

# EVENTI DAL VIVO E ACCESSIBILITÀ

## Uno studio di caso sul respeaking interlinguistico

ANNALISA SANDRELLI

UNIVERSITÀ DEGLI STUDI INTERNAZIONALI DI ROMA-UNINT

**Abstract** – Respeaking is commonly used to subtitle live TV programmes for the benefit of deaf and hard-of-hearing viewers (SDH subtitles): it is based on the interaction between the respeaker's competences and a speech recognition software trained to recognise his/her voice. In some countries this technique is also used to subtitle live events, such as conferences, debates, and so on. In recent years there have been attempts to use the interlingual variant of the technique, thus making it possible to translate a speech from one language to another (as you do in simultaneous interpreting) and obtain real-time subtitles. The respeaker listens to the incoming message and interprets it simultaneously to the software, adding punctuation verbally by means of voice commands; then, the respeaker's audio input is processed by the software, which produces the transcript; the respeaker checks the output and corrects any errors, and the subtitles are finally made available to the audience. After a short overview of the experimental research carried out in this field and the available theoretical models, the present paper assesses the accuracy of the interlingual subtitles (English into Italian) produced via IRSP in a live event (a symposium). The analysis outlines the specific difficulties of this technique when applied to a live event, with the aim of producing useful data and suggestions for the development of practice guidelines.

**Keywords:** interlingual respeaking (IRSP); live events; subtitle accuracy; training.

### 1. Introduzione

Il respeaking è una modalità di produzione di sottotitoli in tempo reale che sfrutta l'interazione tra le competenze di un operatore umano e un software di riconoscimento del parlato *speaker-dependent*. Il respeaker ascolta il messaggio da sottotitolare e lo ripete al software (precedentemente addestrato a riconoscere la sua voce), aggiungendo la punteggiatura mediante dei comandi vocali e riformulando il discorso ove necessario.

In Europa questa tecnica si è sviluppata a partire dal 2001 in un contesto televisivo: i pionieri del respeaking sono stati i servizi di sottotitolazione della BBC (Regno Unito) e della VRT (Belgio, regione delle Fiandre), che hanno avviato una sperimentazione più o meno nello stesso periodo (Romero-Fresco 2011). Vent'anni dopo, il respeaking è ormai la tecnica più diffusa in

Europa per produrre sottotitoli intralinguistici in tempo reale per i programmi televisivi in diretta.<sup>1</sup> I sottotitoli prodotti in questo modo sono pensati per la popolazione sorda e audiolesa, ma sono utili a chiunque voglia seguire un programma ascoltando e leggendo contemporaneamente la trascrizione, come ad esempio gli appartenenti alle minoranze linguistiche, le persone con deficit cognitivi, gli studenti stranieri, e così via.

In alcuni paesi, da qualche anno il *respeaking* viene proposto anche fuori dal contesto televisivo, per sottotitolare “eventi dal vivo”, come conferenze, lezioni universitarie, congressi e così via. Più precisamente, un evento dal vivo è

an event happening in real-time, where the audience attends in person, and which is not watched in its entirety through a screen, although parts of it (the subtitles and certain visual elements such as PowerPoint slides or video clips) will necessarily be displayed on one (or more). (Moore 2020, p. 179)

Tali servizi esistono già nelle Fiandre, in Polonia, in Germania, Austria, Svizzera<sup>2</sup> e negli Stati Uniti.<sup>3</sup> Nel Regno Unito il progetto *Respeaking at Live Events* (Moore 2018, 2020) ha l’obiettivo di promuovere l’utilizzo del *respeaking* in varie tipologie di eventi dal vivo (visite ai musei, conferenze, dibattiti, eccetera) e di contribuire a definire delle linee guida.

Negli ultimi cinque anni alcuni progetti di ricerca hanno iniziato a valutare la fattibilità del *respeaking* in versione interlinguistica, anche perché esistono già alcune realtà in cui il *respeaking* interlinguistico (*interlingual respeaking* o *IRSP*) viene occasionalmente proposto, come nelle Fiandre. Questa tecnica permette di superare contemporaneamente le barriere sensoriali e quelle linguistiche, garantendo così un’accessibilità universale ai contenuti.

In Italia il *respeaking* è ancora poco diffuso. In televisione il *respeaking* intralinguistico è stato introdotto in via sperimentale nel 2008 nel programma contenitore *Cominciamo bene estate* su RAI Tre (Romero-Fresco 2011), e negli anni successivi è stato usato in alcune edizioni del telegiornale, talk show e altri programmi in diretta. Tuttavia, questa tecnica è ancora poco sfruttata dalla RAI, che preferisce affidarsi prevalentemente alla stenotipia soprattutto per i programmi nelle fasce orarie più seguite (Astuto 2014-15). Al di fuori del contesto televisivo, mediante i siti delle poche aziende attive in

<sup>1</sup> In Gran Bretagna la BBC, che ormai sottotitola il 100% del proprio palinsesto, si affida al *respeaking* per tutti i programmi in diretta (previdoni del tempo, talk show, e così via); inoltre, anche le reti commerciali si sono ormai adeguate all’esempio della BBC, per non perdere spettatori.

<sup>2</sup> Paesi in cui il servizio viene chiamato *speech-to-text interpreting* e viene svolto con l’ausilio di speciali tastiere.

<sup>3</sup> Negli USA si chiama *Live Event Captioning*.

questo ambito si ha notizia dell'utilizzo del respeaking in assemblee degli azionisti, riunioni, dibattiti, e così via, ma non ci sono dati ufficiali sull'effettiva diffusione del servizio.<sup>4</sup> Per quanto riguarda la variante interlinguistica, ci sono dei casi di studio documentati in alcune tesi di laurea: ad esempio, Marchionne (2010-11) ha svolto un esperimento di sottotitolazione mediante IRSP tra il francese e l'italiano in un contesto televisivo, mentre Serafini (2014-15) ha organizzato un servizio di IRSP (inglese-italiano) in un festival del cinema, descrivendo la configurazione del servizio e il gradimento del pubblico presente. Poiché sono ancora pochissimi i contesti in cui è presente un servizio di sottotitolazione dal vivo mediante IRSP, è importante descrivere anche dei piccoli casi di studio per poter raccogliere quanti più dati possibile su questa prassi così "giovane": il presente contributo va proprio in questa direzione, analizzando l'accuratezza del servizio di IRSP svoltosi in un evento dal vivo, cioè un simposio accademico. Nella seconda sezione viene descritto il processo di IRSP, facendo una breve rassegna delle ricerche più rilevanti; vengono poi illustrati i nostri dati e la metodologia di analisi (sezione 3), i risultati dello studio (sezione 4) e infine le conclusioni (sezione 5).

## 2. Il respeaking interlinguistico: una panoramica

Le ricerche sul respeaking sono ancora limitate: in questo paragrafo si fa una panoramica dei pochi studi esistenti sulla variante interlinguistica, con brevi rimandi alla variante intralinguistica quando è rilevante.

### 2.1. Il processo di sottotitolazione in tempo reale mediante IRSP

Il respeaking interlinguistico è una modalità di traduzione ibrida che condivide alcuni aspetti con l'interpretazione simultanea e altri con la sottotitolazione dei programmi preregistrati: “[W]ith regard to the process, ‘interlingual respeaking’ [...] is really a form of simultaneous interpreting, while the product, in the intralingual as well as the interlingual variant, is a set of subtitles” (Romero-Fresco, Pöchhacker 2017, p. 158).<sup>5</sup>

Dato che ci sono molte variabili in gioco, la configurazione del servizio di IRSP può variare nei diversi contesti: il recente modello proposto da Pöchhacker and Remael (2019, p. 137) ha il merito di evidenziare gli

<sup>4</sup> Si vedano ad esempio <https://culturabile.it/sottotitolazione/> e <http://www.artis-project.it/servizi-per-accessibilita/respeaking/> (16.11.2020).

