dell'oratore (158,2 parole al minuto). La collocazione "organic farming" corrisponde invece all'interpretato "biological agriculture". Quest'ultimo è più vicino al st ("agricoltura biologica"), ma risulta molto meno diffuso rispetto alla collocazione impiegata dal traduttore. Infatti, il corpus di riferimento fornisce 321 occorrenze per la collocazione "organic farming" contro le 2 ottenute per "biological agriculture". Si può dunque affermare che, forse di nuovo per mancanza di tempo, l'interprete abbia fatto un calco dell'espressione presente nel st.

Per quanto riguarda la collocazione "geographical regions", confrontando la traduzione con l'interpretazione il caso sembra essere lo stesso visto in precedenza, ossia semplificazione da parte dell'interprete. Infatti, i passaggi dei due st risultano essere perfettamente speculari (3).

3 TT-TR-EN: a valid contribution to food security because they often represent the most effective way of utilising land in diverse **geographical regions** of the EU through the use of methods

TT-IN-EN: a valid contribution to food security. Often, they represent the most

efficient way of using the land in different **regions** through methods

ST-TR-IT: un valido contributo alla sicurezza alimentare perché spesso rappresentano il modo più efficace di utilizzare terreni in diverse **regioni geografiche** dell'Unione europea attraverso l'uso di metodi

ST-IN-IT: un valido contributo alla sicurezza alimentare perché spesso rappresentano un modo il più efficace di utilizzare terreni in diverse **regioni geografiche** dell'Unione europea attraverso l'uso di metodi

Testo 004

In questo caso l'interpretato è più collocazionale rispetto al tradotto.

Anche in questo testo una delle collocazioni ("legislative proposals") è stata estratta sia per l'interpretazione che per la traduzione e si è dunque deciso di non analizzarla, a maggior ragione poiché deriva dalla stessa espressione utilizzata nei due st ("proposte legislative").

Per quanto riguarda le collocazioni "dangerous substances" e "hazardous substances", essendo queste presenti nello stesso frammento di testo elaborato nelle due modalità di mediazione, la differenza risiede nella scelta degli aggettivi da utilizzare (4).

4 TT-IN-EN: first of all as far as **dangerous substances** is concerned

TT-TR-EN: Firstly, with regard to the issue of **hazardous substances**

ST-IN-IT: in primo luogo, per quanto riguarda la questione inerente le **sostanze pericolose**

ST-TR-IT: in primo luogo, per quanto riguarda la questione inerente le **sostanze pericolose**

Seppure entrambi i bigrammi trasmettano lo stesso significato voluto dall'oratore, quello utilizzato dal traduttore ("hazardous substances") risulta più diffuso nella lingua inglese corrente rispetto alla collocazione estratta per l'interpretazione ("dangerous substances"), in quanto nel corpus di riferimento se ne trovano rispettivamente 249 e 112 occorrenze. Inoltre, "dangerous" è l'equivalente traduttivo più standard per l'aggettivo italiano "pericoloso" ed è quindi più facilmente accessibile all'interprete rispetto all'opzione scelta dal traduttore.

Ci sono poi diversi casi in cui l'interprete utilizza collocazioni non reperibili nel testo tradotto. In 5 è esemplificato un caso in cui l'interprete, con la collocazione "various forms", amplia il testo originale attraverso l'aggiunta dell'aggettivo "various", mentre il traduttore resta più aderente al relativo st, scegliendo di impiegare il sostantivo semplice "forms".

5 TT-IN-EN: but now we are moving things forward and making it easier for them to fill in the **various forms** that are involved and all of this through legislative simplification.

TT-TR-EN: with tools that help them rather than the exact opposite, that is to say, **forms** to fill out, bureaucratic procedures and extra obligations.

ST-IN-IT: finalmente la normativa comunitaria offrirà strumenti abilitanti

anziché l'esatto contrario, ossia mobili ehm **moduli** da compilare, trafile burocratiche e obblighi aggiuntivi.

ST-TR-IT: finalmente la normativa comunitaria offrirà strumenti abilitanti anziché l'esatto contrario, ossia **moduli** da compilare, trafile burocratiche e obblighi aggiuntivi.

Con l'aggiunta dell'aggettivo, l'interprete sembra voler sottolineare il filo conduttore che si sviluppa lungo tutto il testo, ossia la pesantezza e l'inutilità delle lungaggini burocratiche che devono sperimentare le aziende in tutta Europa.

La stessa connotazione negativa vista in 5 sembra essere condivisa dalla collocazione dell'interpretazione esemplificata in 6 ("excessive bureaucracy").

6 TT-IN-EN: We need to make certain that we don't oblige companies, particularly the very small ones, to undergo **excessive bureaucracy** ehm... or any unnecessary extra work and we've made this clear in our conclusion

TT-TR-EN: In particular, we will have to avoid subjecting enterprises - and particularly micro-enterprises - to **unjustified** or, in any case, unnecessary **administrative obligations**.

ST-IN-IT: In particolare, dovremo evitare di sottoporre le imprese, soprattutto ehm le più piccole, ad **obblighi amministrativi ingiustificati** o comunque non necessari.

ST-TR-IT: In particolare, dovremo evitare di sottoporre le imprese, soprattutto le più

piccole, ad **obblighi amministrativi**
ingiustificati o comunque non necessari.

Tale connotazione è trasmessa non solo dall'aggettivo "excessive", ma anche dal sostantivo "bureaucracy", come si può notare grazie alla definizione ripresa dal dizionario monolingue e riportata di seguito.

bureaucracy noun 1 [uncountable] (often disapproving). The system of official rules and ways of doing things that a government or an organization has, especially when these seem to be too complicated

unnecessary/excessive bureaucracy

We need to reduce paperwork and bureaucracy in the company.

Inoltre, confrontando il frammento di testo con il corrispondente del tradotto, si può osservare che si tratta di un esempio di condensazione, in quanto mentre l'interprete utilizza la collocazione analizzata, la traduzione e i due st contengono gruppi nominali più estesi.

La stessa connotazione negativa presente negli esempi precedenti ritorna con la collocazione "red tape" estratta per il testo interpretato (7).

7 TT-IN-EN: my final point is, that the need to cut **red tape** and tests is going to be important

TT-TR-EN: Lastly, various parts of the text recognise the need to reduce the **administrative** and testing **obligations** on enterprises

ST-IN-IT: Infine, in varie parti del testo si riconosce la necessità di ridurre gli **obblighi amministrativi** delle imprese e quelli legati ai testi

ST-TR-IT: Infine, in varie parti del testo si riconosce la necessità di ridurre gli **obblighi amministrativi** delle imprese e quelli legati ai testi

Anche in questo caso, così come visto nell'esempio sopra, l'interprete probabilmente ha scelto di utilizzare una strategia di condensazione per guadagnare tempo. Inoltre, come già osservato in precedenza, tra la collocazione utilizzata dall'interprete e la scelta effettuata dal traduttore si può notare uno scarto di registro, riconducibile ai diversi contesti di produzione: infatti, per l'interprete il pubblico è *in praesentia*, mentre il traduttore avverte maggiormente la distanza che lo separa dal pubblico e la traspone in un testo di registro più alto. Si può anche presumere che l'interprete utilizzi la collocazione in maniera automatica (si noti la distanza tra l'espressione nel st – “obblighi amministrativi” – e la collocazione stessa).

Nell'elaborare la collocazione in 8 (“successive phases”), l'interprete sembra essere influenzato dall'espressione del st (“fase successiva”).

8 TT-IN-EN: we want to continue to work closely with Parliament in **successive phases** of the adoption of all of this.

TT-TR-EN: I intend to continue to work closely with you, including with regard to the **next phase** in the adoption process.

ST-IN-IT: intendo continuare a collaborare da vicino anche per quel che riguarda la **fase successiva** della adozione

ST-TR-IT: intendo continuare a collaborare da vicino anche per quel che riguarda la **fase successiva** dell'adozione.

L'ipotesi di calco da parte dell'interprete è avvalorata dal fatto che l'aggettivo "successive" in inglese ha più il significato di "consecutivo" che di "seguinte", come si può osservare dalla definizione riportata sotto.

successive adjective. Following immediately one after the other

Synonym **consecutive**

This was their fourth successive win.

Successive governments have tried to tackle the problem.

There has been low rainfall for two successive years.

Si trova poi un caso ("exigent circumstances") in cui l'interprete, trovandosi probabilmente in un momento di difficoltà, si vede costretto a riformulare la frase del st, causando quindi una non corrispondenza tra la collocazione del suo testo e quelle impiegate nel st ("circostanze particolari") e nella traduzione ("special circumstances"), come si può vedere dall'esempio riportato in 9.

9 TT-IN-EN: Commission believes that unless... there are... **exigent circumstances** we really shouldn't be looking at any derogations at all.

TT-TR-EN: In this case, the Commission holds the view that there are no **special circumstances** that can justify the violation of this principle.

ST-IN-IT: Nel caso in questione, la Commissione ritiene che non sussistano **circostanze particolari** che possano giustificare una deroga a tale principio.

ST-TR-IT: Nel caso in questione, la Commissione ritiene che non sussistano **circostanze particolari** che possano giustificare una deroga a tale principio.

La collocazione estratta è inoltre di registro formale e ricalca la stessa espressione utilizzata negli Stati Uniti in linguaggio legale per indicare condizioni particolari che possono dare origine a permessi speciali per prendere determinati provvedimenti. Qui di seguito si riportano due definizioni per esemplificare il registro formale e l'utilizzo della collocazione in ambito legale.

Exigent circumstances are exceptions to the general requirement of a warrant under the Fourth Amendment searches and seizures. Exigent circumstances occur when a law enforcement officer has a probable cause and no sufficient time to secure a warrant.¹²

exigent adjective. formal. Pressing; demanding

the exigent demands of her contemporaries' music took a toll on her voice

Passando alle collocazioni proprie unicamente del testo tradotto (“exact opposite”, “sincere thanks” e “bureaucratic procedures”), la prima, esemplificata in 10, rispecchia fedelmente sia il relativo st che il st dell'interpretazione, dato che entrambi utilizzano il bigramma “esatto contrario”. La collocazione non risulta presente nel testo interpretato in quanto l'interprete riformula la frase.

10 TT-TR-EN: with tools that help them rather than the **exact opposite**, that is to say, forms to fill out

TT-IN-EN: but now we are moving things forward and making it easier for them to fill in the various forms that are involved

ST-TR-IT: finalmente la normativa comunitaria offrirà strumenti abilitanti

¹² http://www.law.cornell.edu/wex/exigent_circumstances

anziché l'**esatto contrario**, ossia moduli da compilare

ST-IN-IT: finalmente la normativa comunitaria offrirà strumenti abilitanti anziché l'**esatto contrario**, ossia mobili ehm moduli da compilare

L'omissione delle altre due collocazioni ("sincere thanks" e "bureaucratic procedures") da parte dell'interprete è dovuta molto probabilmente a una volontà di condensare il testo per motivi di tempo a disposizione (11-12).

