

## 1949-2019: 70 anni di TA visti attraverso i dati utilizzati

Randy Scansani, Alberto Barrón-Cedeño – Dipartimento di Interpretazione e Traduzione, Università di Bologna, Campus di Forlì

---

Citation: Randy Scansani, Alberto Barrón-Cedeño (2020) “1949-2019: 70 anni di TA visti attraverso i dati utilizzati”, in Adriano Ferraresi, Roberta Pederzoli, Sofia Cavalcanti, Randy Scansani (eds.) *Metodi e ambiti nella ricerca sulla traduzione, l’interpretazione e l’interculturalità – Research Methods and Themes in Translation, Interpreting and Intercultural Studies*, *MediAzioni* 29: A124-A147, <http://www.mediazioni.sitlec.unibo.it>, ISSN 1974-4382.

---

La traduzione automatica (TA) ha subito diversi mutamenti dal 1940 ad oggi. Come in molti altri campi dell’informatica e dell’intelligenza artificiale, si è passati da risorse sviluppate ad hoc manualmente ad approcci basati sempre di più su dati preesistenti. Il presente contributo si propone di offrire una panoramica delle diverse architetture di TA e dei dati da esse richiesti, partendo dagli approcci *rule-based* e arrivando alle architetture statistiche, *example-based* e neurali. Ognuno di questi cambiamenti ha influito sulla tipologia di dati richiesti per la costruzione di motori di TA. Se i primi approcci non richiedevano l’utilizzo di frasi allineate, con la TA statistica è diventato imprescindibile poter fare affidamento su una grande quantità di dati paralleli. Oggi, grazie all’utilizzo delle reti neurali, è possibile ottenere una traduzione di buona qualità persino per combinazioni per cui non sono disponibili dati in entrambe le lingue.

### 1. Introduzione

A partire dalla sua nascita fino ad arrivare ad oggi, diversi cambiamenti hanno interessato la traduzione automatica (TA), portando allo sviluppo di diverse architetture. Da una parte, lo sviluppo di nuove architetture ha influito sul

prodotto finale dei sistemi, ovvero l'output. Dall'altra, con i sistemi più recenti è radicalmente cambiata la tipologia di dati richiesti da un sistema di TA per produrre l'output. Dai primi sistemi, che integravano informazioni linguistiche esplicite, si è passati a sistemi che sfruttano grandi quantità di testi precedentemente tradotti.

Il passaggio ad architetture basate su dati è stato reso possibile dall'espansione di internet e da un mercato sempre più globale, con la conseguente necessità di tradurre un volume sempre maggiore di testi. Inoltre, l'incremento nell'utilizzo di strumenti di traduzione assistita ha reso disponibile una più ampia quantità di testi allineati.

L'idea di applicare il computer alla traduzione di frasi risale al XX secolo. Uno dei primi metodi è quello proposto da Peter Troyanskii (Hutchins 1995). La sua idea scompone il processo traduttivo in tre fasi: nella prima un umano che conosce la lingua di partenza individua le forme base delle parole di un testo e descrive le loro proprietà sintattiche e grammaticali. Nella seconda fase una macchina traduce le forme base nella lingua di arrivo, e nell'ultima fase l'intervento umano permette di assemblare la traduzione grazie alle parole tradotte dalla macchina ed alle informazioni raccolte nella prima fase. Benché rudimentale, questo approccio è indubbiamente pionieristico nella sua scelta di integrare il lavoro di una macchina con informazioni linguistiche individuate da un umano.

Nel 1949, data convenzionalmente utilizzata per indicare la nascita della TA, Warren Weaver pubblica un Memorandum in cui propone l'utilizzo dei computer per tradurre tra due lingue<sup>1</sup>. Secondo Weaver, i progressi nella codifica di codici (ottenuti soprattutto durante la Seconda guerra mondiale) e l'individuazione di principi logici e linguistici comuni a diverse lingue avrebbero potuto aiutare nello sviluppo di sistemi di TA, nonostante l'ostacolo delle ambiguità lessicali (Hutchins e Somers 1992).

---

<sup>1</sup> Il Memorandum di Weaver può essere consultato al seguente link: <http://www.mt-archive.info/Weaver-1949.pdf>.