<sup>5</sup> “Per quanto riguarda il processo, ‘il respeaking interlinguistico’ [...] è di fatto una forma di interpretazione simultanea, mentre per quanto riguarda il prodotto si tratta di sottotitoli, sia quando si parla di respeaking intralinguistico che interlinguistico.”

aspetti che accomunano un po' tutte le situazioni e si basa sul Modello degli Sforzi proposto da Gile (2015) per l'interpretazione simultanea, arricchito con le componenti specifiche del *respeaking* interlinguistico. Prima di tutto, si mette in rilievo che ogni servizio di IRSP comprende tre macro-fasi, cioè il pre-, il peri- e il post-processo. Nel pre-processo avviene la preparazione del *respeaker* (cioè l'approfondimento di temi e terminologia dell'incarico) e la preparazione del software di riconoscimento del parlato, con la creazione di eventuali macro<sup>6</sup> e l'arricchimento del vocabolario interno mediante aggiunta di vocaboli, nomi propri o sigle (per permetterne la trascrizione). Nel peri-processo avviene il *respeaking* vero e proprio, che comprende l'ascolto e comprensione, la riformulazione strategica (cioè pensata per le esigenze del software), la dettatura e il monitoraggio della produzione orale; dopodiché, questo input orale viene elaborato dal software che esegue il riconoscimento e lo trasforma in un testo scritto. Il *respeaker* controlla ed eventualmente corregge la trascrizione, trasmessa poi al pubblico mediante un software di sottotitolazione o appositi programmi per gli eventi dal vivo (si veda la sezione 3.1). Infine, nella fase post-processo si fa un controllo sulla qualità del lavoro svolto, si corregge il testo (soprattutto se i sottotitoli devono essere riutilizzati, ad esempio in registrazioni del programma televisivo o della conferenza) e si arricchisce ulteriormente il vocabolario del software, inserendo i nuovi termini eventualmente emersi durante il lavoro.

In questo modello base del processo di IRSP possono variare molti fattori, riassunti in una breve panoramica da Davitti, Sandrelli (2020). Prima di tutto, il servizio può essere svolto da una sola persona o da più persone che si suddividono i compiti (rispettivamente, modello Mono, Duo e Multi; Remael *et al.* 2016): più specificamente, un *respeaker* può essere impegnato a tradurre da una lingua all'altra, enunciando chiaramente per evitare problemi di riconoscimento da parte del software e aggiungendo la punteggiatura orale, mentre una o due altre persone possono occuparsi del controllo della trascrizione, di eventuali correzioni e della formattazione e trasmissione dei sottotitoli. Il *respeaker* può essere fisicamente presente nel luogo dell'evento oppure fornire il servizio da remoto. Il flusso di lavoro può essere gestito dal *respeaker* col solo ausilio del software di riconoscimento del parlato o con ulteriori componenti, quali un software di traduzione automatica. Inoltre, ci sono molte possibili variazioni anche dal punto di vista della fruizione da parte del pubblico: in televisione i sottotitoli possono essere trasmessi in chiaro o attivabili mediante il telecomando, mentre in un evento dal vivo il testo può essere proiettato su uno schermo visibile a tutti o inviato su un

<sup>6</sup> Una macro è un comando vocale breve che si usa per trascrivere espressioni frequenti e lunghe, ad esempio per far precedere il sottotitolo da un'etichetta che identifica il parlante (come GIORNALISTA) o per ottenere una formattazione particolare del testo.

dispositivo personale (smartphone, PC, tablet). Il testo può essere di svariate righe o formattato come dei sottotitoli veri e propri, può essere a scorrimento (*scrolling*) o a blocchi, con latenza o quasi sincronizzato con il discorso originale, e così via. Infine, ci possono essere variazioni relative al genere di programma TV e tipologia di evento dal vivo, al formato comunicativo (monologico, dialogico, dibattito con più partecipanti), alle tematiche (da generaliste a specializzate) e alle caratteristiche di ciascun parlante (accento, velocità di eloquio e così via). Tutte queste variabili rendono difficile confrontare tra di loro i vari contesti: in tutti, però, è sicuramente fondamentale valutare la qualità del servizio di IRSP.

## 2.2. La valutazione della qualità

Dato la complessità del processo, si capisce come la valutazione della qualità dei sottotitoli realizzati mediante IRSP sia piuttosto complicata. Per una valutazione complessiva, sarebbe necessario analizzare almeno tre aspetti: la latenza, la velocità e, soprattutto, l'accuratezza dei sottotitoli, che è l'oggetto del presente lavoro.

La latenza è il ritardo dei sottotitoli rispetto al discorso originale ed è legata a vari fattori. Prima di tutto c'è il *décalage* del respeaker che, come l'interprete simultaneo, rimane sempre indietro di qualche parola per capire il senso del messaggio e trovare le soluzioni traduttive più adeguate: si tratta quindi di un ritardo causato dal fattore umano. Rispetto alla simultanea, c'è poi un altro aspetto che dipende invece dall'interazione uomo-macchina, cioè la necessità di enunciare molto chiaramente perché il software recepisca ed esegua: “the intermediary of speech recognition software demands that the words spoken conform to the capabilities and limitations of the speech recognition tool in use, rather than the human ear” (Moore 2020, p. 178).<sup>7</sup> Ciò richiede uno sforzo supplementare da parte del respeaker che, mentre traduce, deve anche immaginare la propria frase in forma scritta e decidere dove sono necessari un segno di interpunzione o una formattazione particolare e dettare i comandi vocali o le macro necessarie. Infine, c'è il ritardo causato dal software che riceve il “testo intermedio” prodotto dal respeaker (Pöchhacker, Remael 2019) e deve elaborarlo per produrre la trascrizione. Quando il respeaker lavora in modalità Mono, a tutto ciò bisogna aggiungere il tempo necessario per il controllo della trascrizione, l'eventuale correzione manuale e la trasmissione del testo mediante appositi programmi. Il ritardo accumulato a causa di tutti questi fattori deve essere

<sup>7</sup> “L'intermediazione del software di riconoscimento del parlato esige che le parole pronunciate siano conformi alle capacità e ai limiti dello strumento di riconoscimento usato, invece che all'orecchio umano.”

tenuto sotto controllo, per evitare che i sottotitoli siano troppo fuori sincrono rispetto al parlante o a eventuali informazioni visive.<sup>8</sup> Nel 2015 l'Autorità britannica delle telecomunicazioni ha commissionato uno studio a campione sulla qualità dei sottotitoli intralinguistici sui maggiori canali televisivi del Regno Unito (Ofcom 2015): in tale studio la latenza media era di circa 5,6 secondi, molto superiore alla soglia raccomandata di 3 secondi. Nello studio di Moores (2020) sul respeaking intralinguistico in eventi dal vivo la latenza media si è attestata sui 5,8 secondi, ma con picchi fino a 19,4 secondi e con grosse differenze da un evento all'altro (in ragione delle diverse caratteristiche tecnico-logistiche di ogni evento). Poiché nel respeaking interlinguistico c'è anche la componente traduttiva, ci si può aspettare una latenza mediamente superiore a quella dei sottotitoli intralinguistici, ma non esistono ancora linee guida in merito.