11 TT-TR-EN: I also wish to express my **sincere thanks** to all the shadow rapporteurs

TT-IN-EN: I'd also like to **thank** all the shadow rapporteurs

ST-TR-IT: Voglio inoltre **ringraziare vivamente** tutti i relatori ombra

ST-IN-IT: voglio inoltre **ringraziare vivamente** tutti i relatori ombra

12 TT-TR-EN: that is to say, forms to fill out, **bureaucratic procedures** and extra obligations.

TT-IN-EN: making it easier for them to fill in the various forms that are involved and **all of this** through legislative simplification.

ST-TR-IT: ossia moduli da compilare, **trafile burocratiche** e obblighi aggiuntivi.

ST-IN-IT: ossia mobili ehm moduli da compilare, **trafile burocratiche** e obblighi aggiuntivi.

In entrambi i casi la riduzione porta inoltre alla perdita di alcuni elementi nel testo interpretato, ossia la connotazione positiva dell'aggettivo "sincere" in 11 e l'esemplificazione in 12.

5.3.3.2 Italiano

Testo 007

Il testo è maggiormente collocazionale nella traduzione che nell'interpretazione.

Come già in precedenza, alcune collocazioni ("Unione europea", "medie imprese", "libera circolazione" e "crisi economica") sono state estratte sia per l'interpretazione che per la traduzione come equivalenti traduttivi delle stesse espressioni utilizzate in entrambi i st ("EU", "medium-sized enterprises", "free movement" e "economic crisis"). Tali collocazioni non sono quindi state considerate nell'analisi qualitativa.

Nel caso in 1 l'interprete è più specifico del traduttore, aggiungendo l'aggettivo "burocratici" al sostantivo in questione, ossia "oneri".

1 TT-IN-IT: Ehm e la marcatura CE che ehm deve essere introdotta affinché li piccoli impresa affrontino gli stessi **oneri burocratici** delle grandi imprese..

TT-TR-IT: la dichiarazione di prestazione o marcatura CE, in merito alla quale si è raggiunto un accordo soddisfacente, volto a garantire che le microimprese non debbano andare incontro agli stessi **oneri** che spettano alle imprese più grandi.

ST-IN-EN: the declaration of performance or CE marking, where a successful agreement was

made to ensure micro-enterprises do not face the same **burdens** as larger companies.

ST-TR-EN: the declaration of performance or CE marking, where a successful agreement was made to ensure micro-enterprises do not face the same **burdens** as larger companies.

L'aggiunta sembra immotivata, dato che anche il st dell'interpretazione è vago al pari della traduzione e del relativo st. Molto probabilmente l'interprete sceglie di specificare per aiutare il pubblico a seguire meglio il filo del discorso, mentre il traduttore non ne ha bisogno, visto che, essendo il testo scritto, il lettore può tornare indietro e rileggere. Inoltre, questa maggiore specificità dell'interprete rispetto al traduttore e all'oratore potrebbe essere dovuta a una volontà di prendere tempo aspettando la continuazione del discorso da parte dell'oratore ed evitando di riempire il vuoto con pause piene ("ehm").

Per quanto riguarda le collocazioni estratte per la traduzione, ma non per l'interpretazione, nel caso di "sostanze pericolose" l'interprete, probabilmente in difficoltà, non rende un'intera parte di frase in cui dovrebbe trovarsi la collocazione in questione (2).

2 TT-TR-IT: agli stessi oneri che spettano alle imprese più grandi. Per quanto concerne l'esame e l'identificazione delle **sostanze pericolose** nei prodotti da costruzione, siamo riusciti a trovare un compromesso, evitando qualsiasi conflitto con le normative REACH già esistenti.

TT-IN-IT: gli stessi oneri burocratici delle grandi imprese... e dobbiamo cercare di ehm evitare conflitti con la legislazione in vigore.

ST-TR-EN: the same burdens as larger companies. On tackling and identifying **hazardous substances** in construction products, we managed to broker a compromise ensuring no conflict with existing REACH regulations.

ST-IN-EN: the same burdens as larger companies. And on tackling and identifying **hazardous substances** in construction products, we managed to broker a compromise ensuring no conflict with existing REACH regulations.

Un altro caso di omissione da parte dell'interprete è quello
esemplificato in 3.

3 TT-TR-IT: Spero che questo accordo in seconda lettura possa sortire **risultati concreti**, promuova la libera circolazione dei prodotti da costruzione, migliori gli standard in seno all'Unione europea e incentivi il riciclaggio

TT-IN-IT: Spero che nella seconda lettura si riesc- si riesca veramente a incoraggiare la libera circolazione dei prodotti, il miglioramento dei standard nell'Unione europea anche per i prodotti futuri

ST-TR-EN: I hope that this second-reading agreement will have **real results** and will encourage the free movement of construction products, improve standards across the EU, and promote the recycling

ST-IN-EN: I hope that this second-reading agreement will have **real results** and encourage the free movement of construction

products, improve standards across the EU,
and promote the recycling

Questa volta, a differenza dell'esempio precedente, l'omissione non compromette il senso del discorso. Infatti, il significato insito nella collocazione "risultati concreti" e nel corrispettivo "real results" presente in entrambi i st viene trasmesso dalla serie di esempi riportata sia dall'interprete che dal traduttore.

Gli esempi 4, 5, 6 e 7 presentano casi in cui il traduttore utilizza collocazioni estese mentre l'interprete adotta strategie di condensazione ("oneri burocratici" vs. "onere", "imprese produttrici" vs. "produttori", "importanza vitale" vs. "vitale" e "cittadini europei" vs. "cittadini").

4 TT-TR-IT: evitando, soprattutto, di tradursi di un aumento degli **oneri burocratici** a carico delle aziende o di creare confusione tra i consumatori.

TT-IN-IT: senza aumentare la l'**onere** sugli imprenditori, soprattutto i piccoli, o la confusione dei consumatori.

ST-TR-EN: whilst, most importantly, not increasing **bureaucratic burdens** on businesses or confusing consumers.

ST-IN-EN: whilst, most importantly, not increasing **bureaucratic burdens** on businesses or confusing consumers.

5 TT-TR-IT: circa il 92 per cento delle **imprese produttrici** di materiali da costruzione

TT-IN-IT: circa 92 per cento dei **produttori** dei materiali di s- costruzione

ST-TR-EN: approximately 92 percent of the **manufacturers** of construction materials

ST-IN-EN: approximately 92 percent of the **manufacturers** of construction materials

6 TT-TR-IT: Chiaramente, l'edilizia riveste un'**importanza vitale** per l'economia europea

TT-IN-IT: Quindi, il settore della costruzione edile è **vitale** per l'economia europea

ST-TR-EN: Clearly, the construction sector is **vital** for the European economy

ST-IN-EN: Clearly, the construction sector is **vital** for the European economy

7 TT-TR-IT: Attualmente, in Europa, il settore edile dà lavoro direttamente a 12 milioni di **cittadini europei;**

TT-IN-IT: In Europa, oggi, il settore della costruzione dà lavoro a 12 milioni di **cittadini;**

ST-TR-EN: In Europe today, the construction sector directly employs 12 million **EU citizens.**

ST-IN-EN: In Europe today, the construction sector employs directly 12 million **EU citizens.**

In realtà, nei casi in 5 e 6 il traduttore espande il st ("imprese produttrici" per "manufacturers", "riveste un'importanza vitale" per "is vital"), mentre l'interprete si mantiene più aderente al relativo st rendendo "manufacturers" con il termine "produttori" e "is vital" con "è vitale".

In un solo caso (8), l'interprete espande la collocazione rispetto al traduttore (mentre il traduttore utilizza la collocazione "settore edile", l'interprete rende con "settore della costruzione").

8 TT-TR-IT: Attualmente, in Europa, il **settore edile** dà lavoro direttamente a 12 milioni di cittadini europei;

TT-IN-IT: In Europa, oggi, il **settore della costruzione** dà lavoro a 12 milioni di cittadini;

ST-TR-EN: In Europe today, the **construction sector** directly employs 12 million EU citizens.

ST-IN-EN: In Europe today, the **construction sector** employs directly 12 million EU citizens.

L'interprete in questo caso resta più vicino al st, in cui l'oratore si riferisce al settore edile con l'espressione "construction sector". L'interprete utilizza questa strategia forse per non perdere tempo prezioso pensando alla collocazione migliore da inserire. Ricercando le due espressioni nel corpus di riferimento risulta che "settore edile" è molto più diffusa rispetto a "settore della costruzione" (rispettivamente 112 e 14 occorrenze). Effettuando una ricerca anche per "settore delle costruzioni" si osserva però che è questa l'espressione più diffusa con 169 occorrenze. È quindi probabile che l'interprete sia stato influenzato dal st, dove il nome "construction" è utilizzato al singolare.

Testo 026

Il testo è maggiormente collocazionale nell'interpretazione che nella traduzione.

Anche per questo testo ci sono collocazioni estratte sia per la traduzione che per l'interpretazione e derivate dallo stesso st ("EU"). Si tratta di tre occorrenze di "Unione europea", che quindi non sono state analizzate.

Nell'esempio 9 si può osservare come la traduzione contenga due occorrenze di "Unione europea", mentre nell'interpretazione la collocazione ricorre una sola volta. A prima vista sembra che sia il traduttore a ripetere la collocazione. In realtà, confrontando i tt con i relativi st, risulta che anche in questi ultimi sono presenti due occorrenze di "EU". È quindi l'interprete a ridurre il relativo st, forse in un momento di difficoltà, visto anche l'errore relativo alla cifra (0,06 viene reso in interpretazione con 0,6).

9 TT-IN-IT: Secondo, l'attuale accordo riguarda solamente una parte degli scambi commerciali con l'**Unione europea**, siamo solo allo 0,6 per cento.

TT-TR-IT: In secondo luogo, l'accordo attuale riguarda solo una piccola parte del commercio dell'**Unione europea**, poiché il commercio dell'intera regione del Pacifico rappresenta solo lo 0,06 per cento del commercio dell'**Unione europea**.