Anche la prima dimostrazione di un sistema di TA (tenuta nel 1954 e frutto della collaborazione tra la Georgetown University e IBM) sfrutta informazioni linguistiche. Nella dimostrazione, infatti, 49 frasi vengono tradotte dal russo in inglese utilizzando un dizionario di 250 parole e sei regole linguistiche (Garvin 1968). Va specificato che, nell'ambito della TA, quando si fa riferimento a dizionari si intende una raccolta di corrispondenze tra una parola e una o più informazioni linguistiche (ad esempio il traduttore, oppure informazioni sintattiche, morfologiche, ecc.) in base all'architettura e al sistema, come si vedrà in seguito. La combinazione linguistica utilizzata per questo esperimento suggerisce quale fattore abbia dato un ulteriore impulso alla ricerca nel campo della TA, ovvero la Guerra Fredda.

I sistemi basati su regole che permettono il trasferimento del testo dalla lingua di partenza a quella di arrivo vengono definiti *rule-based* (RBMT). Risulta intuitivo comprendere come questi sistemi non facciano uso di dati bilingue nel senso che generalmente oggi viene attribuito a questo termine nel campo della TA, cioè raccolte di testi o frasi in due lingue. Poco più di 30 anni dopo le prime implementazioni, la TA cambia radicalmente pelle con lo sviluppo delle prime architetture statistiche (Berger *et al.* 1994; Koehn 2010), che si servono di corpora analizzati tramite algoritmi per creare modelli statistici. Questo momento segna il passaggio alle architetture data-driven, cioè basate su dati preesistenti. Dell'universo data-driven fanno parte anche i sistemi *example-based* (EBMT) (Nagao 1984) e i più recenti sistemi neurali (NMT) (Kalchbrenner e Blunsom 2013; Sutskever *et al.* 2014) in cui reti neurali vengono addestrate utilizzando corpora bilingue per "imparare" a tradurre una frase nella lingua di arrivo.

Se è vero che i mutamenti nelle architetture di TA sono stati accompagnati (e per certi versi resi possibili) da un aumento della quantità di dati a disposizione, dall'altra parte questo incremento non si è verificato per tutte le combinazioni linguistiche o per tutti i domini testuali. Ad esempio, la dimensione dei corpora disponibili nella raccolta Opus<sup>2</sup> (Tiedemann 2012), presumibilmente la più

---

<sup>2</sup> <http://opus.nlpl.eu/>.

utilizzata per estrarre dati per la costruzione di motori di TA, varia notevolmente in base alla combinazione linguistica selezionata. Inoltre, alcuni domini (ad esempio legale, medico, informatico) sono ben rappresentati, ma la varietà è ridotta (sono assenti, ad esempio, corpora specifici nei domini energetico, astronomico o della chimica). L'assenza di ingenti quantità di dati in alcuni domini si scontra con la necessità, da parte dei sistemi di TA data-driven, di essere addestrati su grandi volumi affinché i modelli costruiti siano rappresentativi dei testi da tradurre (Koehn e Knowles 2017). Per questo il mondo della ricerca sulla TA ha proposto diversi metodi di *domain adaptation*, cioè processi per adattare ad un dominio specifico un motore di TA addestrato su un'ampia quantità di dati generici.

Il presente contributo si propone di riattraversare i circa 70 anni trascorsi dalla data scelta convenzionalmente come anno di nascita della TA, focalizzandosi sull'evoluzione dei dati utilizzati. Se infatti numerosi contributi si sono concentrati sulla storia della traduzione automatica per ripercorrerne le tappe fondamentali o introdurre i principali meccanismi (Hutchins 1986, 1995; Gaspari e Hutchins 2007; Hutchins e Somers 1992), gli autori non sono a conoscenza di contributi che abbiano provato a mettere in luce le diverse tipologie di dati utilizzati dalle differenti architetture, discutendo come questi abbiano contribuito a rendere più o meno realizzabile lo sviluppo di un'architettura.

Nel presente contributo verrà offerta una panoramica dei dati necessari per addestrare i vari tipi di sistemi di TA: la Sezione 2 si occupa dei sistemi basati su regole, la Sezione 3 descrive i sistemi example-based. Le due architetture stato dell'arte nella TA degli ultimi anni, cioè l'architettura statistica e quella neurale, vengono descritte rispettivamente nelle Sezioni 4 e 5, insieme ai metodi adottati per applicare la TA anche a scenari caratterizzati da una scarsità di dati.

## **2. Sistemi *rule-based***

I sistemi *rule-based* basano il proprio funzionamento su regole. Secondo Hutchins e Somers (1992), il primo approccio appartenente a questa categoria

è conosciuto come approccio diretto e consiste in una semplice sostituzione delle parole nella lingua di partenza tramite i loro equivalenti nella lingua di arrivo servendosi di un dizionario. I limiti dell'approccio diretto sono evidenti: ad esempio, non vengono prese in considerazione le differenze sintattiche tra due lingue e non è previsto alcun metodo per risolvere eventuali ambiguità lessicali.