Per quanto riguarda la velocità dei sottotitoli, nel respeaking intralinguistico è stato stimato che, se la velocità del discorso originale è molto alta e il respeaker cerca di trascrivere tutto nel modo più fedele possibile, i sottotitoli diventano troppo veloci e quasi illeggibili (Romero-Fresco 2009, 2011). Nel già citato rapporto Ofcom si è visto che nel campione di sottotitoli analizzati ce n'erano alcuni che superavano le 200 parole al minuto (wpm), sebbene per i sottotitoli preregistrati si raccomandi di non superare le 160-180 wpm. Anche in questo caso non esistono ancora direttive precise per il respeaking interlinguistico, ma è evidente che, per fare in modo che i sottotitoli non siano troppo veloci per il pubblico, l'approccio traduttivo prevalente deve prevedere la comunicazione del messaggio in modo sintetico. Una certa riduzione è fisiologica e legata alla natura stessa di questa prassi traduttiva, poiché, oltre a tradurre, il respeaker deve contemporaneamente svolgere anche altre complesse operazioni: è importante quindi saper sintetizzare, individuando ridondanze e informazioni secondarie che possono essere tagliate senza conseguenze per il messaggio.

Tali considerazioni ci portano a parlare della valutazione dell'accuratezza, che è una questione assai complessa. Per il respeaking intralinguistico si è ormai affermato nel settore il modello NER, che tiene conto del fatto che gli errori riscontrati nei sottotitoli possono essere stati causati dal respeaker o dal software. Attribuendo un punteggio a ciascun errore in base alla sua gravità, il modello permette di calcolare la percentuale di accuratezza (si vedano Romero-Fresco 2011; Romero-Fresco, Martínez 2015). Per il respeaking interlinguistico è stato proposto il modello NTR, che arricchisce il modello NER con input provenienti dagli studi sull'interpretazione simultanea (Romero-Fresco, Pöchhacker 2017). Il

<sup>8</sup> Ad esempio, nel contesto di una conferenza le informazioni presenti nella presentazione multimediale dell'oratore.

modello calcola il rapporto tra il numero di parole presenti nei sottotitoli (N) e gli errori presenti negli stessi, distinguendo tra gli errori del software (cioè R, gli errori di riconoscimento del parlato) e gli errori umani (cioè T, gli errori di traduzione veri e propri).

Modello NTR			
N: numero di parole			
Punteggio errori <i>minor</i> : -0,25 <i>major</i> : - 0,50 <i>critical</i> : - 1			
T: e. di traduzione	e. di contenuto		omissioni
			aggiunte
sostituzioni			
	e. di forma	correttezza	
		stile	
R: e. di riconoscimento			
EE: interventi efficaci			

Tabella 1  
Il modello NTR.

Come si vede dalla Tabella 1, gli errori di traduzione possono essere di forma o di contenuto e, mentre i primi si suddividono ulteriormente in errori relativi alla correttezza grammaticale e in errori stilistici, i secondi si suddividono in omissioni, aggiunte e sostituzioni. Le omissioni sono i casi in cui elementi informativi del testo di partenza vengono tagliati, le aggiunte i casi in cui vengono inseriti elementi non pertinenti nei sottotitoli e le sostituzioni i casi in cui viene modificato il significato del testo originale, sostituendo un elemento con un altro completamente diverso (gli errori di traduzione in senso classico). La formula prevede che ogni errore venga penalizzato con una perdita di punteggio di -0,25 se l'errore è lieve (*minor*) e non ostacola la comprensione, -0,50 se è di gravità intermedia (*major*) e causa confusione o perdita di informazione; e -1 punto se è grave (*critical*), cioè se introduce informazioni fuorvianti o false. Dal numero totale di parole nei sottotitoli si sottrae il valore di tutti gli errori di traduzione (T) e di quelli di riconoscimento (R), si divide il punteggio ottenuto per il numero di parole e lo si moltiplica per 100, ottenendo la percentuale dei sottotitoli priva di errori.

Il modello classifica come errori i casi in cui gli interventi del respeaker sul testo hanno causato danni, ma non quando sono andati a buon fine: ad esempio, se il respeaker è riuscito a condensare in modo efficace il testo eliminando parti ridondanti dell'originale, non viene considerata un'omissione, ma un intervento corretto (*effective edition* - EE). A questi ultimi non viene attribuito un punteggio, ma vengono annotati durante

l'analisi NTR, perché sono indicativi della capacità del respeaker di padroneggiare il testo.

Come si intuisce, l'applicazione del modello NTR ha un certo margine di soggettività, sia nella classificazione degli errori che, soprattutto, nell'attribuzione dei punteggi. Essendo un modello relativamente recente, ha bisogno di essere validato in vari contesti, anche se i suoi ideatori riferiscono di un elevato grado di correlazione nei loro dati fra i punteggi assegnati alle stesse prestazioni da parte di valutatori diversi (Romero-Fresco, Pöchhacker 2017). Per adesso, il modello è stato applicato in alcuni studi sperimentali, cioè Dawson (2019), Dawson, Romero-Fresco (2021) e Davitti, Sandrelli (2020), brevemente riassunti qui di seguito.

Dawson (2019) e Dawson, Romero-Fresco (2021) sono due studi relativi al progetto ILSA (*Interlingual Live Subtitling for Access*).<sup>9</sup> Il primo è un esperimento pilota servito a preparare il secondo; in entrambi i partecipanti hanno fornito delle prestazioni di IRSP nella coppia di lingue inglese-spagnolo. In Dawson (2019) i 10 soggetti avevano un percorso formativo vario (con competenze di sottotitolazione, interpretazione, respeaking, anche in combinazione tra di loro) e hanno ricevuto una formazione minima in IRSP (due ore), per poi sottotitolare tre brevi video (di circa due minuti) dallo spagnolo all'inglese e tre dall'inglese allo spagnolo; alcuni video erano di genere televisivo e altri assimilabili agli eventi dal vivo (discorsi presidenziali). L'obiettivo era testare la fattibilità del respeaking interlinguistico e la metodologia per il progetto più ampio (Dawson, Romero-Fresco 2021), nel quale 44 partecipanti hanno seguito un corso online di 4 settimane in IRSP (inglese-spagnolo). Anche in questo caso gli studenti avevano una formazione precedente variabile (solo in sottotitolazione, solo in interpretazione o un po' in entrambe). Il modello NTR è stato applicato in entrambi gli studi: i risultati sembrano indicare che, con la formazione adeguata, è possibile raggiungere anche in IRSP un tasso di accuratezza del 98%, cioè lo standard minimo di qualità che è stato indicato per i sottotitoli intralinguistici prodotti in tempo reale. Inoltre, gli interpreti sembrano avere un certo vantaggio sugli altri studenti in termini di apprendimento, anche se la loro formazione non è sufficiente di per sé a garantire un buon risultato in IRSP.

Davitti e Sandrelli (2020) hanno realizzato uno studio pilota per testare parte degli strumenti metodologici da usare nel progetto *SMART*,<sup>10</sup> il cui

<sup>9</sup> Programma Erasmus+ (2017-1-ES01-KA203-037948): <http://ka2-ilsa.webs.uvigo.es/> (16.11.2020).

<sup>10</sup> SMART (*Shaping Multilingual Access through Respeaking Technology*) è un progetto iniziato a luglio 2020 e finanziato dallo Economic and Social Research Council (ESRC), con capofila l'Università del Surrey e partner l'Università di Roehampton, l'Università degli Studi

obiettivo è definire le competenze, le abilità cognitive e le caratteristiche personali che favoriscono una rapida ed efficace acquisizione delle competenze di IRSP. 25 studenti di tre università (UNINT, Surrey e Roehampton) hanno partecipato a un corso accelerato di IRSP della durata di 6-8 ore. Nel corso, svoltosi in presenza in un solo giorno, gli studenti hanno ricevuto un'infarinatura di respeaking (dall'italiano all'italiano), prima di procedere alla variante interlinguistica nella direzione inglese-italiano, lavorando su testi appartenenti al genere "conferenza". Anche in questo caso il background degli studenti era variabile, con alcuni soggetti che avevano frequentato moduli di sottotitolazione, di interpretazione, di entrambe e/o anche di respeaking intralinguistico. Nessuno degli studenti ha raggiunto la soglia del 98% di accuratezza; tuttavia, quelli con le competenze più complete e variegata (con interpretazione + sottotitolazione e interpretazione + sottotitolazione + respeaking) hanno ottenuto valori medi più alti del resto del gruppo, con picchi tra il 95 e il 96%. Tenendo conto che per tutti i soggetti era la prima esperienza di IRSP, i risultati sono abbastanza incoraggianti e sembrano suggerire che avere alle spalle una formazione più composita sia di aiuto nell'acquisizione della tecnica.