ST-IN-EN: Secondly, the current ehm agreement concerns only a small part of **EU** trade, since the share of **EU's** trade with the whole of the Pacific region is just 0.06 percent

ST-TR-EN: Secondly, the current agreement concerns only a small part of **EU** trade, since the share of **EU** trade accounted for by

the whole of the Pacific region is just 0.06 percent

Quattro delle collocazioni estratte per l'interpretazione (due occorrenze di “prodotti ittici”, “industria ittica” e “situazione attuale”) non sono utilizzate nella traduzione in quanto quest'ultima si serve di espressioni più estese e più aderenti al st dal punto di vista formale. Negli esempi 10 e 11, la collocazione “prodotti ittici”, scelta dall'interprete, risulta più diffusa rispetto all'espressione “prodotti della pesca” presente nel testo tradotto: nel corpus di riferimento se ne trovano rispettivamente 126 e 76 occorrenze. Tuttavia, la collocazione è meno aderente al st (“fishery products”) e meno corretta rispetto all'espressione impiegata dal traduttore, poiché il termine “prodotti ittici” include anche i prodotti di allevamento ittico.

10 TT-IN-IT: Innanzitutto la questione della deroga rispetto alle norme di origine per i **prodotti ittici**.

TT-TR-IT: soprattutto a causa della deroga alle norme di origine per i **prodotti della pesca**.

ST-IN-EN: most notably on the issue with the derogation from the r- rules of origin for **fishery products**.

ST-TR-EN: most notably on the issue of the derogation from the rules of origin for **fishery products**.

11 TT-IN-IT: potranno beneficiare di questa deroga e potenzialmente esportare maggiori quantità di **prodotti ittici** lavorati nell'Unione europea, potenzialmente danneggiando gli interessi della nostra industria.

TT-TR-IT: possano sfruttare a proprio vantaggio questa deroga per esportare nell'Unione europea grandi quantità di **prodotti della pesca** trasformati, con il rischio di nuocere agli interessi dell'industria europea di trasformazione e conservazione del pesce.

ST-IN-EN: will be able to take advantage of this derogation and potentially export large quantities of processed **fisheries... products** to the EU, potentially harming the interests of the EU fish processing and canning industry.

ST-TR-EN: will be able to take advantage of this derogation and potentially export large quantities of processed **fisheries products** to the EU, potentially harming the interests of the EU fish processing and canning industry.

Negli esempi 12 e 13, invece, l'interprete decide di ridurre il st ("fishing processing and canning industry" e "The agreement, as it is now") utilizzando "industria ittica" e "situazione attuale", forse per motivi di tempo, togliendo qualche sfumatura di significato a ciò che l'oratore voleva trasmettere. Il traduttore dal canto suo resta più fedele al st rendendo le due espressioni con "industria [...] di trasformazione e di conservazione del pesce" e "L'accordo, nella sua forma attuale".

12 TT-IN-IT: visto che c'è un effetto destabilizzante che è stato dimostrato sulla nostra **industria ittica** e quindi chiediamo la sospensione della deroga relativa alle norme di origine in questo accordo.

TT-TR-IT: qualora la valutazione dimostrasse l'effetto destabilizzante sull'**industria europea di trasformazione e di conservazione del pesce** di avviare la procedura per sospendere gli accordi eccezionali concernenti le norme di origine, contenuti nell'accordo.

ST-IN-EN: in the event that the assessment report demonstrates a destabilising effect on the unfinished on the EU's **fishing processing and canning industry**, to initiate the procedure leading to the suspension of the exceptional arrangements we regard in the rules of origin in this agreement.

ST-TR-EN: in the event that the assessment demonstrates a destabilising effect on the EU's **fish processing and canning industry**, for it to initiate the procedure leading to suspension of the exceptional arrangements in this agreement regarding rules of origin.

13 TT-IN-IT: La **situazione attuale** non è ideale: ehm questo infatti è un accordo solo interinale.

TT-TR-IT: L'**accordo, nella sua forma attuale**, non è perfetto, non da ultimo perché non si tratta di un APE completo, ma solo di un accordo interinale.

ST-IN-EN: The **agreement, as it is now**, is not a perfect one, not least because it is not a comprehensive EPA, but only an interim one.

ST-TR-EN: The **agreement, as it is now**, is not a perfect one, not least because it is

not a comprehensive EPA, but only an interim one.

Nel caso in 14, l'interprete utilizza la collocazione "scambi commerciali" mentre il traduttore decide di utilizzare un termine semplice ("commercio"), restando aderente al st ("trade").

14 TT-IN-IT: Secondo, l'attuale accordo riguarda solamente una parte degli **scambi commerciali** con l'Unione europea, siamo solo allo 0,6 per cento.

TT-TR-IT: In secondo luogo, l'accordo attuale riguarda solo una piccola parte del **commercio** dell'Unione europea, poiché il commercio dell'intera regione del Pacifico rappresenta solo lo 0,06 per cento del commercio dell'Unione europea.

ST-IN-EN: Secondly, the current ehm agreement concerns only a small part of EU **trade**, since the share of EU's trade with the whole of the Pacific region is just 0.06 percent

ST-TR-EN: Secondly, the current agreement concerns only a small part of EU **trade**, since the share of EU trade accounted for by the whole of the Pacific region is just 0.06 percent

Molto probabilmente, come nel caso della collocazione "red tape" analizzata in precedenza (si veda 5.3.3.1), si tratta di un automatismo dell'interprete, di una collocazione che ricorda e impiega facilmente perché parte di un lessico familiare. Inoltre, anche in questo caso, l'interprete potrebbe aver deciso di impiegare una collocazione più esplicita rispetto al st e alla resa del traduttore per prendere tempo in attesa di ascoltare il resto del discorso dell'oratore.

L'unica collocazione estratta per la traduzione ma non per l'interpretazione è "grandi quantità" (15).

15 TT-TR-IT: per esportare nell'Unione europea **grandi quantità** di prodotti della pesca trasformati

TT-IN-IT: potenzialmente esportare **maggiori quantità** di prodotti ittici lavorati nell'Unione europea

ST-TR-EN: potentially export **large quantities** of processed fisheries products to the EU

ST-IN-EN: potentially export **large quantities** of processed fisheries... products to the EU

Come si può notare dall'esempio, anche l'interprete utilizza un bigramma: l'aggettivo impiegato, ossia "maggiori", denota però una sfumatura di significato diversa sia rispetto all'aggettivo "grandi" della traduzione sia rispetto all'aggettivo "large" presente in entrambi i st.

Per il parametro alto MI, si può osservare una maggiore tendenza all'omissione da parte degli interpreti e una maggiore tendenza all'esplicitazione da parte dei traduttori, sebbene quest'ultimo fenomeno non sia assente nei testi prodotti dagli interpreti.

Il prossimo paragrafo riassumerà le tendenze osservate nel corso dell'analisi qualitativa, proponendone una classificazione.

5.3.4 Riassunto risultati analisi qualitativa

La tabella 5.10 riporta i risultati dell'analisi qualitativa classificati a seconda delle tendenze osservate e che hanno portato a una differenze di collocazionalità tra tradotto e interpretato. La

categorizzazione è stata ripresa da Bernardini (2007:97) e adattata alla prospettiva intermodale applicata all'analisi qualitativa oggetto del presente capitolo.

Testo	Lingua target	Tendenza	Fq nel testo	Fq totale	Esempi	
048	EN	Altro significato	2	7	TT-IN: <u>mere politicking</u> TT-IN: <u>communal plant</u>	TT-TR: simply to stand up for my region TT-TR: municipal [...] facilities
		Omissione	1	16	TT-TR: <u>underused areas</u>	TT-IN: //
		Preferenza lessicale	1	5	TT-TR: <u>European veto</u>	TT-IN: Community ban
034	EN	Straniamento	1	4	TT-IN: <u>favourable context</u>	TT-TR: favourable environment
		Maggiore informalità	2	6	TT-IN: <u>favourable context</u> TT-IN: <u>broad palette</u>	TT-TR: environment that is conducive TT-TR: diversification
		Minore appropriatezza	2	3	TT-IN: <u>young businessmen</u> TT-IN: <u>one-man companies</u>	TT-TR: young entrepreneurs TT-TR: micro [...] enterprises
		Condensazione	1	5	TT-IN: <u>consolidated framework</u>	TT-TR: consolidated political framework
		Maggiore esplicitzza	1	17	TT-TR: <u>consolidated framework</u>	TT-IN: framework

		Maggiore vaghezza	1	5	TT-TR: <u>same enterprises</u>	TT-IN: SMEs
		Omissione	1	16	TT-TR: <u>repayable aid</u>	TT-IN: //
		Maggiore precisione	1	4	TT-TR: <u>electronic observatory</u>	TT-IN: other ways
017	IT	Maggiore vaghezza	1	5	TT-IN: <u>legname illegale</u>	TT-TR: prodotti di legno illegale
		Condensazione	1	5	TT-TR: <u>legname illegale</u>	TT-IN: legname che è abbattuto [...] illegalmente
		Omissione	1	16	TT-TR: <u>legname tropicale</u>	TT-IN: //
		Maggiore precisione	1	4	TT-TR: <u>legname tropicale</u>	TT-IN: questo tipo di legname
		Altro significato	1	7	TT-TR: <u>base vitale</u>	TT-IN: ospitano
		Straniamento	1	4	TT-TR: <u>gestione accorta</u>	TT-IN: attenta gestione
006	IT	Maggiore vaghezza	1	5	TT-IN: <u>buoni steccati</u>	TT-TR: <u>buone recinzioni</u>
		Altro significato	1	7	TT-IN: <u>piccole corporazioni</u>	TT-TR: imprese più piccole
		Preferenza lessicale	1	5	TT-IN: <u>regolamentazione migliore</u>	TT-TR: <u>normativa migliore</u>
		Straniamento	1	4	TT-IN: <u>condizioni separate</u>	TT-TR: condizioni speciali
		Omissione	2	16	TT-TR: <u>signor Haerynck</u> TT-TR: <u>diverse correlazioni</u>	TT-IN: // TT-IN: //
041	EN	Maggiore vaghezza	1	5	TT-IN: <u>other options</u>	TT-TR: virtuous alternatives
		Omissione	3	16	TT-TR: <u>European</u>	TT-IN: //