Il sistema utilizzato per la dimostrazione Georgetown-IBM (Sezione 1), sfruttava un dizionario russo-inglese composto da 250 parole. Le parole nella lingua di partenza erano divise in radici e desinenze, ciascuna delle quali appariva solo una volta e con una sola traduzione. A dimostrazione del fatto che i problemi di ambiguità sono di difficile risoluzione con un approccio di questo tipo, ai fini di questa dimostrazione le frasi da tradurre vennero accuratamente selezionate per evitare espressioni ambigue.

Un esempio simile, ma con un dizionario più ricco (170.000 voci) viene descritto da Reifler (1961). Nonostante un dizionario più ricco e un processo più dispendioso, l'autore definisce l'output ottenuto come "comprensibile".

Nei sistemi transfer, il testo di partenza viene prima analizzato dal punto di vista morfologico e sintattico, risolvendo eventuali ambiguità. Sulla base dell'analisi, il testo viene trasferito nella lingua di arrivo e, prima di produrre la traduzione, i vari elementi della frase vengono riordinati in base alla sintassi della lingua di arrivo. È sorprendente osservare come l'approccio transfer si basi su alcune delle idee precedentemente proposte da Troyanskii e Weaver (Sezione 1). Un esempio di sistema transfer è Systran (Toma *et al.* 1970; 1972), sviluppato a partire dagli anni 70 per la combinazione linguistica russo-inglese e utilizzato, a partire dal 1976, dalla Commissione Europea con l'aggiunta di altre combinazioni linguistiche, come francese-inglese e italiano-inglese (Hutchins 1986). Systran si compone di algoritmi che si occupano dell'analisi della lingua di partenza e della generazione nella lingua di arrivo. Alcuni di questi sono applicabili a qualsiasi combinazione linguistica, mentre altri sono sviluppati ad hoc per determinate combinazioni. Il processo traduttivo è basato su di un ampio e complesso dizionario con corrispondenze tra la lingua di partenza e quella di arrivo, che nel 1972 (per la combinazione russo-inglese) si componeva di 144.000 radici, 160.000 parole composte e 3.300 tra forme idiomatiche ed

espressioni composte da più parole (Toma *et al.* 1972). In ogni voce erano inoltre immagazzinate informazioni linguistiche (morfologia, semantica, parti del discorso, ecc.).

Un esempio peculiare di architettura *rule-based* è Météo (Chandioux 1976; Chandioux e Guéraud 1981), sistema sviluppato dal gruppo di ricerca TAUM di Montréal per tradurre bollettini metereologici dall'inglese al francese. Météo non è definibile come architettura transfer in senso stretto (Hutchins e Somers 1992; Hutchins 1986). Infatti, il suo processo traduttivo comprendeva la consultazione dei dizionari (come negli approcci diretti) e alcuni passaggi per l'analisi dell'input e la generazione dell'output. Météo si componeva di due dizionari (Chandioux 1976): uno, con 300 voci, dedicato alla traduzione di idiomi (che vengono definiti come gruppi di parole nel testo di partenza che vanno tradotti con una parola singola nel testo di arrivo) e alla traduzione di nomi propri. Nel dizionario principale, invece, ad ogni parola inglese corrispondeva la traduzione francese con le sue informazioni grammaticali, semantiche e morfologiche. Il dizionario principale conteneva una voce per ogni forma flessa, per un totale di 1.200 voci. In tutto, quindi, Météo faceva riferimento a 1.500 voci divise in due dizionari, un numero che, se confrontato con i più di 300.000 elementi nei dizionari di Systran, rende l'idea di quanto le differenze tra domini (o il numero di domini a cui i sistemi sono applicati) influenzino la composizione dei dizionari delle architetture RBMT. A partire dagli anni 80, la ricerca nel campo della TA si è focalizzata sempre di più sulla possibilità di creare sistemi che fossero completamente indipendenti dalla combinazione linguistica. È il caso dell'approccio basato sull'interlingua (Tomita e Carbonell 1987; Nyberg III e Mitamura 1992). L'interlingua è un linguaggio ideato con l'intento di creare una rappresentazione astratta di un testo a livello semantico. L'approccio interlingua si fondava quindi sull'idea di trasferire il contenuto del testo di partenza in una rappresentazione semantica astratta che potesse poi essere trasferita alla lingua di arrivo. Da un punto di vista pratico, l'interlingua è composta da un insieme di mappature a livello di parola tra una lingua di partenza e l'interlingua stessa e tra quest'ultima e una lingua target (Hutchins 1986).