Nei tre studi succitati il modello NTR sembra aver dato buoni frutti, ma è stato applicato a dati sperimentali. Non ci risultano applicazioni del modello a dati provenienti da contesti di lavoro reali, data anche la loro attuale scarsità. Il presente studio si propone di dare un contributo in questa direzione, applicando il modello NTR a un piccolo campione di dati provenienti da un evento dal vivo svoltosi a Roma nel 2015. La sezione seguente illustra i dati e la metodologia utilizzata per l'analisi.

### 3. Metodologia e dati

Come si è visto, il respeaking interlinguistico è ancora raro in Italia, il che rende la raccolta dati estremamente difficoltosa. Il servizio non è ancora utilizzato in televisione e non esiste un vero e proprio mercato congressuale, come invece avviene per la ben più radicata prassi dell'interpretazione simultanea. Per questi motivi, quando nel 2015 il nostro ateneo ha organizzato il Quinto Simposio Internazionale *Respeaking, sottotitolazione in diretta e accessibilità* (12 giugno 2015),<sup>11</sup> il Comitato Organizzatore Locale ha ritenuto opportuno offrire sia un servizio di interpretazione simultanea che uno di sottotitolazione in diretta mediante respeaking, registrando il tutto a

Internazionali di Roma – UNINT e l'Università di Vigo, oltre ad alcuni partner aziendali (codice ES/T002530/1).

<sup>11</sup> Si veda: <https://www.unint.eu/it/calendario-eventi/-22.html> (16.11.2020).

scopo di ricerca (ovviamente previo consenso di tutti gli interessati, cioè oratori, interpreti e respeaker). Il Simposio è stato descritto in modo dettagliato in Sandrelli (2020). Qui verranno ricordate solo le caratteristiche principali dell'evento, per poi illustrare il corpus di dati analizzati nel presente lavoro.

### **3.1. Il Simposio**

Il Simposio è stato un evento di un giorno, con la partecipazione di 20 oratori tra studiosi, ingegneri e sviluppatori di software, provenienti da diversi paesi europei. Le due lingue ufficiali erano l'inglese e l'italiano, ma tutte le relazioni congressuali (su tematiche legate alla ricerca e sviluppo) sono state presentate in inglese. La maggior parte dei relatori non era di madrelingua inglese, ma usava questa lingua quotidianamente per lavoro; l'italiano è stato utilizzato invece nei discorsi di apertura e chiusura, nella tavola rotonda finale e in alcuni annunci di carattere organizzativo e logistico.

Il Simposio si è svolto nell'Aula Magna dell'università, un auditorium su due piani dotato di computer, schermo gigante, proiettore e un impianto per la simultanea, con 4 cabine insonorizzate in galleria. Il servizio di IRSP è stato offerto da uno dei nostri sponsor, l'associazione *onA.I.R.* di respeaking, grazie a 4 respeaker che si sono alternati durante la giornata. Il software utilizzato per il respeaking era *Dragon Naturally Speaking* (v. 12) in versione italiana, installato su computer portatili personali. A causa della conformazione dell'auditorium, non c'era la possibilità di appendere un secondo schermo accanto, sopra o sotto quello principale per i sottotitoli, come a volte si fa nei festival del cinema. Si è deciso perciò di proiettare i sottotitoli direttamente sulle diapositive degli oratori mediante il software *Text-on-Top*: a tutti gli oratori, perciò, è stato inviato un modello per le presentazioni, con uno spazio vuoto pari a tre righe di testo nella parte inferiore di ogni diapositiva (Figura 1).



Figura 1

La configurazione del servizio IRSP nel Simposio.

Per questioni di praticità, nel presente lavoro si fa riferimento ai testi prodotti in diretta nel Simposio come *sottotitoli*, nonostante il formato un po' anomalo (su tre righe, con scorrimento a blocchi), anche perché in effetti comparivano *sotto* il testo dell'oratore; in altri casi, come fa rilevare Moores (2020, p. 179-180), il testo può essere localizzato altrove e la dicitura *titles* ("didascalie") potrebbe essere più corretta. Inoltre, come si è accennato nella sottosezione 2.1, esiste la possibilità di inviare i sottotitoli direttamente ai dispositivi personali degli utenti, in modo che siano visibili solo ai diretti interessati. Nel nostro caso, i sottotitoli dovevano essere visibili a tutto il pubblico, poiché potevano essere utili a anche a coloro che ascoltavano la simultanea con gli auricolari.

Infine, i respeaker hanno prescelto il modello *Mono*, cioè con un solo operatore che si occupava di tutte le operazioni necessarie (dalla traduzione al controllo della trascrizione, eventuali correzioni mediante tastiera e trasmissione dei sottotitoli); ovviamente i respeaker si alternavano al lavoro. Bisogna anche sottolineare che gli operatori avevano una buona esperienza di respeaking, ma non in modalità interlinguistica. Faceva eccezione uno di loro che ha anche una formazione da interprete e aveva già fornito un servizio di IRSP in qualche occasione: sono i sottotitoli prodotti da quest'ultimo a essere oggetto della nostra analisi nella Sezione 4.

### 3.2. L'archivio multimediale e i dati per l'analisi

Le relazioni presentate dagli oratori sono state filmate con una telecamera fissa; le versioni interpretate simultaneamente dagli interpreti ed i testi prodotti dai respeaker sono stati registrati dai diretti interessati mediante registratori digitali posti dentro le cabine. A causa di problemi tecnici durante l'evento, una parte degli interventi dei relatori non è stata registrata correttamente, come pure alcune parti del lavoro degli interpreti e dei respeaker. Nel complesso, però, è stato possibile raccogliere una buona quantità di dati. Alla fine del Simposio tutto il materiale raccolto è stato salvato e trascritto per poter creare un archivio multimediale, anche grazie all'aiuto di una laureanda che ha scelto di confrontare nella sua tesi il servizio di respeaking e l'interpretazione simultanea (Luppino 2016-2017). Grazie al software *ELAN* (v. 4.9.4), per ogni intervento congressuale è stato creato un file multimediale che comprende la registrazione video dell'oratore originale, i due file audio della resa dell'interprete e del respeaker, le trascrizioni delle registrazioni e i sottotitoli proiettati durante il Simposio.

In Sandrelli (2020) una griglia di analisi sviluppata appositamente è stata applicata a un piccolo campione (4 testi in lingua di partenza, due relazioni scientifiche e due interventi di moderatori), per confrontare il testo in lingua d'arrivo prodotto in simultanea con i sottotitoli prodotti mediante IRSP. Nel presente lavoro si è invece deciso di focalizzarsi solamente sul servizio di IRSP e in particolare sulla tipologia testuale della *relazione congressuale*, escludendo gli interventi dei moderatori, le domande del pubblico, eccetera. Sono quindi state prese in esame 6 relazioni scientifiche presentate al Simposio, come da tabella 2.