					<u>legislation</u> TT-TR: <u>renewable</u> <u>energy</u> TT-TR: <u>environmental</u> <u>impact</u>	TT-IN: // TT-IN: //
		Preferenza lessicale	1	5	TT-TR: <u>European</u> <u>legislation</u>	TT-IN: Community provisions
		Differenza sintattica	1	2	TT-TR: <u>public</u> <u>health</u> [...] <u>impact</u>	TT-IN: work against citizen's health
22	EN	Preferenza lessicale	1	5	TT-IN: <u>human</u> <u>rights</u> law	TT-TR: humanitarian law
		Maggiore informalità	2	6	TT-IN: <u>international</u> <u>level</u> TT-IN: <u>same</u> <u>time</u>	TT-TR: worldwide TT-TR: taking into account
		Addomesticazione	1	4	TT-TR: <u>dramatic</u> <u>increase</u>	TT-IN: exponential increase
061	IT	Differenza sintattica	1	2	TT-IN: diamo <u>grande</u> <u>attenzione</u>	TT-TR: siamo particolarmente attenti
		Straniamento	1	4	TT-IN: <u>caso</u> <u>concreto</u>	TT-TR: <u>caso</u> <u>specifico</u>
		Addomesticazione	1	4	TT-IN: <u>attuale</u> <u>fase</u>	TT-TR: fase attuale
		Maggiore esplicitzza	3	17	TT-TR: <u>sistema</u> <u>giudiziario</u> TT-TR: <u>prossime</u> <u>elezioni</u> TT-TR: <u>Parlamento</u> <u>europeo</u>	TT-IN: giudiziario TT-IN: elezione TT-IN: Parlamento
		Maggiore precisione	1	4	TT-TR: <u>sviluppo</u> <u>tecnologico</u>	TT-IN: punto di vista tecnologico
		Omissione	1	16	TT-TR: <u>Unione</u> <u>europea</u>	TT-IN: //

		Minore appropriatezza	1	3	TT-TR: <u>fondamentalismo islamico</u>	TT-IN: radicalismo islamico
034	IT	Maggiore esplicitezza	4	17	TT-IN: <u>maggior parte</u> TT-IN: <u>certi casi</u> TT-IN: <u>genitori adottivi</u> TT-IN: <u>genitori adottivi</u>	TT-TR: molti TT-TR: molti TT-TR: adozione TT-TR: adottanti
		Omissione	2	16	TT-IN: <u>stragrande maggioranza</u> TT-TR: <u>buona fede</u>	TT-TR: // TT-IN: //
		Maggiore informalità	1	6	TT-IN: <u>giusto equilibrio</u>	TT-TR: delicato equilibrio
023	EN	Maggiore vaghezza	1	5	TT-IN: <u>long periods</u>	TT-TR: <u>long periods of time</u>
		Omissione	1	16	TT-TR: <u>small-scale farming</u>	TT-IN: //
		Addomesticazione	1	4	TT-TR: <u>organic farming</u>	TT-IN: biological agriculture
		Maggiore esplicitezza	1	17	TT-TR: <u>geographical regions</u>	TT-IN: regions
004	EN	Preferenza lessicale	1	5	TT-IN: <u>dangerous substances</u>	TT-TR: <u>hazardous substances</u>
		Maggiore esplicitezza	2	17	TT-IN: <u>various forms</u> TT-TR: <u>sincere thanks</u>	TT-TR: forms TT-IN: thank
		Condensazione	1	5	TT-IN: <u>excessive bureaucracy</u>	TT-TR: unjustified administrative obligations
		Maggiore informalità	1	6	TT-IN: <u>red tape</u>	TT-TR: administrative obligations

		Altro significato	1	7	TT-IN: <u>successive phases</u>	TT-TR: next phase
		Maggiore formalità	1	1	TT-IN: <u>exigent circumstances</u>	TT-TR: special circumstances
		Omissione	1	16	TT-TR: <u>exact opposite</u>	TT-IN: //
		Maggiore precisione	1	4	TT-TR: <u>bureaucratic procedures</u>	TT-IN: all of this
007	IT	Maggiore esplicitezza	5	17	TT-IN: <u>oneri burocratici</u> TT-TR: <u>oneri burocratici</u> TT-TR: <u>imprese produttrici</u> TT-TR: <u>importanza vitale</u> TT-TR: <u>cittadini europei</u>	TT-TR: oneri TT-IN: onere TT-IN: produttori TT-IN: vitale TT-IN: cittadini
		Omissione	2	16	TT-TR: <u>sostanze pericolose</u> TT-TR: <u>risultati concreti</u>	TT-IN: // TT-IN: //
		Addomesticazione	1	4	TT-TR: <u>settore edile</u>	TT-IN: settore della costruzione
026		Omissione	1	16	TT-TR: <u>Unione europea</u>	TT-IN: //
		Altro significato (sfumatura)	2	7	TT-IN: <u>prodotti ittici</u>	TT-TR: prodotti della pesca
		Altro significato			TT-TR: <u>grandi quantità</u>	TT-IN: maggiori quantità
		Condensazione	2	5	TT-IN: <u>industria ittica</u> TT-IN:	TT-TR: industria [...] di trasformazione e di conservazione del pesce

					<u>situazione attuale</u>	TT-TR: L'accordo, nella sua forma attuale
		Maggiore esplicitzza	1	17	TT-IN: <u>scambi commerciali</u>	TT-TR: commercio

Tabella 5.10 - Tendenze prospettiva intermodale

Come si può notare dalla tabella 5.10, le tendenze maggiormente diffuse nei testi analizzati sono la maggiore esplicitzza e l'omissione, che presentano rispettivamente 17 e 16 casi.

Una tendenza alla maggiore esplicitzza dovrebbe intuitivamente essere più diffusa nella traduzione piuttosto che nell'interpretazione, dato che l'interprete tende a condensare rispetto al st per non perdere tempo prezioso. In realtà, sebbene nel caso in esame la maggiore esplicitzza sia più diffusa nel tradotto che nell'interpretato, si possono osservare molti casi di esplicitazione da parte dell'interprete che probabilmente, come affermato in precedenza, si serve della "strategia" per prendere tempo in attesa di sentire la continuazione del discorso dell'oratore. Un'altra causa alla base di tale tendenza da parte dell'interprete potrebbe essere il fatto di restare più aderente al st rispetto al traduttore che tende ad allontanarsene per non commettere calchi.

La tendenza all'omissione è caratteristica quasi esclusiva degli interpreti ed è distribuita omogeneamente essendo presente nella maggior parte dei testi. Diverse sono le ragioni che spingono un interprete a tagliare parti del discorso: nella maggior parte dei casi, l'interprete si trova in difficoltà, a causa dell'impiego di termini troppo specifici, di alcune caratteristiche nel parlato dell'oratore, come ad esempio un particolare accento, o di st

contenenti costruzioni complesse e intricate, poiché, sebbene siano testi orali, nascono come testi scritti e vengono letti come tali. In alcuni casi le omissioni corrispondono a riformulazioni del testo da parte dell'interprete, per cui la traduzione e l'interpretazione si trovano a non combaciare più. Altre volte, l'interprete si trova a omettere per ridurre il testo originale, con o senza conseguenze per il risultato finale: in certe occasioni, infatti, l'omissione per riduzione porta a perdite di significato sia rispetto al st che rispetto al tradotto.

Anche la tendenza denominata "altro significato" è abbastanza diffusa (7 casi) e corrisponde a calchi commessi dall'interprete oppure riduzioni del st da parte dello stesso, con conseguente produzione di significati che il st non trasmette. Talvolta l'interprete è anche portato a una sorta di esplicitazione del st aggiungendo una particolare sfumatura al senso originariamente trasmesso dall'oratore (come nel caso di "maggiori quantità" per "large quantities").

Le restanti tendenze individuate per i testi oggetto della presente analisi sono diffuse in maniera abbastanza omogenea e spesso una tendenza specifica di una modalità di mediazione è la controparte di una tendenza riscontrata nell'altra modalità.

È il caso di addomesticazione e straniamento, la prima specifica della traduzione, mentre il secondo si riscontra maggiormente nell'interpretazione. Infatti, il traduttore risulta avere un comportamento più addomesticante in quanto prova ad allontanarsi maggiormente dal st, mentre l'interprete, essendo più aderente al st o addirittura formando dei calchi della lingua di partenza, risulta anche maggiormente straniante. Ci sono tuttavia delle eccezioni: in alcuni casi il traduttore si allontana sì dal st, ma la sua scelta risulta meno diffusa nella lingua di arrivo rispetto a

quella dell'interprete, nonostante quest'ultima sia più legata al st o alla lingua di partenza. A volte, invece, traduttore e interprete impiegano espressioni o collocazioni la cui differenza di diffusione nella lingua corrente non risulta estremamente marcata: in questi casi si è preferito non parlare di addomesticazione o straniamento, in quanto tali scelte non risultano completamente errate o corrette. Per indicare tali fenomeni state infatti adottate le categorie "preferenza lessicale" e "differenza sintattica".

Anche la tendenza a una maggiore precisione costituisce la controparte di un'altra tendenza, quella a una maggiore vaghezza. Il tradotto è generalmente più preciso rispetto all'interpretato, in quanto il traduttore dispone di più tempo per pensare alla resa migliore per tutte le sfumature di senso e tutti gli elementi contenuti nel st. Inoltre, nel caso di termini specifici o di citazioni il traduttore, a differenza dell'interprete, ha la possibilità di documentarsi. L'unico caso in cui il traduttore risulta più vago dell'interprete ("same enterprises" vs. "SMEs") è dovuto alla volontà del traduttore di non ripetere il termine specifico SMEs utilizzato poco prima nel testo.

Si riscontrano poi casi di maggiore formalità per il tradotto e maggiore informalità per l'interpretato, sempre da un punto di vista intermodale. Tali categorie sono legate al contesto di produzione in cui operano le due modalità di mediazione. Infatti, l'interprete tende a essere meno formale perché più vicino fisicamente al pubblico, mentre la formalità tipica dello scritto riflette la distanza che separa il traduttore dai lettori.

Per quanto riguarda la tendenza alla condensazione, ne sono stati individuati 4 casi nell'interpretazione, dovuti probabilmente alla volontà degli interpreti di risparmiare tempo nell'atto di mediazione. In un caso, invece, è il traduttore a

condensare rispetto alla resa contenuta nel testo interpretato (“legname illegale” vs. “legname che è abbattuto [...] illegalmente”). In questo caso il traduttore si serve della condensazione per allontanarsi dal relativo st (come si è visto nel corso dell’analisi qualitativa, i testi tradotti sono più inclini ad allontanarsi dalla forma del st).

Sono stati inoltre osservati 3 casi di minore appropriatezza: due di questi sono attribuibili all’interprete che non ha tempo a disposizione per pensare a collocazione neutre in termini di genere (“young businessmen” e “one-man companies”). Si può tuttavia notare come in un caso tale tendenza sia adottata dal traduttore in maniera inconscia: esso infatti utilizza un’espressione più comune, che però ha connotazione negativa (“fondamentalismo islamico”), mentre l’interprete, restando più fedele al st, impiega un’espressione più neutra e più corretta.

In conclusione, si può affermare che, grazie alla presente analisi, sono state individuate strategie tipiche di una modalità o dell’altra di mediazione, tuttavia non senza eccezioni. Data la varietà di testi e oratori non è possibile fare generalizzazioni universalmente valide. È dunque auspicabile che lo studio venga ampliato in futuro per confermare o rifiutare le tipicità e le diversità riscontrate attraverso la presente analisi.