Tuttavia, la complessità di sviluppare un'interlingua in grado di rappresentare a

livello semantico qualsiasi contenuto in qualsiasi lingua fu un ostacolo per i progetti di ricerca che perseguivano questo obiettivo, e nessun sistema basato su questo approccio venne sviluppato e utilizzato su ampia scala (Hutchins 1986; Hutchins e Somers 1992). Un esempio di sistema interlingua è quello descritto da Nyberg III e Mitamura (1992), chiamato KANT e che traduceva dall'inglese in giapponese, francese e tedesco. KANT è stato sviluppato per tradurre esclusivamente nel dominio della manualistica elettronica, e si basava su un dizionario limitato di 14.000 concetti generici, su un set di diverse centinaia di termini di dominio e su un numero limitato di strutture sintattiche traducibili. Per l'utilizzo di KANT era previsto l'uso di un linguaggio controllato<sup>3</sup>. Per la costruzione dei suoi dizionari, KANT faceva affidamento su un procedimento semi-automatico che, partendo da corpora, creava rappresentazioni di una parola che contenessero informazioni semantiche e grammaticali. Una serie di regole di mapping, in seguito, collegava queste rappresentazioni ad una rappresentazione nell'interlingua. Il processo veniva ripetuto in ordine inverso per il passaggio dall'interlingua alla lingua di arrivo. Se da una parte l'interlingua ha il vantaggio di poter essere applicata a qualsiasi lingua, dall'altra parte non solo svilupparla è molto complesso, ma le mappature che collegano una lingua all'interlingua sono ovviamente diverse per ogni lingua e anche il loro sviluppo ha un alto coefficiente di difficoltà. Non a caso, il sistema KANT è stato sviluppato per tradurre in un dominio ben definito e utilizzando il linguaggio controllato.

Dopo aver analizzato i dizionari in tutte le architetture *rule-based*, risulta più semplice comprendere che il passaggio ad architetture *data-driven* è stato motivato non solo dalla crescente diffusione di una quantità sempre maggiore di dati bilingue, ma anche dalla difficoltà nel migliorare sensibilmente la qualità dell'output utilizzando approcci basati su regole linguistiche esplicite. Inoltre, i sistemi diretti e *transfer* non sono scalabili (cioè applicabili a diverse combinazioni linguistiche senza doverne rivoluzionare la struttura) poiché implicano lo sviluppo di regole e dizionari differenti per ogni combinazione o

---

<sup>3</sup> Il linguaggio controllato è una lingua artificiale in cui la variabilità sintattica, grammaticale e semantica di una lingua è tenuta sotto controllo per facilitare la leggibilità del testo (Gobbi 2014: 7-19).

addirittura direzione linguistica<sup>4</sup>. I sistemi interlingua, che avrebbero potuto risolvere i problemi di scalabilità, si sono dimostrati di difficile realizzazione da un punto di vista pratico.

### **3. Sistemi *example-based***

Aggirando l'ostacolo della creazione di regole e dizionari, a partire dalla metà degli anni 80 circa nacquero i primi progetti di ricerca mirati a riutilizzare frammenti di traduzioni esistenti per tradurre nuovi testi in modo automatico. Un esempio è la TA *example-based* (Nagao 1984).

Questo approccio si fondava sull'uso di ampi corpora bilingue paralleli, ovvero allineati a livello di segmento. I sistemi EBMT cercavano analogie tra la frase che doveva essere tradotta e le frasi – nella stessa lingua – presenti nel corpus o nei corpora paralleli a loro disposizione. Per scovare queste analogie il sistema, oltre ai corpora paralleli, si serviva di dizionari e database con informazioni semantiche e lessicali (come thesauri o ontologie). In uno degli esempi riportati da Nagao (1984), il sistema doveva tradurre la frase inglese “A man eats vegetables” in giapponese. Nei corpora paralleli era presente la frase inglese “He eats potatoes” con la corrispondente traduzione giapponese. Nonostante il soggetto e il complemento oggetto siano diversi, le due frasi sono equivalenti da un punto di vista sintattico e simili da un punto di vista semantico. Il sistema procede per questo ad utilizzare la traduzione giapponese di “He eats potatoes” presente nel corpus parallelo, servendosi del dizionario per sostituire la traduzione di “A man” e “vegetables” con la traduzione giapponese di “He” e “potatoes” rispettivamente. Rispetto ai sistemi *rule-based* (Sezione 2), i sistemi EBMT facevano affidamento su database già esistenti e non richiedevano processi manuali per la costruzione di dizionari o la formulazione di regole linguistiche esplicite. Come si vedrà nelle sezioni 4 e 5, questo vantaggio è comune a tutte le architetture data-driven. L'altro grande vantaggio portato dalle