	<b>durata</b>	<b>n. parole TP</b>	<b>velocità (wpm)</b>
<b>S1</b>	17' 55''	1,788	100
<b>S2</b>	6' 59''	970	139
<b>S3</b>	10' 12''	1,534	150
<b>S4</b>	22' 48''	2,902	127
<b>S5</b>	17' 17''	2,596	149
<b>S6</b>	19' 18''	2,946	153
<b>TOTALE</b>	1h 34' 29''	12,736	136*

Tabella 2  
Corpus di analisi (\* = velocità media).

Il corpus analizzato comprende oltre 90 minuti di registrazioni e un totale di quasi 13,000 parole di trascrizione in lingua inglese. Come si vede, la durata degli interventi è stata molto variabile, da un minimo di circa 7 minuti a un massimo di circa 23. Inoltre, poiché ogni oratore aveva un suo ritmo di eloquio

(da un minimo di 100 a un massimo di 153 wpm), anche la lunghezza della trascrizione di discorsi di durata simile (come S1 e S5) varia parecchio. Vale la pena di ricordare che, nella letteratura sull'interpretazione simultanea si stima che una velocità media del testo di partenza comoda per l'interprete sia di circa 100-120 wpm (Pöchhacker 2004, p. 129): sebbene entrino in gioco numerosi altri fattori a determinare la difficoltà di un incarico (familiarità con l'argomento, tempo e materiale disponibile per la preparazione, familiarità con l'accento dell'oratore e così via), si ritiene che in generale a questa velocità un interprete professionista riesca a tradurre abbastanza agevolmente. Poiché il respeaking interlinguistico comprende anche altre operazioni oltre alla traduzione in tempo reale, si può ipotizzare che quasi tutti i discorsi (ad eccezione forse di S1) siano stati di difficile gestione per il respeaker.

## 4. Risultati

I principali risultati dello studio vengono qui presentati partendo da una visione d'insieme e poi scendendo nel dettaglio degli errori riscontrati più frequentemente nel nostro corpus.

Il primo aspetto su cui porre l'attenzione è un dato meramente quantitativo, cioè il numero di parole presenti nei sottotitoli italiani prodotti mediante IRSP. Generalmente quando si traduce dall'inglese all'italiano il testo di arrivo (TA) tende ad essere un po' più lungo dell'originale un po' in tutti i tipi di traduzione. In particolare, inoltre, la riduzione testuale è connaturata sia alla sottotitolazione (a causa dei vincoli di tempo e spazio) sia all'interpretazione simultanea (a causa del fattore tempo): il respeaking interlinguistico, che condivide aspetti dell'una e dell'altra attività, non fa eccezione. Nella Tabella 3, insieme alla velocità media dell'originale (OSR) e alla lunghezza di ciascun testo di partenza (TP), vengono presentati i dati relativi alla lunghezza dei testi di arrivo e la percentuale di riduzione.

	durata	OSR (wpm)	n. parole TP	n. parole TA	% riduzione
<b>S1</b>	17' 55''	100	1,788	931	48%
<b>S2</b>	6' 59''	139	970	468	52%
<b>S3</b>	10' 12''	150	1,534	748	52%
<b>S4</b>	22' 48''	127	2,902	1,342	54%
<b>S5</b>	17' 17''	149	2,596	1,153	56%
<b>S6</b>	19' 18''	153	2,946	1,331	55%
<b>MEDIA</b>	15' 1''	136	12,736	5,973	53%

Tabella 3

Durata, velocità, lunghezza TP (testo di partenza),  
lunghezza TA (testo di arrivo), riduzione testuale.

Come si vede, in media il testo dei sottotitoli è lungo poco meno della metà rispetto al testo originale; inoltre, la percentuale di riduzione è relativamente costante, tra il 48% di S1 e il 56% di S5. È interessante notare che la percentuale di riduzione più bassa si registra sul discorso pronunciato alla velocità media più bassa (S1), mentre sui discorsi pronunciati a velocità maggiori (S6, S3 ed S5) la percentuale aumenta, anche se non in modo direttamente proporzionale (ci sono anche discorsi pronunciati a una velocità intermedia, come S4, in cui la percentuale di riduzione è stata piuttosto alta). Chiaramente, i fattori di difficoltà nel *respeaking* sono vari e la velocità è solo uno di essi: tuttavia, questi dati empirici sembrano suggerire che un'alta velocità dell'originale possa indurre il *respeaker* a tagliare il testo in lingua di arrivo. Naturalmente, dire che in media la riduzione del testo è del 53% non equivale a dire che i sottotitoli contengono solo il 47% del messaggio originale, poiché i due testi sono in lingue diverse. Di conseguenza, il dato grezzo non dice nulla sull'effettivo contenuto: per poter dire qualcosa in merito, è necessaria un'analisi dell'accuratezza dei sottotitoli mediante l'applicazione del modello NTR.

#### 4.1. Applicazione del modello NTR

La tabella 4 riporta le percentuali di accuratezza NTR ottenute per i 6 discorsi del corpus. Come si vede, in nessun caso il *respeaker* si è avvicinato alla soglia del 98%, considerata lo standard minimo di qualità nei pochi studi esistenti. Tuttavia, è interessante notare la grande continuità di rendimento del professionista, dato che le percentuali di accuratezza variano veramente di poco da un discorso all'altro, attestandosi intorno al 94% in quasi tutti i casi. L'altro aspetto da commentare è che l'accuratezza non sembra dipendere dalla velocità del discorso originale, dato che il tasso NTR più alto (94,63%) è stato ottenuto nel discorso più veloce di tutti (S6) e tassi simili sono stati raggiunti su discorsi pronunciati a velocità diverse (ad esempio, S3 e S4).

	<b>durata</b>	<b>OSR (wpm)</b>	<b>% NTR</b>
<b>S1</b>	17' 55''	100	94,09
<b>S2</b>	6' 59''	139	94,23
<b>S3</b>	10' 12''	150	94,35
<b>S4</b>	22' 48''	127	94,47
<b>S5</b>	17' 17''	149	93,84
<b>S6</b>	19' 18''	153	94,63
<b>MEDIA</b>	15' 1''	136	94,27

Tabella 4  
Percentuali NTR raggiunte nel corpus.

Scendiamo adesso nel dettaglio per quanto riguarda gli errori principali riscontrati durante l'analisi NTR. Se calcoliamo i punti detratti dal conteggio per errori di tutti i tipi, troviamo che quelli che hanno influito maggiormente sull'abbassamento della percentuale di accuratezza NTR nel corpus sono stati, nell'ordine, le omissioni e le sostituzioni, seguiti dagli errori di correttezza grammaticale (che comprende vari aspetti, tra cui ad esempio la mancata concordanza tra soggetto e verbo o tra sostantivo e aggettivo, l'uso errato di tempi verbali, e soprattutto l'uso errato o l'assenza della punteggiatura). Come era prevedibile, quindi, gli errori di contenuto hanno avuto un peso molto superiore rispetto a quelli di forma e di riconoscimento (Grafico 1).

### % detrazione punti per tipologia di errore

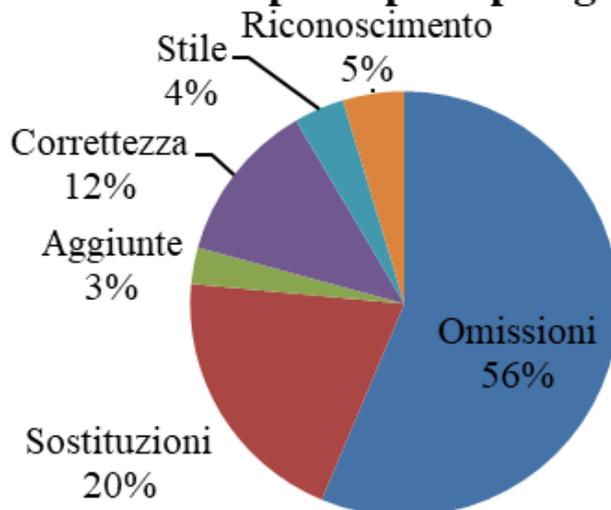


Grafico 1  
Distribuzione errori.