Nel prossimo capitolo si passerà a una discussione dei risultati dell’analisi quantitativa e dei risultati derivati dall’analisi qualitativa alla luce anche degli shift osservati rispetto ai st. Verrà poi fatto un confronto con gli studi simili effettuati in passato. Infine si proporrà una panoramica sui possibili studi da portare avanti in futuro attraverso il corpus EPTIC_01_2011.

6 Discussione

6.1 Introduzione al capitolo

Il presente capitolo ha come obiettivo la discussione dei risultati delle analisi presentate nel capitolo 5, proponendo una classificazione dei bigrammi e delle collocazioni analizzati qualitativamente in base agli shift osservati confrontandoli con i relativi st (6.2 e 6.3). Nella seconda parte del capitolo verranno invece presentati gli eventuali studi da portare avanti in futuro sul corpus EPTIC_01_2011 o su un ampliamento dello stesso (6.4).

6.2 Discussione analisi quantitativa

Come già descritto in dettaglio in 5.2, l'analisi quantitativa intermodale condotta sulle due lingue di lavoro per i bigrammi non frequenti/non attestati, per le collocazioni ad alto *t*-score e per quelle ad alto MI non ha restituito risultati significativi. Tali risultati evidenziano quindi un'assenza di differenze significative nell'utilizzo di bigrammi non frequenti/non attestati da parte di interpreti e traduttori. Allo stesso modo, l'interpretazione non risulta essere significativamente più collocazionale della traduzione e viceversa.

Tali risultati possono essere messi a confronto con lo studio effettuato da Ferraresi e Miličević (submitted), che tuttavia si proponeva di investigare non solo la prospettiva intermodale (fraseologia utilizzata da interpreti e traduttori), ma anche le eventuali differenze tra testi paragonabili (st della traduzione vs. st dell'interpretazione e tt interpretato vs. tt tradotto). Per il confronto di cui sotto verranno quindi presi in considerazione unicamente i risultati relativi all'indagine intermodale.

A differenza di quanto osservato nell'analisi quantitativa oggetto della presente tesi, lo studio di Ferraresi e Miličević (submitted) rivela una maggiore collocazionalità del tradotto rispetto all'interpretato e un maggiore utilizzo da parte di quest'ultima di bigrammi non frequenti/non attestati. Per quanto riguarda il parametro *t*-score, invece, come nel caso presentato in questo lavoro, Ferraresi e Miličević (submitted) non raggiungono risultati significativi.

Ragionando sulle cause che possono essere all'origine delle discrepanze nei risultati di tale analisi, si può concludere che la non corrispondenza tra i risultati ottenuti in questa tesi e quelli ottenuti da Ferraresi e Miličević (submitted) non può essere dovuta alle lingue prese in esame. Infatti, entrambi gli studi sono basati sulla combinazione italiano-inglese. Le differenze potrebbero quindi essere dovute al numero modesto di testi raccolti per questo studio sperimentale o forse alle diversità che intercorrono tra tali testi (per esempio in termini di lunghezza, di temi, ecc.).

6.3 Discussione analisi qualitativa

La tabella 6.1 riportata sotto riprende nuovamente la classificazione elaborata da Bernardini (2007) e la applica questa volta agli shift che possono essere osservati tra interpretato e relativo st e tra tradotto e relativo st e che possono essere alla base delle tendenze individuate nell'analisi qualitativa, come riportata nel capitolo 6.

Shift	Totale casi	Casi per		Esempio
		IN	TR	
Altro significato	3	3	/	TT-IN: prodotti ittici ST-IN: fishery products

Calco	3	3	/	TT-IN: piccole corporazioni ST-IN: smaller corporations
Maggiore esplicitezza	8	4	4	TT-TR: a sum from the Fund for <u>Underused Areas</u> (FAS) ST-TR: una quota dei FAS
Straniamento	5	4	1	TT-IN: condizioni separate ST-IN: special conditions
Maggiore informalità	2	2	/	TT-IN: red tape ST-IN: obblighi amministrativi
Condensazione	4	3	1	TT-TR: legname illegale ST-TR: timber that is illegally harvested
Maggiore vaghezza	4	3	1	TT-TR: same enterprises ST-TR: piccole e medie imprese
Minore appropriatezza	1	/	1	TT-TR: fondamentalismo islamico ST-TR: radical Islamism

Tabella 6.1 - Shift dimensione parallela

Nonostante gli shift riportati in tabella e le tendenze individuate nel corso dell'analisi qualitativa presentino pochi casi per ognuno di essi, è comunque possibile confrontare i risultati ottenuti nella presente tesi con quelli raggiunti in altri lavori.

Gli shift in tabella 6.1 sono generalmente in linea con le tendenze individuate nei comportamenti di interpreti e traduttori.

I calchi commessi dagli interpreti, così come la loro tendenza a ricorrere, sebbene inconsciamente, alla “strategia” dello straniamento, sono sintomo di maggiore aderenza al st (Shlesinger e Ordan, 2012), a differenza del testo tradotto che è solito allontanarsene. Tuttavia si può notare come anche in traduzione sia presente un caso di straniamento: infatti, come già accennato in precedenza, alcuni traduttori, per allontanarsi a tutti i

costi dal st, rischiano di produrre espressioni meno diffuse nella lingua di arrivo. La maggiore informalità, invece, non dovrebbe comparire tra gli shift del testo interpretato rispetto al st. In realtà, i testi pronunciati dagli oratori delle istituzioni europee nascono come testi scritti e spesso vengono letti in aula. Di conseguenza, gli st presi in considerazione spesso presentano le costruzioni tipiche dei testi scritti e naturalmente vengono resi più semplici e più informali dagli interpreti. Nel confronto intermodale, invece, la maggiore formalità è stata attribuita al diverso contesto di produzione in cui si trovano interpreti e traduttori, con l'interprete a diretto contatto con il pubblico e il traduttore lontano da esso.

I casi di maggiore condensazione e maggiore vaghezza rispetto al st, ma anche rispetto al tradotto, sono in linea con quanto affermato in Bernardini et al. (accepted), nel quale viene individuata e confermata una tendenza alla semplificazione da parte dell'interprete. Infatti, il testo interpretato risulta più vago o contiene un numero minore di informazioni quando l'interprete non ha tempo a disposizione per elaborare esattamente ciò che viene detto nel st. Quando la vaghezza e la condensazione vengono utilizzate dal traduttore è perché quest'ultimo vuole allontanarsi il più possibile dalla forma del st.

La stessa motivazione, ovvero la volontà di distanziarsi dall'originale è alla base dell'esplicitezza da parte del traduttore. Come si può notare dai risultati riportati in tabella e dall'analisi qualitativa del capitolo precedente, anche l'interprete talvolta tende a esplicitare. La causa è però diversa: infatti, l'interprete lo fa per prendere tempo attendendo che l'oratore del st continui il proprio discorso o a causa di automatismi che inserisce nella sua produzione in maniera spontanea.

Dai dati riportati nella tabella 6.1 risulta inoltre che l'allontanamento del traduttore dal st non sempre porta a risultati vincenti, in quanto si possono incontrare casi di minore appropriatezza di una determinata espressione rispetto a un'altra che viene erroneamente ritenuta equivalente.

Anche nel confronto parallelo tra tt e st, così come nel confronto intermodale sono stati rinvenuti casi di cambiamento di significato, per cui tuttavia non è possibile dare una spiegazione universale. Come visto nel capitolo precedente, infatti, un diverso significato può essere causato da maggiore aderenza al st o a riformulazioni.

Si può concludere che lo studio oggetto della presente tesi, per quanto modesto e sperimentale, atto a un'indagine esplorativa del corpus intermodale EPTIC_01_2011, porta a confermare le ipotesi avanzate dagli studi sull'intermodalità sviluppati in passato. È comunque auspicabile un ampliamento del corpus per condurre ulteriori analisi e approfondire i risultati ottenuti fino ad oggi.

6.4 Prospettive future

Come affermato più volte nel corso della presente tesi, nonostante i risultati raggiunti grazie al presente studio, il progetto può essere ulteriormente sviluppato su diversi fronti.

Innanzitutto è auspicabile un ampliamento del corpus EPTIC_01_2011, non solo in termini di dimensioni, con l'aggiunta di nuovi testi, ma anche in termini di lingue su cui poter condurre lo stesso studio o studi simili. Oltre all'analisi intermodale applicata alla presente tesi, si potranno effettuare analisi sia comparabili che parallele.

Come prima sperimentazione, sono stati ricercati i pattern nome+aggettivo e aggettivo+nome. Lo stesso metodo potrà comunque essere replicato al fine di studiare l'impiego della fraseologia sotto forma di altri pattern e sviluppare nuove ipotesi sulle strategie di mediazione e di costruzione dei tt.

Un ulteriore sviluppo potrebbe consistere nell'analisi dei comportamenti di interpreti e traduttori da un punto di vista direzionale, ossia osservare se ci sono differenze nelle strategie utilizzate in una o nell'altra modalità di mediazione in una specifica lingua di arrivo.

Infine, sarebbe utile e produttivo utilizzare le informazioni fornite dai metadati per filtrare i testi in modo più accurato di modo da poter portare avanti analisi su materiale simile e riuscire quindi a individuare norme comuni a testi con caratteristiche simili (testi caratterizzati da stessa lunghezza, stessa velocità di esposizione, ecc.). In questo senso, è stata effettuata una breve analisi delle strategie utilizzate dai diversi interpreti. L'analisi è stata possibile solamente per la lingua inglese, in cui alcuni testi erano comuni a uno stesso interprete, in quanto i tt italiani sono stati tutti prodotti da interpreti diversi. I tt in lingua inglese prodotti dallo stesso interprete sono lo 048 e lo 041 da una parte e lo 034, lo 022 e lo 023 dall'altra. I primi due testi hanno in comune il fatto che l'interprete tende a omettere, forse a causa delle caratteristiche dell'oratore o del st, e a restare fedele al st attraverso calchi o espressioni e costruzioni stranianti. Gli altri tre testi presentano molti altri tipi di shift rispetto ai due testi precedenti, avendo tuttavia in comune con questi ultimi lo shift "straniamento" e, in due testi su tre, l'omissione. Non si può quindi parlare di strategie proprie a un determinato interprete, quanto di strategie comuni alla modalità interpretata da applicare a seconda del st e delle

caratteristiche dell'oratore. Naturalmente, tale analisi è da approfondire in futuro su un maggior numero di testi.

7 Conclusione

La tesi qui presentata è stata focalizzata su uno studio del corpus EPTIC_01_2011 per reperire le differenze nell'utilizzo della fraseologia da parte di traduttori e interpreti.