---

<sup>4</sup> La direzione linguistica esplicita da e verso quale lingua si traduce all'interno di una combinazione. Ogni combinazione linguistica include due direzioni.



architetture data-driven è la scalabilità dei suoi sistemi. Un sistema può essere applicato a qualsiasi combinazione linguistica perché non integra informazioni linguistiche esplicite, ma le estrapola dai dati a disposizione. Infine, i sistemi EBMT basano la traduzione su corrispondenze tra frammenti di una frase invece che su singole parole, prendendo maggiormente in considerazione il contesto rispetto a sistemi basati sulla traduzione di parole.

Dall'altra parte, ogni frase può essere formulata in molti modi diversi. Questo, volendo basare la traduzione sulle analogie tra frasi, implica la necessità di avere a disposizione uno o più corpora particolarmente ampi e con una grande varietà semantica e sintattica, cosa che non sempre è possibile e non sempre risulta sufficiente. Inoltre, così come le architetture transfer, le architetture basate su esempi cercano di risolvere il complesso problema dell'ambiguità lessicale servendosi di dizionari. Nonostante questi limiti, l'idea di usare traduzioni già esistenti per produrne di nuove è stata indubbiamente innovativa e un importante primo passo verso le architetture descritte nelle prossime sezioni, che sono diventate lo stato dell'arte nel campo della TA.

#### **4. Sistemi statistici**

Come accennato nella Sezione 1, nei sistemi statistici corpora bilingue paralleli vengono sfruttati per costruire modelli statistici. L'unità fondamentale delle prime architetture statistiche, dette *word-based*, era la parola. In questi modelli venivano immagazzinate corrispondenze a livello di parola: dizionari che contenevano corrispondenze tra parole nella lingua di partenza e parole nella lingua di arrivo e la loro probabilità di traduzione. Successivamente, l'architettura statistica più utilizzata è diventata quella *phrase-based* (PBMT). Anche in quest'ultima si parte da allineamenti a livello di parola, ma questi allineamenti vengono poi sfruttati per creare corrispondenze tra frasi nelle due lingue. Si noti che per frasi si intendono semplicemente sequenze (di lunghezza variabile) di parole attigue, formate spezzando frasi di senso compiuto in più parti e non sulla base di regole linguistiche. L'utilizzo di corrispondenze a livello di frasi anziché a livello di parole permette alle

architetture *phrase-based* di tenere in considerazione anche il contesto in cui una parola compare, come visto per le architetture basate su esempi (Sezione 3).

Tuttavia, la disponibilità di ampi corpora bilingue allineati a livello di parola è indubbiamente molto limitata, e d'altra parte la creazione di questi allineamenti sarebbe decisamente dispendiosa. Il problema di partenza che i sistemi di TA statistica sono chiamati a risolvere è quindi quello di stimare modelli statistici per la traduzione tra due lingue in assenza di un dato fondamentale: l'allineamento a livello di parola. Questo problema è stato superato grazie ad un algoritmo in grado di calcolare, partendo dai dati bilingue a disposizione, la probabilità che una parola nella frase di partenza sia allineata ad una parola nella frase di arrivo. Il prodotto dell'algoritmo è un modello statistico per l'allineamento tra parole.

Una volta disponibile il modello per stimare gli allineamenti, viene creato il *translation model*, che è alla base sia dei modelli *word-based* (Berger *et al.* 1994) che di quelli *phrase-based* (Koehn *et al.* 2003b). Il *translation model* contiene tutte le occorrenze bilingue recuperate dai corpora insieme alla loro probabilità statistica. Nel caso dei modelli *word-based* (Brown *et al.* 1993) le occorrenze saranno a livello di parola, mentre nel caso dei *phrase-based* (Koehn *et al.* 2003a) saranno a livello di frasi di diversa lunghezza, ma sempre allineate a livello di parola. I sistemi statistici si servono inoltre di corpora monolingue nella lingua di arrivo (generalmente vengono usati il lato target dei corpora bilingue disponibili e ampi corpora monolingue se disponibili) per costruire il *language model*. Questo modello contiene gruppi di parole di diversa lunghezza individuati nella lingua di arrivo insieme alla probabilità che quella frase possa comparire nella lingua di arrivo.