La tabella 5 riporta il dettaglio dei punti persi in ciascun discorso per ciascuna categoria di errore, il totale e la media. Chiaramente, occorre ricordare anche la diversa lunghezza degli interventi nel confrontare un discorso con l'altro (si veda la Tabella 2).

	S1	S2	S3	S4	S5	S6	TOT.	MEDIA
<b>Errori di contenuto</b>								
Omissioni	-30,25	-19,75	-21,75	-44,25	-40,25	-46,75	<b>-203</b>	<b>-33,71</b>
Sostituzioni	-13,25	-3,75	-11	-10,75	-16,75	-17,25	<b>-72,75</b>	<b>13,25</b>
Aggiunte	-1	-0,75	-2,25	-2,75	-2,5	-0,75	<b>-10</b>	<b>1,5</b>
<b>Errori di forma</b>								
Correttezza	-5,75	-2	-5,5	-11,75	-6,25	-13	<b>-44,25</b>	<b>7,38</b>
Stile	-2,75	0	-0,5	-4,75	-2	-3,75	<b>-13,75</b>	<b>2,29</b>
<b>Riconoscimento</b>								
	-2,75	-1,25	-1,25	-2,75	-3,25	-5,5	<b>-16,75</b>	<b>2,79</b>
<b>EE</b>								
	-7	0	0	-9	-3	-15	<b>-34</b>	<b>5,66</b>

Tabella 5

Detrazioni di punteggio per tipologia di errore.

I dati relativi agli errori di riconoscimento sono molto positivi, e dimostrano un'ottima interazione con il software: grazie alla sua ricca esperienza di *respeaking* intralinguistico, il professionista era abituato ad articolare molto chiaramente i suoni, aveva un buon ritmo ed è riuscito a evitare molti problemi; inoltre, il software era stato ben addestrato in precedenza, anche con l'inserimento nel vocabolario alcuni termini specifici per il Simposio, grazie ai materiali ricevuti in anticipo per la preparazione.

Più significativo invece il peso degli errori relativi alla correttezza grammaticale, e in particolar modo gli errori di punteggiatura: l'aggiunta della punteggiatura orale è infatti uno degli scogli maggiori del *respeaking*, che può mandare in saturazione le risorse cognitive del *respeaker* a causa del *multi-tasking*. Si tratta di un aspetto su cui bisogna insistere nei corsi di formazione: è utile focalizzarsi su questo aspetto prima in esercizi intralinguistici e poi aggiungere anche l'ostacolo della traduzione.

Per quanto riguarda gli errori di contenuto, come si vede, le omissioni hanno fatto perdere quasi il triplo dei punti delle sostituzioni (-203 contro -72,75). Inoltre, si nota che nel corpus esaminato il numero di *effective editions* (EE) è molto variabile e non è direttamente correlato con la durata dei discorsi o la loro velocità: il numero più alto di EE è stato riscontrato nel discorso S6 e non in S4, più lungo di ben 3 minuti e mezzo. È però interessante che S6 sia anche il discorso più veloce del corpus (153 wpm) e anche quello in cui la quantità di punti persi a causa delle omissioni è la più alta (-46,75). Parrebbe quindi che in questo caso il *respeaker* sia stato costretto a tagliare molto il testo, a volte sacrificando parte delle informazioni (omissioni) e a volte riuscendo invece a sintetizzare con successo (EE); un meccanismo simile sembra esserci stato anche nella sottotitolazione di S3, con un alto numero di punti persi a causa delle omissioni ma anche il secondo numero più alto di EE.

Dato il peso che omissioni e sostituzioni hanno avuto nel calcolo della percentuale di accuratezza NTR, nel sottoparagrafo seguente le analizziamo più dettagliatamente.

#### **4.2. Analisi delle omissioni e delle sostituzioni**

La tabella 6 riporta la distribuzione delle omissioni e delle sostituzioni tra i tre livelli di gravità (da minore a critico): i valori in tabella questa volta non rappresentano le penalizzazioni causate da tali errori nel conteggio NTR, ma proprio il numero di errori per ciascuna tipologia.

	S1	S2	S3	S4	S5	S6	TOT.	media
<b>Omissioni</b>								
<i>minori</i>	47	23	51	69	51	41	<b>282</b>	<b>47</b>
<i>maggiori</i>	37	24	18	50	53	71	<b>253</b>	<b>42,17</b>
<i>critiche</i>	0	2	0	2	1	1	<b>6</b>	<b>1</b>
<b>Sostituzioni</b>								
<i>minori</i>	19	7	10	21	11	14	<b>82</b>	<b>13,67</b>
<i>maggiori</i>	3	0	5	5	8	5	<b>26</b>	<b>4,33</b>
<i>critiche</i>	7	2	6	3	10	9	<b>37</b>	<b>6,17</b>

Tabella 6  
Omissioni e sostituzioni per livello di gravità.

Come sempre, il numero degli errori va rapportato con la lunghezza dei discorsi: non è un caso che il numero più basso di errori di tutti i tipi sia stato riscontrato in S2, che durava soltanto 7 minuti. A parte questo aspetto ovvio, è positivo notare che sia per le omissioni che per le sostituzioni la maggior parte degli errori è stata di gravità minore, cioè con un impatto lieve (282 e 82, rispettivamente).

Tuttavia, sono stati relativamente frequenti gli errori di traduzione molto gravi, cioè le sostituzioni critiche (37); inoltre, c'è un alto numero di omissioni di gravità maggiore, cioè intermedia (253), che generalmente corrispondono a intere unità di significato saltate (subordinate o coordinate). Non avendo avuto modo di intervistare il respeaker subito dopo la prestazione, non abbiamo informazioni sugli eventuali problemi che può aver incontrato durante lo svolgimento dell'incarico ed è quindi impossibile individuare le cause degli errori. Alcuni potrebbero essere stati causati da caratteristiche specifiche dei discorsi in esame (la velocità di eloquio elevata, l'oggettiva complessità dei temi proposti, la densità di informazioni in alcuni passaggi, la durata stessa di alcuni interventi, e così via), mentre altri da difficoltà del respeaker (la mancata comprensione di qualche termine, cali di concentrazione eccetera). Certamente la presenza massiccia di omissioni e di sostituzioni di entità critica indica una perdita di significato piuttosto netta.

Inoltre, nell'analisi si è notato che a volte l'omissione di intere unità di significato del testo originale si collega a errori di traduzione importanti nell'unità successiva e vice versa. Gli esempi qui di seguito riportati illustrano proprio questi due meccanismi.

L'esempio 1 della Tabella 7 (dal discorso S2) mostra un caso di sostituzione critica seguita da un'omissione maggiore. L'oratore originale sta parlando del funzionamento di un software per il riconoscimento del parlato in lingua ceca e delle difficoltà specifiche di questa lingua. I sottotitoli, invece di riportare il concetto delle ore di addestramento necessarie al software, fanno riferimento al vocabolario che deve essere più ampio di quello della lingua inglese (83): si tratta di una sostituzione critica, poiché il messaggio espresso in LA differisce in modo sostanziale dall'originale e il pubblico non ha modo di accorgersi di questa mancanza di fedeltà. Subito dopo, la parte in cui l'oratore descrive l'ampiezza del vocabolario in lingua ceca scompare completamente dai sottotitoli (omissione maggiore; 84), che nel complesso comunicano un messaggio lacunoso e fuorviante.