Nel capitolo 1 è stato fornito un background sui principali studi e le principali teorie riguardo alla linguistica dei corpora e alle collocazioni. Dopo aver proposto una definizione di corpus, ne sono state presentate le varie tipologie, soffermandosi in particolare sul corpus intermodale, ossia il tipo di corpus più recente nonché quello utilizzato per lo studio in questione. La seconda parte si è concentrata sulle collocazioni, offrendone le principali descrizioni proposte da alcuni importanti linguisti. Infine, sono stati proposti alcuni casi in cui la ricerca e lo studio delle collocazioni si sono avvalsi della linguistica dei corpora.

Nel capitolo 2 è stato introdotto EPTIC_01_2011 e ne sono state descritte forma e struttura. Il capitolo ha riservato particolare attenzione al processo di preparazione dei testi inclusi nella versione finale del corpus, riportando dettagliatamente le convenzioni di editing applicate e analizzando i metadati che introducono i testi.

Nel capitolo 3 è stata presentata la metodologia, ripresa e adattata da Durrant e Schmitt (2009), impiegata per la ricerca delle collocazioni e per un'analisi quantitativa e qualitativa delle stesse.

Nel capitolo 4 sono stati riportati i risultati delle suddette analisi con il supporto di esempi e grafici. Un'analisi qualitativa approfondita ha fatto luce sui fenomeni osservati per alcuni

bigrammi contigui aggettivo+nome/nome+aggettivo estratti grazie alla metodologia presentata nel capitolo 3.

Nel capitolo 5, infine, sono stati riassunti e classificati i fenomeni osservati nel capitolo 4 e sono state discusse le possibili cause di tali fenomeni. Nell'ultima parte del capitolo sono state proposte idee per future analisi basate su EPTIC_01_2011 nonché per un suo ampliamento.

Bibliografia e sitografia

- Altenberg, B. (1998). "On the phraseology of spoken English: the evidence of recurrent word-combinations". In A. P. Cowie (1998). 101-122.
- Bahns, J. (1993). "Lexical collocations: a contrastive view". *ELT Journal*, 47 (1): 56-63.
- Baker, M. (1993). "Corpus linguistics and translation studies: implications and applications". In M. Baker, G. Francis e E. Tognini-Bonelli (1993). 233-250.
- Baker, M. (1995). "Corpora in translation studies. An overview and some suggestions for future research". *Target*, 7 (2): 223-243.
- Baker, M., Francis, G. e E. Tognini-Bonelli a cura di (1993). *Text and technology: in honour of John Sinclair*. Amsterdam-Philadelphia: John Benjamins.
- Baroni, M., Bernardini, S., Ferraresi, A. e E. Zanchetta (2009). "The WaCky Wide Web: a collection of very large linguistically processed web-crawled corpora". *Language resources and evaluation*, 43 (3): 209-226.
- Bendazzoli, C. (2010). *Corpora e interpretazione simultanea*. Bologna: Asterisco.
- Bendazzoli, C. e A. Sandrelli (2009). "Corpus-based Interpreting Studies: early work and future prospects". *Revista tradumàtica*, 7: 1-9.
<http://www.fti.uab.es/tradumatica/revista/num7/articles/08/08central.htm> (consultato il 08/10/2014)
- Bernardini, S. (2007). *Collocations in translated text. A corpus-based study*. Ph. D. thesis. Middlesex University.
- Bernardini, S., Ferraresi, A. e M. Miličević (accepted). *From EPIC to EPTIC – exploring simplification in interpreting and translation from an intermodal perspective*.
- Centre National de Ressources Textuelles et Lexicales.
<http://www.cnrtl.fr/definition/>
- Cowie, A. P. (1991). "Multiword units in newspaper language". In S. Granger (1991). 101-116.
- Cowie, A. P. a cura di (1998). *Phraseology: theory, analysis and applications*. New York: Oxford University Press.
- Dayrell, C. (2007). "A quantitative approach to compare collocational patterns in translated and non-translated texts". *International journal of corpus linguistics*, 12 (3): 375-414.
- Durrant, P. e N. Schmitt (2009). "To what extent do native and non-native writers make use of collocations?". *IRAL*, 47: 157-177.

- Fernandes, L. (2006). "Corpora in translation studies: revisiting Baker's typology". *Fragmentos*, 30: 87-95.
- Ferraresi, A. e M. Miličević (submitted). *Phraseological patterns in interpreting and translation: similar or different?*
- Foley, J. A. a cura di (1996). *J. M. Sinclair on lexis and lexicography*. Singapore: UniPress.
- Fontenelle, T. (1998). "Discovering significant lexical functions in dictionary entries". In A. P. Cowie (1998). 189-207.
- Gandin, S. (2009). "Linguistica dei corpora e traduzione: definizioni, criteri di compilazione e implicazioni di ricerca dei corpora paralleli". *Annali della facoltà*, 5. Università degli studi di Sassari. 133-152.
- Gellerstam, M. (1986). "Translationese in Swedish novels translated from English". In W. Wollin e H. Lindquist (1986). 88-95.
- Granger, S. (1998). "Prefabricated patterns in advanced EFL writing: collocations and formulae". In A. P. Cowie (1998). 145-160.
- Granger, S. a cura di (1991). *Perspectives on the English lexicon: a tribute to Jacques van Roey*. Louvain-la-Neuve: Cahiers de l'Institut de Linguistique de Louvain.
- Hansen, G., Chesterman, A. e H. Gerzymisch-Arbogast a cura di (2009). *Efforts and models in interpreting and translation research: a tribute to Daniel Gile*. Amsterdam: John Benjamins Publishing Company.
- Howarth, P. (1998). "The phraseology of learners' academic writing". In A. P. Cowie (1998). 161-186.
- Jones, S. e J. Sinclair (1974/1996). "English lexical collocations". In J. A. Foley (1996). 21-54.
- Laviosa, S. (2002). *Corpus-based Translation Studies: theory, findings, applications*. Amsterdam-New York: Rodopi.
- Legal Information Institute. <http://www.law.cornell.edu/>
- Manuale interistituzionale di convenzioni redazionali*.
<http://bookshop.europa.eu/it/manuale-interistituzionale-di-convenzioni-redazionali-2011-pbOA3110655/>
- Masini, F. (2012). *Parole sintagmatiche in italiano*. Cesena-Roma: Caissa Italia.
- Moon, R. (1998). "Frequencies and forms of phrasal lexemes in English". In A. P. Cowie (1998). 79-100.
- Oxford Learner's Dictionaries*. <http://www.oxfordlearnersdictionaries.com/>
- Rover, R. V. (1995). "Rischio sociale e sicurezza nazionale: il radicalismo islamico. Analisi della sua origine, teorizzazioni e possibili riflessi sulla sicurezza nazionale". *Per aspera ad veritatem*, 2. Roma.

<http://gnosis.aisi.gov.it/sito%5CRivista2.nsf/servnavig/4> (consultato il 17/11/2014).

- Russo, M., Bendazzoli, C., Sandrelli, A. e N. Spinolo (2010). *The European Parliament Interpreting Corpus (EPIC): implementation and developments*. Paper presented at the international conference "Emerging topics in translation and interpreting/Nuovi percorsi in traduzione e interpretazione". 16-18 giugno 2010, SSLMIT, Università di Trieste. 1-25.
- Sandrelli, A. e C. Bendazzoli (2005). "Lexical patterns in simultaneous interpreting: a preliminary investigation of EPIC (European Parliament Interpreting Corpus)". *Proceedings from the Corpus Linguistics Conference Series 1*. Birmingham: University of Birmingham. 1-19. <http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx> (consultato il 09/09/2014)
- Sandrelli, A., Bendazzoli, C. e M. Russo (2010). "European Parliament Interpreting Corpus (EPIC): methodological issues and preliminary results on lexical patterns in simultaneous interpreting". *International journal of translation*, 22: 165-203.
- Shlesinger, M. (1998). "Corpus-based Interpreting Studies as an offshoot of Corpus-based Translation Studies". *Meta: translators' journal*, 43 (4): 486-493.
- Shlesinger, M. (2009). "Towards a definition of Interpretese: an intermodal, corpus-based study". In G. Hansen, A. Chesterman e H. Gerzymisch-Arbogast (2009). 237-253.
- Shlesinger, M. e N. Ordan (2012). "More spoken or more translated? Exploring a known unknown of simultaneous interpreting". *Target*, 24 (1): 43-60.
- Siepmann, D. (2005). *Discourse markers across languages: a contrastive study of second-level discourse markers in native and non-native text with implications for general and pedagogic lexicography*. Londra-New York: Routledge.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- SSLMIT Dev Online. <http://dev.sslmit.unibo.it/>
- Stubbs, M. (1995). "Collocations and semantic profiles: on the cause of the trouble with quantitative methods". *Functions of language*, 2 (1): 1-33.
- Teubert, W. (1996). "Comparable or parallel corpora?". *International journal of lexicography*, 9 (3): 238-264.
- Van Roey, J. (1990). *French-English contrastive lexicology: an introduction*. Louvain-la-Neuve: Peeters.
- Vocabolario Treccani: definizioni, etimologia e citazioni*.
<http://www.treccani.it/vocabolario/>

Wollin, W. e H. Lindquist a cura di (1986). *Translation studies in Scandinavia: proceedings from the Scandinavian Symposium On Translation Theory (SSOTT) II, Lund 14-15 June 1985*. Svezia: CWK Gleerup.

Zanettin, F. (2001). *IperGrimus. Ipermedia e traduzione*. inTRAlinea Ipermedia.
http://www.intralea.it/intra/ipermedia/IperGrimus/_private/default.htm
(consultato il 08/08/2014)

Zanettin, F. (2001). *Tipi di corpora per lo studio della traduzione*.
<http://www.intralea.org/monographs/zanettin/ipertesi/4.3.htm>
(consultato il 08/08/2014)

Appendice

EPTIC_01_2011 Guidelines

1. Introduction

EPTIC, the *European Parliament Translation and Interpreting Corpus*, is an intermodal corpus created by transcribing European Parliament speeches and their interpretations, and by subsequently pairing the transcripts of interpreted speeches and their source texts with the corresponding translated versions and respective source texts. EPTIC's construction is made possible by the fact that, for each plenary session, the European Parliament publishes so-called “verbatim” reports of proceedings consisting of transcripts of the speeches and their translations into all EU official languages; despite being called “verbatim”, these reports are often considerably edited (punctuation is added, context-related comments are removed, mistakes such as false starts, unfinished sentences or mispronunciations are corrected). The translations of the proceedings are then the result of an independently performed translation process based on the verbatim reports, without any reference to the interpreters' outputs.