I dizionari descritti nella Sezione 2 lasciano quindi il posto a diversi modelli statistici. Al momento della traduzione, il decoder si serve del *translation model* per recuperare tutte le occorrenze che potrebbero essere utilizzate per tradurre una frase. Il *language model* viene sfruttato per scegliere da una lista di possibili traducenti la parola corretta da generare sulla base delle parole precedentemente generate. L'obiettivo del *language model* è quello di formulare

un output più scorrevole possibile e di utilizzare la parola corretta in quel contesto, tenendo cioè in considerazione le parole precedentemente generate. Ad esempio, traducendo in italiano la frase “Click here to accept cookies” il *language model* dovrebbe essere in grado di assegnare un punteggio più alto al termine “cookies” in prestito dall’inglese rispetto alla traduzione italiana “biscotti” che sarebbe ovviamente sbagliata in questo contesto, ma dovrebbe anche evitare che vengano generate frasi sintatticamente errate come “Qui cliccare per accettare i cookies”.

Rispetto ai sistemi *rule-based* (Sezione 2), i sistemi statistici sono scalabili e quindi permettono di applicare lo stesso sistema a diverse combinazioni linguistiche, a patto che sia disponibile una quantità sufficiente di dati. Non è quindi necessario sviluppare una serie di regole linguistiche esplicite che permettano il trasferimento da una lingua di partenza ad una di arrivo. Dall’altra parte, i sistemi statistici presentano principalmente due problematiche che sono inerenti ai sistemi data-driven in generale. Per costruire modelli statistici affidabili è necessario avere un ampio quantitativo di dati. A causa della grande quantità di differenti sistemi sviluppati (per alcuni dei quali non sono disponibili informazioni approfondite sull’architettura) e dei divari nella quantità disponibile di dati nelle diverse combinazioni linguistiche, per i sistemi data-driven è decisamente più complicato dare un’indicazione precisa del quantitativo di dati utilizzati come fatto nella Sezione 2. Tuttavia è possibile affermare che i sistemi *phrase-based* che hanno raggiunto lo stato dell’arte sono stati addestrati almeno su diverse centinaia di migliaia di coppie di frasi (Bojar *et al.* 2016).

Oltre al problema relativo alla quantità di dati necessaria, c’è un problema legato alla loro variabilità linguistica in base al dominio. Se il testo da tradurre è diverso dai testi utilizzati per addestrare un sistema data-driven (ad esempio nello stile, nella sintassi o nella terminologia), diventa probabile che la qualità finale dell’output non sia accettabile. A differenza dei sistemi basati su regole, la cui complessità rendeva difficile l’adattamento ad un dominio, con i sistemi statistici diverse tecniche sono state sperimentate in questo senso. La procedura per adattare un sistema a un nuovo dominio è chiamata *domain adaptation*. Nel caso della TA statistica la *domain adaptation* ha mostrato

risultati convincenti (Pecina *et al.* 2015; Eck *et al.* 2004; Foster e Kuhn, 2007; Bertoldi e Federico 2009; Zhao *et al.* 2004). In generale, la *domain adaptation* avviene integrando grandi corpora generici con una minore quantità di dati appartenenti ad un dominio specifico. Eck *et al.* (2004); Zhao *et al.* (2004) hanno adattato un sistema *phrase-based* traducendo frasi di dominio con il *language model* generico ed estraendo automaticamente da ampi corpora frasi simili a quelle appena tradotte. Le frasi così estratte venivano incluse in un *language model* aggiuntivo, più piccolo del *language model* già esistente, ma popolato di dati più rilevanti.

Avendo a disposizione solo dati monolingue, la *domain adaptation* rimane limitata al *language model*. D'altra parte, la disponibilità di dati bilingue rilevanti per un dominio specifico è spesso limitata (Sezione 1). Per questo, alcuni metodi di *domain adaptation* si sono focalizzati sulla generazione di dati bilingue sintetici, e cioè traducendo un corpus monolingue di dominio con un sistema *phrase-based*. Questo metodo viene chiamato *backtranslation*. Nell'approccio descritto da Bertoldi e Federico (2009), sistemi *phrase-based* preesistenti sono stati utilizzati per tradurre verso l'inglese dati monolingue scritti in spagnolo e, viceversa, per tradurre verso lo spagnolo dati monolingue scritti in inglese, simulando una situazione in cui in un dominio specifico si disponga di soli dati monolingue. Questo contributo ha evidenziato che l'uso di dati bilingue sintetici può migliorare l'output, ma che i risultati migliori vengono ottenuti quando il solo *language model* viene adattato tramite l'uso di dati monolingue sintetici nella lingua di arrivo.