TESTO IN LP (LINGUA DI PARTENZA)	SOTTOTITOLI IN LA (LINGUA DI ARRIVO)
You have to know that Czech language is higher inflectional	ma visto che la lingua ceca fa molto uso di declinazioni e coniugazione [82]
so we need much more hours than English English language	ovviamente abbiamo bisogno di un vocabolario [83] più ampio rispetto alla lingua inglese.
although we u-we have over one million words in the vocabulary some words have are not covered by this vocabulary	[84]

Tabella 7

Sostituzione critica seguita da omissione maggiore (S2).

Ovviamente, non si può affermare che i due errori siano legati da un rapporto di causa-effetto, cioè che la sostituzione di gravità critica abbia causato l'omissione maggiore successiva, poiché non sappiamo cosa stesse pensando il respeaker in quel momento. Lo stesso dicasi per l'esempio 2 della Tabella 8 (dal discorso S4), in cui vediamo un'omissione maggiore seguita da una sostituzione di gravità critica. L'oratore sta descrivendo dei software per la sottotitolazione assistita sviluppati dalla sua azienda, e spiega che è stato creato un apposito modello per la valutazione dell'accuratezza, in cui si analizzano anche la segmentazione e la punteggiatura introdotte automaticamente dal software (senza intervento umano). Nel brano qui riportato, il respeaker ha tagliato completamente il riferimento ai due modelli già esistenti (WER e NER), producendo due omissioni maggiori (115 e 116). Subito dopo c'è una sostituzione critica (118): mentre l'originale sottolineava

che tali modelli non contemplano i parametri che a loro interessavano, i sottotitoli italiani parlano genericamente del fatto che i software dell'azienda *producono sottotitoli automatici* e che *siamo interessati a ottenere diversi risultati*. Al pubblico non arriva l'informazione relativa alla creazione di un nuovo modello di valutazione specifico per i sottotitoli automatici e il testo rimane oscuro.

TESTO IN LP (LINGUA DI PARTENZA)	SOTTOTITOLI IN LA (LINGUA DI ARRIVO)
but the point is that when you came to apply this kind of metrics to assisted subtitling there are limitations [114]	Per quanto riguarda questi sottotitoli assistiti, però, [114]
because in a way Word Error Rate is only taking into considerations the transcription errors	[115]
and NER is a little bit better because it takes also into account edition errors	[116]
but again there is this a limited scope for for transcription and for the parameters that we were interested into [117]	bisogna prendere in considerazione il fatto che i software producono sottotitoli automatici e quindi [117] visto che noi siamo interessati a ottenere diversi risultati [118] [119]
So that's why we introduced into into NER a number of new parameters	ovviamente dobbiamo anche adattare questi strumenti di valutazione [120] per quelli che sono i nostri scopi. [121]

Tabella 8  
Omissione maggiore seguita da sostituzione critica (S4).

## 5. Conclusioni

Come si è visto, il respeaking interlinguistico è una tipologia di sottotitolazione che ha molti aspetti in comune con l'interpretazione simultanea. Tuttavia, c'è un'importante differenza: sebbene anche l'interpretazione simultanea sia resa possibile dalla tecnologia (l'impianto con la console, la cabina, i microfoni, i ricevitori per il pubblico), il testo di arrivo è prodotto interamente da un essere umano e gli strumenti sono solo il mezzo attraverso il quale la traduzione dell'interprete arriva al pubblico. Al contrario, nell'IRSP il respeaker traduce simultaneamente il discorso originale, ma questo testo intermedio orale è pensato per il software di riconoscimento del parlato: il testo scritto finale che arriva al pubblico può quindi contenere alcuni errori di traduzione del respeaker, ma anche errori introdotti dal software in fase di riconoscimento.

Per tenere conto di questa particolarità, è stato proposto di recente il modello NTR per la valutazione dell'accuratezza dei sottotitoli prodotti in tempo reale mediante IRSP, finora applicato solo in studi sperimentali. Il presente lavoro ha presentato uno studio di caso relativo a un evento dal vivo nel quale è stato fornito un servizio di sottotitolazione in diretta mediante IRSP dall'inglese all'italiano. Dopo aver creato un archivio multimediale contenente tutte le registrazioni e trascrizioni del Simposio, sono state selezionate 6 relazioni congressuali in inglese con relativi sottotitoli italiani realizzati mediante IRSP, ed è stato applicato il modello NTR.

L'analisi dei dati ha rivelato un tasso medio di accuratezza del 94,27%, ben al di sotto del 98% suggerito come soglia minima. Tali percentuali sono più basse rispetto a quelle ottenute dai migliori studenti che hanno frequentato il corso online descritto da Dawson, Romero-Fresco (2021), in cui circa il 40% dei partecipanti ha raggiunto o superato un'accuratezza del 98%. Risultano invece simili alle percentuali ottenute dai migliori studenti che hanno partecipato allo studio pilota per il progetto *SMART* (Davitti, Sandrelli, 2020), attestatesi appunto sul 94%: in questo caso si trattava di studenti privi di esperienza che avevano seguito un corso di respeaking di poche ore. Invece, nel presente lavoro il respeaker aveva esperienza pluriennale di respeaking e di interpretazione simultanea, ma prima del Simposio aveva provato pochissime volte la variante interlinguistica: la mancanza di esperienza può avere perciò influito sull'accuratezza. Complessivamente, i risultati di questi tre studi sembrano indicare che, per raggiungere un livello di accuratezza paragonabile allo standard minimo nei sottotitoli intralinguistici, è imprescindibile una formazione apposita in IRSP.

Tuttavia, occorre far notare che in realtà i tassi NTR raggiunti nel presente studio non sono direttamente paragonabili con quelli dei due studi precedenti, poiché ci sono alcune importanti differenze. La prima è la durata dei testi utilizzati, che varia da quelli brevissimi di *ILSA* (circa due minuti) a quelli di durata intermedia di *SMART* (rispettivamente, 8 e 9 minuti e mezzo), fino ad arrivare a quelli del nostro corpus che andavano dai 7 ai quasi 23 minuti. Banalmente, in un testo brevissimo si hanno anche minori possibilità di commettere errori, il che potrebbe aver contribuito agli alti tassi di accuratezza registrati in *ILSA*. Inoltre, nel nostro studio abbiamo analizzato la prestazione di un unico respeaker su un corpus di 6 discorsi, mentre negli altri due studi sono state analizzate le prestazioni di 44 e 25 soggetti diversi (rispettivamente per *ILSA* e *SMART*). Infine, la maggiore differenza è la natura dei testi di partenza: mentre i due studi sperimentali hanno utilizzato video pedagogici (cioè discorsi realizzati appositamente a scopo didattico per la formazione di interpreti), quelli del presente studio erano relazioni congressuali autentiche, con conseguente maggiore complessità dal punto di vista della tematica e della densità informativa. Inoltre, nel nostro caso i testi

degli oratori erano in realtà multimodali, cioè comprendevano sia il discorso orale da essi pronunciato che le diapositive *PowerPoint* mostrate al pubblico: di conseguenza, il pubblico leggeva i sottotitoli, ma guardava anche le figure, i grafici e così via; allo stesso modo, il respeaker doveva anche monitorare la proiezione dei sottotitoli sulle diapositive. Per poter valutare più precisamente l'efficacia del servizio di IRSP dal punto di vista delle informazioni effettivamente trasmesse, sarebbe stato necessario fare anche uno studio di ricezione sul pubblico presente al Simposio quel giorno: ad esempio, si sarebbe potuto valutare se l'omissione di alcuni dettagli dai sottotitoli era stata compensata in qualche caso dalla presenza delle stesse informazioni nelle diapositive mostrate dall'oratore. Di conseguenza, per una valutazione più completa del servizio di IRSP in eventi dal vivo di questo genere si potrebbe forse pensare a un adattamento del modello NTR che tenga conto della natura multimodale dei testi di partenza.