An initial version of EPTIC (Bernardini et al. 2014) was based on the English <> Italian part of EPIC (Sandrelli and Bendazzoli 2005, Russo et al. 2010, Sandrelli et al. 2010), based on interpreting sources and targets (81 text¹ pairs in the English > Italian and 17 in the Italian > English direction, for a total of about 175,000 words); this version contained texts from the Parliament's part-session held in February 2004. Thanks to the work of several students at the University of Bologna, in 2014 the Italian > English portion was extended and an initial set of French texts were added. Texts from 2011 were used for this enlargement (the verbatim reports available on the European Parliament website in November 2013 went up to the plenary session of June 2011), leading to a total of 62 text sets for English > Italian, 68 text sets for Italian > English, 65 text sets for French > Italian and 68 text sets for Italian >

¹The word “text” refers to both transcripts of speeches and interpretations and verbatim reports and their translations.

French, for a total of about 280,000 words. Further enlargement is currently being worked on; these guidelines are intended to help those involved in the process of transcription and coding.

2. Obtaining the material

The speeches and the associated verbatim reports can be accessed via the European Parliament web page <http://www.europarl.europa.eu/plenary/en/debates-video.html>, which allows search by parliamentary term and date(s) of the sittings. Clicking on the search results (bottom of page), organised by topic-based sections, opens a page containing the verbatim report of a given section (in HTML format), as well as links to video recordings (see Figures 1 and 2). The report is in the language in which the page is consulted (meaning that for talks originally delivered in that language it constitutes a transcript, while it is a translation for the others), as is the video recording (consisting of original and interpreted speeches); the language of the page needs to be changed to access other language versions (a language code for the source is given in brackets if a speech is translated).

The screenshot displays the European Parliament website interface. At the top, the 'European Parliament' logo is visible. Below it, a navigation bar includes 'Index', 'Previous', 'Next', and 'Full text' buttons. A 'Procedure' section shows the document ID '2009/2214(INI)' and 'Document stages in plenary'. A table lists the document's history: 'Texts tabled: A7-0377/2010', 'Debates: PV 20/01/2011 - 4 (CRE 20/01/2011 - 4)', 'Votes: PV 20/01/2011 - 7.5 (Explanations of votes)', and 'Texts adopted: P7_TA(2011)0024'. The main section is titled 'Debates' and 'Thursday, 20 January 2011 - Strasbourg'. It features a sub-section '4. A sustainable EU policy for the High North (debate)' with a 'Video of the speeches' link. The report text begins with the President's introduction and the opening remarks of Michael Gahler, rapporteur, who discusses the report on a sustainable EU policy for the High North.

Figure 1. An example of a verbatim report page with links to video recordings

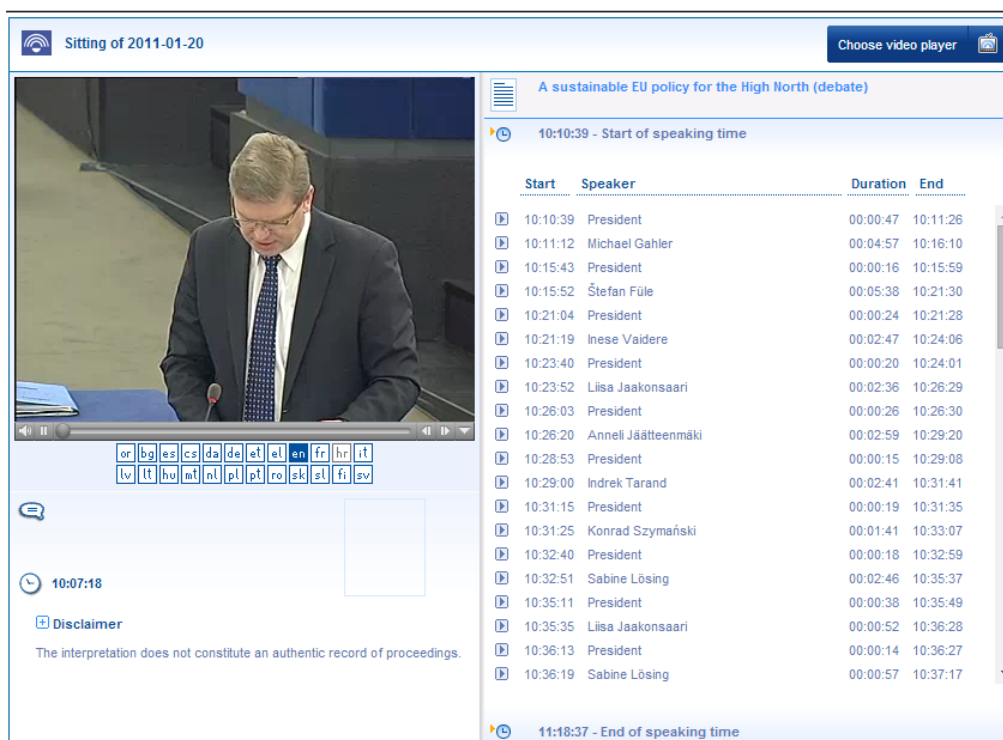


Figure 2. An example of a video recording window

The verbatim reports (sorted by date) can also be downloaded in .pdf or .odt format from <http://www.europarl.europa.eu/RegistreWeb/search/typedoc.htm?codeTypeDocu=PCRE>. Obtaining the reports (and saving them as .txt documents, see Section 3) is a central step in the creation of translation subcorpus components. For oral data, the reports can serve as a useful basis for a proper verbatim transcription, described in detail in Section 5.

To make the corpus useful for linguistic research, only texts reaching a minimum length of 60 words are to be included. Due to being very short and repetitive, speeches by Presidents, Vice-presidents or Chairmen when moderating the debate should be excluded too.

3. File format and naming conventions

All files should be saved in plain text format. We recommend using an advanced text editor such as Notepad++ (<http://notepad-plus-plus.org/>). Encoding should be set to ANSI and end of line to UNIX. The files should contain a metadata header, as described in Section 4, and a body of text. The transcription conventions outlined in Section 5 should be strictly adhered to,

and any language-specific additional issues should be dealt with in a systematic manner, and documented.

Daily sessions for each language are to be saved as separate files. File names should include the following information, ordered as below:

- st/tt → indicating whether the text is source or target
- in/tr → indicating whether the text belongs to the interpretation or translation subcorpus
- en/fr/it... → the language the text is in
- the date of the session in ddmmyy format

The first three elements should be separated with dashes, while the date follows an underscore. For instance, a source text of an interpretation from English delivered on 17 January 2011 should be named *st-in-en_170111.txt*, and its interpretation into Italian *tt-in-it_170111.txt*. Note that only the language the text is in is coded in the name.

4. Metadata

Each text needs to be preceded by a set of metadata useful for limiting the search in the corpus. The metadata contain a number of attribute/values pairs. The values should be entered in inverted commas and cannot contain any non-alphanumerical symbols (for instance, the political group “S&D” should be indicated as “SD”). Multiple words within inverted commas should be separated by dashes, while contiguous attributes are to be separated by an underscore. A full example is shown in Table 1, with all attributes defined below.

<pre>#text_id="001" _date="17-01-11- a" _length="short" _lengthw="232" _durat ion="short" _durations="76" _speed="hig h" _speedwm="182.7" _delivery="read" _to pic="Politics" _topicspec="Order-of- business" _type="st-in- en" _comments="NA"</pre>	text data
--	-----------

<pre> speaker_name="Tavares- Rui" _gender="M" _country="Portugal" _na tive="n" _politfunc="MEP" _politgroup=" GUE-NGL" st_length="NA" _lengthw="NA" _duration= "NA" _durations="NA" _speed="NA" _speedw m="NA" _delivery="NA" interpreter_gender="NA" _native="NA" </pre>	<p>speaker data</p> <p>source text data</p> <p>interpreter data</p>
---	---

Table 1. An example of metadata header (the column to the right explains the blocks)

The metadata attributes are defined in the following way:

- **text_id** → a number identifying the source text and its target, both in the interpretation and the translation subcorpus; the first text of each source language has the id “001”
- **date** → the date of the session the text belongs to, followed by “m” if the speech was pronounced in the morning (before 1 PM), or by “a” if it was pronounced in the afternoon (after 1 PM); the required date format is dd-mm-yy-m/a
- **length** → a broad classification of the text based on the number of words it contains (see Table 2)

length	short: <300 words medium: 301-1000 words long: >1000 words
duration	short: <2 minutes medium: 2-6 minutes long: >6 minutes
speed	low: <130 words per minute (w/m) medium: 131-160 w/m high: >160 w/m

Table 2. Values for the attributes length, duration and speed

- **lengthw** → text length in words, i.e. the exact number of words in the text, obtained using the following regular expression in Notepad++,

with the plugin TextFX → TextFX Quick → Find/Replace → tick on Match case and on Regular Expr → click on Count

```
[a-zA-Z0-9\xC0-\xFF]+([-,:][a-zA-Z0-9\xC0-\xFF]+)*
```

In the interpreting subcorpora the word count omits words enclosed in slashes (see Section 5), but truncated words and filled pauses (“ehm”) are counted. Dash-separated words are counted as one word, which might not be equally suitable for all languages; in English and in Italian these are indeed single words (such as *light-hearted*), while in French they can be either single words (*c’est-à-dire*) or two words (*voulez-vous*). The percentage of dash-separated words which are actually two words is not significant compared to the total number of words, so it is not worth modifying the regular expression or separating the two words. However, to make sure lapses of time and other spans are counted as two words, a space is inserted before and after the dash (*2010 - 2014*).

- **duration** → a broad classification of the text based on its duration in seconds (see Table 2)
- **durations** → speech duration in seconds
- **speed** → a broad classification of the pace of the speech based on the number of words per minute pronounced by the speaker (see Table 2)
- **speedwm** → number of words pronounced per minute, obtained by dividing the total number of words by the duration in minutes; numbers should be rounded up to a single decimal
- **delivery** → the mode of delivery of the speech – read, impromptu or mixed; for interpreting target texts, use “interpreted”
- **topic** → the wider topic that the text deals with, chosen from the list below, based on topicspec (see the next point) or the text itself, in case topicspec is too general

Economics-and-Finance
Employment
Environment
Health
Justice
Politics
Procedure-and-Formalities
Society-and-Culture
Science-and-Technology
Transport

- **topicspec** → the title of the section of the plenary session that the text belongs to; the titles can be found on the European Parliament website (see Section 2)
- **type** → a code showing whether the text is source or target (st/tt), interpreted or translated (in/tr), and which language it is in (en/it/fr...); equals the file name without the date
- **comments** → a field for comments pertaining to the text (e.g. problems with the video, particular types of errors...)
- **speaker_name** → the name of the speaker, in the format “Surname-Name”
- **gender** → the gender of the speaker (F/M)
- **country** → the country in which the speaker was elected as a member of a European institution
- **native** → whether or not the speaker speaks in his/her mother tongue (y/n)
- **politfunc** → the role of the speaker in the European institutions, chosen from the list below; if an MEP is also one of the 14 Vice-presidents of the European Parliament or if he/she is the Chairman of the session, the more important function should be chosen

MEP
MEP-Chairman-of-the-session
President-of-the-European-Parliament
Vice-President-of-the-European-Parliament
European-Commission

European-Council
Guest

- **politgroup** → the political group that the speaker belongs to, as indicated in brackets next to his/her name on the European Parliament website and in the verbatim reports; if the speaker's "politfunc" is "European-Commission", "European-Council" or "Guest", this attribute is not applicable and should be filled with "NA"
- **st_length/lengthw/duration/durations/speed/speedwm/delivery** → attributes concerning the source text (the values are copied from the source) and therefore applicable only to the target texts; this information is useful for a better understanding of the target text at issue (for instance, if the interpreter makes many mistakes, this could be due to the speed or modality of delivery of the source text)
- **interpreter_gender/native** → indicates the gender of the interpreter (F/M) and whether or not he/she is interpreting into his/her mother tongue (y/n); given that the European Parliament does not provide any information about the interpreters, the decision on the values of these attributes needs to be made by the transcriber; if the nativeness status is unclear, the value "unknown" can be used.