Sempre nel tentativo di migliorare l'aderenza dell'output ad un dominio, alcuni sistemi *phrase-based* hanno utilizzato tipologie di dati diverse dai corpora monolingue e bilingue (sintetici e non) citati fino a qui. Diverse tecniche per integrare raccolte di termini bilingue sono infatti state testate. Questi glossari bilingue, seppur di piccole dimensioni, permettono di migliorare l'output di sistemi di TA statistica (Pinnis e Skadinš 2012; Koehn *et al.* 2007; Arcan *et al.* 2014; Langlais 2002). Inserire terminologia di dominio può permettere ad un sistema addestrato su una grande quantità di dati di adattarsi ad un nuovo dominio testuale.

Il passaggio alle architetture data-driven ha quindi portato due principali vantaggi: una migliore qualità dell'output e una maggiore scalabilità. Allo stesso tempo, la qualità dell'output diventa fortemente dipendente dalla tipologia e quantità di dati usati per addestrare un motore. La minore complessità dei modelli utilizzati nella TA *phrase-based* rispetto ai precedenti sistemi ha permesso di testare, anche con risultati soddisfacenti, alcune tecniche per adattare ad un dominio un sistema addestrato su una grande quantità di dati generici.

## 5. Sistemi neurali

Con le architetture neurali la TA non consiste più nel ricombinare porzioni di testo precedentemente tradotte. Il *translation model* e il *language model* lasciano spazio a modelli basati su reti neurali che prendono come input una sequenza (nel caso della TA, una frase) e producono un output sotto forma, anch'esso, di sequenza (Kalchbrenner e Blunsom 2013; Sutskever *et al.* 2014).

Nel caso della TA la sequenza input è la frase di partenza e quella output è il suo equivalente nella lingua di arrivo. Un sistema di TA trasforma la sequenza input, una parola alla volta, in un vettore composto da numeri in grado di catturare le proprietà semantiche e sintattiche di ogni parola. Questo vettore viene poi trasformato nell'output generando una parola alla volta. La produzione di una parola output non dipende però solo dalle informazioni sulla parola stessa, ma anche dalle informazioni riguardanti le parole precedenti. Inoltre, il progredire delle architetture neurali ha portato all'utilizzo di reti neurali bidirezionali, grazie alle quali è stato possibile tenere in considerazione sia le parole precedenti sia quelle successive alla parola che sta per essere generata (Bahdanau *et al.* 2014).

Uno dei progressi portati dalle architetture statistiche *phrase-based* rispetto a quelle *word-based* è stata la capacità di considerare un contesto più ampio rispetto alla singola parola (Sezione 4). Le architetture neurali rappresentano un ulteriore passo avanti, poiché sono in grado di tenere in considerazione l'intero

contesto sia della frase di partenza che della frase di arrivo.

I miglioramenti dell'output ottenuti grazie all'utilizzo di reti neurali hanno portato anche qualche svantaggio in confronto alle precedenti architetture statistiche. Non solo i modelli neurali necessitano di un quantitativo ancora più ampio di dati, ma il loro funzionamento è difficile da decifrare e per molti versi le reti neurali sono tutt'ora una scatola nera. Ad oggi un grande sforzo viene prodotto nel mondo della ricerca per cercare di comprendere meglio e poter influenzare il processo che porta alla generazione dell'output; ad esempio per capire come ottenere una forma maschile anziché femminile o singolare anziché plurale di un aggettivo (Dalvi *et al.* 2019). Questo si contrappone al funzionamento più chiaro delle precedenti architetture: in quelle *rule-based* ogni dizionario aveva una funzione esplicita (ad esempio rimuovere le ambiguità) e nelle architetture statistiche ogni modello si occupava di una funzione (ad esempio il *language model* garantiva che venissero scelte le parole corrette dato il contesto).

Trattandosi di architetture data-driven, la tipologia di dati richiesta per la costruzione di motori di TA neurale non è cambiata molto rispetto ai dati usati per la TA statistica (Sezione 4): è necessaria sempre una buona quantità di corpora paralleli, e la disponibilità di dati di dominio è fondamentale. Aggiungendo al processo di apprendimento delle reti neurali qualche iterazione che coinvolga dati di uno specifico dominio è possibile adattare un motore generico al dominio in questione. È stato dimostrato che un numero limitato (anche qualche centinaio) di coppie di frasi può essere sufficiente per una *domain adaptation* efficace (Sennrich *et al.* 2015).

Mentre con le architetture *phrase-based* è stata dimostrata la possibilità di ottenere una *domain adaptation* sfruttando il solo *language model* (e quindi con dati monolingue, più facilmente reperibili), nelle architetture neurali il *language model* non è presente ed è sempre necessario disporre di dati bilingue. Per ovviare a questo, Gulcehre *et al.* (2015) ha integrato un *language model* basato su reti neurali ad un modello di TA neurale.