Nonostante tutte le differenze summenzionate, è interessante vedere che gli errori più frequenti riscontrati in tutti e tre gli studi si sono rivelati, nell'ordine, le omissioni e le sostituzioni, seppure con un peso diverso.<sup>12</sup> Questo risultato, in linea con tutti gli studi precedenti che indicano nelle omissioni il problema principale del respeaking interlinguistico, suggerisce che nella formazione devono rivestire un ruolo chiave gli esercizi di analisi del testo (per imparare a distinguere gli elementi principali da quelli secondari) e quelli di riformulazione (alla ricerca della soluzione linguisticamente più sintetica). In altre parole, i respeaker devono riuscire a incrementare gli interventi efficaci sul testo (*EE*), per poter liberare energie da dedicare ai problemi traduttivi, al multitasking, alla gestione dello stress e così via.

Trattandosi di uno studio di caso relativo a uno specifico evento dal vivo, il presente contributo non ha certo la pretesa di aver esaurito l'argomento, ma ha l'obiettivo di stimolare la riflessione su queste tematiche, sia per contribuire a sviluppare delle linee guida per chi dovrà organizzare tali servizi, sia per delineare una proposta formativa mirata.

**Bionota:** Annalisa Sandrelli è Ricercatore confermato in Lingua e Traduzione Inglese (LIN/12) presso la Facoltà di Interpretariato e Traduzione dell'Università degli Studi Internazionali di Roma- UNINT. Dopo essersi formata come interprete di conferenza presso l'Università di Trieste, ha insegnato presso le Università di Hull, Trieste e Bologna-Forlì. Ha esperienza di insegnamento di interpretazione consecutiva, dialogica e

<sup>12</sup> In particolare, le omissioni sono quasi il quadruplo delle sostituzioni, mentre nei due studi sperimentali su studenti il rapporto tra i due tipi di errori è più equilibrato: chiaramente, non deve stupire che un professionista abbia fatto, in proporzione, meno errori di traduzione degli studenti.

simultanea, di traduzione audiovisiva, audio descrizione e respeaking. I suoi interessi di ricerca comprendono l'inglese giuridico, i corpora, la traduzione audiovisiva e le nuove tecnologie per l'interpretazione. Ha partecipato a vari progetti finanziati dalla Commissione Europea sulla formazione di traduttori e interpreti giudiziari e al progetto *European Parliament Interpreting Corpus (EPIC)*, Università di Bologna; attualmente dirige l'unità di inglese del progetto *Osservatorio sull'Euroletto*. Dopo aver coordinato i progetti *DubTalk* e *TVTalk* su doppiaggio e sottotitolazione, coordina il progetto *¡Sub!: Localisation Workflows that Work* e partecipa come International Co-Investigator al progetto *SMART (Shaping Multilingual Access with Respeaking Technology)*, finanziato dallo *Economic and Social Research Council* e con capofila l'Università del Surrey.

**Recapito autrice:** [annalisa.sandrelli@unint.eu](mailto:annalisa.sandrelli@unint.eu)

## Riferimenti bibliografici

- Astuto A. 2014-2015, *Il respeaking in Italia: analisi di un genere televisivo*, Tesi di laurea non pubblicata, Università degli Studi Internazionali di Roma-UNINT.
- Davitti E. e Sandrelli A. 2020, *Embracing the complexity: a pilot study on interlingual respeaking*, in “JAT- Journal of Audiovisual Translation” 3 [1], pp. 103-139. <https://www.jatjournal.org/index.php/jat/article/view/135/40> (2.03.2021).
- Dawson H. 2019, *Feasibility, quality and assessment of interlingual live subtitling: a pilot study*, in “Journal of Audiovisual Translation” 2 [2], pp. 36-56. <http://www.jatjournal.org/index.php/jat/article/view/72/24> (16.11.2020).
- Dawson H. e Romero-Fresco P. 2021, *Towards research-informed training in interlingual respeaking: an empirical approach*, in “The Interpreter and Translator Trainer”.
- Gile D. 2015, *Effort models*, in Pöchhacker F. (a cura di), *Routledge Encyclopedia of Interpreting Studies* Routledge, London/New York, pp. 135-137.
- Luppino D. 2016-17, *Interpretazione simultanea e respeaking interlinguistico: due modalità di traduzione a confronto. Un contributo alla ricerca*, Tesi di laurea non pubblicata, Università degli Studi Internazionali di Roma-UNINT.
- Marchionne F. 2010-11, *Il respeaking interlinguistico. Sperimentazioni per una evoluzione della sottotitolazione in diretta*, Tesi di laurea non pubblicata, Università di Macerata.
- Moore Z. 2018, *Respeaking at live events: ensuring quality in diverse settings*, Relazione presentata al 6th International Symposium on Accessibility and Live Subtitling, Milano, 14 settembre 2018.
- Moore Z. 2020, *Fostering Access for All Through Respeaking at Live Events*, in “The Journal of Specialised Translation” 33, pp. 207-226. [https://jostrans.org/issue33/art\\_moore.php](https://jostrans.org/issue33/art_moore.php) (16.11.2020).
- Ofcom 2015, *Measuring Live Subtitling Quality. Results from the Fourth Sampling Exercise*. [https://www.ofcom.org.uk/data/assets/pdf\\_file/0011/41114/qos\\_4th\\_report.pdf](https://www.ofcom.org.uk/data/assets/pdf_file/0011/41114/qos_4th_report.pdf) (16.11.2020).
- Pöchhacker F. 2004, *Introducing Interpreting Studies*, Routledge, London/New York.
- Pöchhacker F. e Remael A. 2019, *New efforts? A competence-oriented task analysis of interlingual live subtitling*, in “Linguistica Antverpiensia” 18, pp. 130-143. <https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/515/471> (16.11.2020).
- Remael A., Van Waes L. e Leijtenet M. 2016, *Live Subtitling with Speech Recognition – How to Pinpoint the Challenges?*, in Abend-David D. (a cura di), *Media and Translation. An Interdisciplinary Approach*, Bloomsbury, London, pp. 120-147.
- Romero-Fresco P. 2009, *More Haste Less Speed: Edited vs. Verbatim Respeaking*, in “Vigo International Journal of Applied Linguistics (VIAL)” 6, pp. 109-133. <http://vialjournal.webs.uvigo.es/pdf/Vial-2009-Article6.pdf> (16.11.2020).
- Romero-Fresco P. 2011, *Subtitling Through Speech Recognition: Respeaking*, St Jerome, Manchester.
- Romero-Fresco P. e Martínez J. 2015, *Accuracy Rate in Live Subtitling- the NER Model*, in Díaz-Cintas J. e Baños Piñero R. (a cura di), *Audiovisual Translation in a Global Context. Mapping. An Ever-Changing Landscape*, Palgrave, London, pp. 28-50.
- Romero-Fresco P. e Pöchhacker F. 2017, *Quality Assessment in Interlingual Live Subtitling: the NTR Model*, in “Linguistica Antverpiensia” 16, pp. 149-167. <https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/438> (11.06.2020).

- Sandrelli A. 2020, *Interlingual Respeaking and Simultaneous Interpreting in a Conference Setting: A Comparison*, in “inTRAlinea”. <http://www.intralinea.org/specials/article/2518> (16.11.2020).
- Serafini G. 2014-2015, *La sottotitolazione interlinguistica in tempo reale: il caso Ericsson Olanda e l'esperimento presso il festival Sedificorto 2014*, Tesi di laurea non pubblicata, Università di Bologna.