There are a number of cases in which some of the attributes are not relevant:

- information about the source text is not relevant for the source texts themselves;
- information about the interpreter is only needed for the interpreting subcorpus;
- the attributes "duration", "durations", "speed", "speedwm" and "delivery" are not relevant for the texts belonging to the translation subcorpus;
- the attribute "comments" is not relevant if the text does not present any problem or peculiarity.

In these cases, the non-relevant attributes are to be filled with “NA” (not applicable), and not to be left blank.

To sum up, table 3 shows all the attributes and all their possible values.

	Attribute	Value
Text data	text_id	<i>number identifying the text</i>
	date	<i>dd-mm-yy-m/a</i>
	length	<i>short/medium/long</i>
	lengthw	<i>number of words</i>
	duration	<i>short/medium/long/NA</i>
	durations	<i>speech duration in seconds/NA</i>
	speed	<i>low/medium/high/NA</i>
	speedwm	<i>number of words per minute/NA</i>
	delivery	<i>read/mixed/impromptu/interpreted/NA</i>
	topic	<i>Agriculture-and-Fisheries Economics-and-Finance Employment Environment Health Justice Politics Procedure-and-Formalities Society-and-Culture Science-and-Technology Transport</i>
	topicspec	<i>title of the section on the website</i>
	type	<i>st/tt-in/tr-language code</i>
	comments	<i>comment on the text/NA</i>
Speaker data	speaker_name	<i>Surname-Name</i>
	gender	<i>M/F</i>
	country	<i>name of the country</i>
	native	<i>y/n</i>
	politfunc	<i>MEP MEP-Chairman-of-the-session President-of-the-European-Parliament Vice-President-of-the-European-Parliament European-Commission European-Council</i>

		<i>Guest</i>
	politgroup	<i>abbreviation of the political group as indicated on the website/NA</i>
Source text data	st_length	<i>short/medium/long</i>
	lengthw	<i>number of words</i>
	duration	<i>short/medium/long/NA</i>
	durations	<i>speech duration in seconds/NA</i>
	speed	<i>low/medium/high/NA</i>
	speedwm	<i>number of words per minute/NA</i>
	delivery	<i>read/mixed/impromptu/interpreted/NA</i>
Interpreter data	interpreter_gender	<i>M/F/NA</i>
	native	<i>y/n/NA</i>

Table 3. Metadata attributes and values

5. Editing conventions

Both the texts belonging to the interpretation subcorpora and those making up the translation subcorpora are modified following some editing conventions. Orality traits that were not retained in the verbatim reports are reinserted in the oral subcorpora (sources and targets of interpreting) based on the video recordings. Common changes include reinserting omissions as well as truncated words, repetitions, disfluencies, pauses, etc. While many of the editing conventions described below are relevant only for the interpreting subcorpus, being closely linked to orality (see points 1-6), others apply both to the interpreting and the translation subcorpora (7-12). The applicable conventions should be followed closely.

1. **Punctuation:** It should follow the syntactic structure of the speech and the speaker's intonation.
2. **Truncated words:** If the speaker does not pronounce the entire word, this should be indicated with a dash (*propo-*). If the speaker makes a pause while pronouncing a word, the word should first be written in its correct form, and then put between slashes with an underscore where the speaker stops (*proposal /pro_posal/*).

3. **Disfluencies:** The word is first written in its correct form and then reported between slashes as effectively pronounced by the speaker (*proposal /preposal/*), indicating, if necessary, mistakes regarding stress (*proposal /pròposal/*) or word merging (*evitare eventuali /eventuare/*). The same convention is used when the speaker insists on one part of the word, extending one particular sound (*proposal /proposal/*). Pronunciation markers due to non-nativeness should not be indicated between slashes, for they do not constitute a mistake but point to the origin of the speaker.
4. **Pauses:** There are two kinds of pauses, filled and empty. Filled pauses should be indicated with “ehm” (coughing does not count as a filled pause). A single “ehm” should always be inserted, regardless of the pause duration or repetition. Empty pauses should be indicated with “...”. Empty pauses are defined as those that last an amount of time which sounds unnatural, and that interrupt the discourse flow in an unnatural way. When “...” is used at the end of a text, it means that the speaker is interrupted and does not finish a meaningful sentence.
5. **Unclear bits:** Use “#” when you cannot hear a portion of the speech because of technical problems, or you do not understand what the speaker says, as well as if the speaker switches to a different language (one that is not the language of the subcorpus at issue)
6. **Calques and “made-up” words:** Non-conventional words which can be understood only based on the source text should be indicated in braces (e.g. when the French interpreter uses the word “manganelle” for the Italian word “manganello”, {manganelle} should be written). Similar interventions should be explained in the “comments” field.
7. **Applause:** If the speech is interrupted by or ends with applause, [applause] should be inserted.
8. **Numbers:** Numbers are written as figures, following the conventions of individual languages. Numbers in the millions and billions are not to be written as figures (*12 million, rather than 12,000,000*). Fractions are

written as words. If numbers are not pronounced in their conventional way, that should be indicated between slashes (*2010 /twenty ten/*). Ordinals are written following the conventions of each language (*er/ère/ème* for French, *st/nd/rd/th* for English, *o/a* for Italian). Ordinals included in the names of important people or historical figures are written in Roman numerals (*John Paul II*). The symbol “%” is always written as a word in the relevant language.

9. **Character issues:** Inverted commas and apostrophes should be standardised using only those set by default in Notepad++ (" / ').
10. **Capitalisation:** To decide whether to use capital letters, refer to the Interinstitutional style guides (<http://bookshop.europa.eu/en/interinstitutional-style-guide-2011-pbOA3110655/>), official websites or monolingual dictionaries: for English the Oxford Learner’s Dictionary (<http://www.oxfordlearnersdictionaries.com/>), for French the CNRTL (<http://www.cnrtl.fr/definition/>), for Italian Vocabolario Treccani (<http://www.treccani.it/vocabolario/>). To distinguish between the European Commission and various parliamentary commissions, use capital letters only for the former. The same applies to the word “commissioner” – a “Commissioner” is a member of the European Commission, a “commissioner” is a member of a parliamentary commission.
11. **Acronyms and names not referring to people:** Acronyms and names of programmes, institutions, regulations, action plans, etc. should be written following the conventions set by the Interinstitutional style guides (see the link above), or relevant official websites. Names of programmes, institutions, regulations, action plans, etc. are to be put in capital letters only if the speaker uses their official names.
12. **Titles:** Titles are written as follows (for English, French and Italian respectively): *Mr, Mrs, Ms, sir; madam, monsieur, madame, mademoiselle; signor, signore, signora.*

Table 4 sums up the editing conventions.

Text Feature	Example	Editing convention
Punctuation		<i>Based on syntax and intonation</i>
Truncated words	propo	<i>propo-</i>
	pro posal	<i>proposal /pro_posal/</i>
Disfluencies	preposal	<i>proposal /preposal/</i>
	pròposal	<i>proposal /pròposal/</i>
	proposal	<i>proposal /proposal/</i>
	eventuare	<i>evitare eventuali /eventuare/</i>
Pauses	filled	<i>ehm</i>
	empty	<i>...</i>
Interrupted speech		<i>...</i>
Numbers	numbers	<i>in figures</i>
	326,000,000	<i>326 million/millions/milioni</i>
	326,000,000,000	<i>326 billion/milliards/miliardi</i>
	34%	<i>34 percent/pour cent/per cento</i>
	years	<i>in figures</i>
	fractions	<i>in words</i>
	ordinals	<i>figures+st/nd/rd/th/er/ère/ème/o /a or Roman numerals</i>
Unclear bits		<i>#</i>
Language switching		<i>#</i>
Calques/Made-up words	manganelle	<i>{manganelle}</i>
Applause		<i>[applause]</i>
Capitalisation		<i>Refer to Interinstitutional style guides, official websites or monolingual dictionaries</i>
	European Commission	<i>Commission/Commission/Com missione</i>

	parliamentary commission	<i>commission/commission/commi ssione</i>
	member of the European Commission	<i>Commissioner/Commissaire/Co mmissario</i>
	member of a parliamentary commission	<i>commissioner/commissaire/com missario</i>
Acronyms and non-personal names		<i>Refer to Interinstitutional style guides or official websites</i>
Titles		<i>Mr/Mrs/Ms/sir/madam monsieur/madame/mademoisell e signor/signore/signora</i>

Table 4. Editing conventions

References

- Bernardini S., Ferraresi A., Miličević M. (2014). *From EPIC to EPTIC: exploring simplification in interpreting and translation from an intermodal perspective*. Submitted.
- Sandrelli A., Bendazzoli C. (2005). “Lexical patterns in simultaneous interpreting: a preliminary investigation of EPIC (European Parliament Interpreting Corpus)”. *Proceedings from the Corpus Linguistics Conference Series 1*; University of Birmingham, Birmingham.
- Russo M., Bendazzoli C., Sandrelli A., Spinolo N. (2010). “The European Parliament Interpreting Corpus (EPIC): implementation and developments”. Paper presented at the international conference *Emerging Topics in Translation and Interpreting/Nuovi percorsi in traduzione e interpretazione*, 16-18 giugno 2010, SSLMIT, Università di Trieste.
- Sandrelli A., Bendazzoli C., Russo M. (2010). “European Parliament Interpreting Corpus (EPIC): methodological issues and preliminary results on lexical patterns in simultaneous interpreting”. *International journal of translation*, 22, 165-203.