Per aumentare la quantità di coppie di frasi a disposizione partendo da dati monolingue, specialmente in contesti caratterizzati da una scarsità di dati, sono

state sperimentate tecniche di backtranslation (Sezione 4) con risultati convincenti. I primi esperimenti si sono focalizzati sulla produzione di frasi sintetiche nella lingua di partenza traducendo piccoli corpora monolingue nella lingua di arrivo (Sennrich *et al.* 2015). Ovviamente, il requisito fondamentale è quello di avere a disposizione un modello per tradurre verso la lingua di partenza. Queste tecniche sono poi state affinate, ad esempio creando diverse traduzioni sintetiche per una singola frase nella lingua di arrivo per diversificare le frasi sintetiche (Edunov *et al.* 2018), oppure etichettando le frasi generate in modo sintetico per segnalare alle reti neurali la necessità di trattare quelle frasi di partenza in modo differente rispetto alle frasi di partenza non sintetiche (Caswell *et al.* 2019).

La Sezione 4 ha illustrato tecniche per reperire frasi bilingue in base alla loro similarità con la frase da tradurre. Farajian *et al.* (2018) ha dimostrato che è possibile adattare un modello neurale addestrato su un grande quantitativo di dati bilingue generici utilizzando un limitato numero di frasi reperite dai dati di training in base alla loro similarità con la frase da tradurre.

La TA neurale permette di ottenere la *domain adaptation* non solo agendo sui dati a disposizione, ma anche agendo sul modello neurale stesso. Ad esempio, è possibile addestrare un classificatore che stimi il dominio di una frase e usi questa informazione per influenzare il modello neurale affinché dia un peso maggiore alle frasi di dominio Chen *et al.* (2017). Molti altri metodi per la *domain adaptation* sono stati testati. Chu e Wang (2018) ne offrono una panoramica esaustiva.

Non solo la TA neurale ha cambiato radicalmente il modo in cui i dati bilingue vengono sfruttati nel training o nella traduzione, ma ha anche reso possibile l'addestramento di un sistema utilizzando dati con diverse lingue di partenza o di arrivo. Questi sistemi, detti multilingue, possono avere diverse configurazioni. La configurazione *many-to-one* prevede che la rete neurale venga addestrata su molti corpora che abbiano la stessa lingua di arrivo, ma diverse lingue di partenza (ad esempio un corpus inglese-italiano e uno spagnolo-italiano). Unendo questi corpora, le reti neurali sono in grado di raccogliere un maggior quantitativo di informazioni sull'output che deve essere prodotto. Sarà il sistema

stesso ad imparare a distinguere le frasi in cui il testo di partenza è in inglese dai casi in cui è in italiano. In modo simile, le reti neurali hanno reso possibile la costruzione di sistemi *many-to-many* (Johnson *et al.* 2017).

Uno degli effetti più interessanti della possibilità di creare sistemi multilingue è la cosiddetta “zero-shot translation” (Johnson *et al.* 2017), che indica la capacità del modello neurale di tradurre in una combinazione linguistica che non ha mai incontrato nei dati di training. Se ad esempio si addestra un singolo sistema neurale per le seguenti combinazioni linguistiche:  $L_1-L_2$ ,  $L_1-L_3$ ,  $L_1-L_4$ ,  $L_2-L_4$ ,  $L_2-L_3$  e  $L_3-L_1$  il sistema non vedrà nessuna coppia di frasi per la combinazione  $L_3-L_4$ . Ciononostante, siccome la rete neurale condivide i parametri per i modelli di ogni combinazione linguistica, sarà in grado di tradurre anche nella combinazione  $L_3-L_4$ . Anche se ancora i sistemi *zero-shot* non producono un output di qualità comparabile ai sistemi addestrati su dati di una singola combinazione linguistica (Koehn 2019: 91), i loro risultati sono comunque promettenti specialmente per la traduzione da o verso lingue per cui la quantità di dati a disposizione è limitata o nulla. Inoltre, Lakew *et al.* (2018) hanno dimostrato che è possibile migliorare l’output di un modello *zero-shot* utilizzando la *zero-shot translation* per ottenere dati sintetici (cioè facendo la *backtranslation* di una parte dei dati di training) e poi riaddestrando il modello utilizzando sia i dati paralleli disponibili per il primo training, sia i dati sintetici.

Un approccio completamente differente da quelli descritti fino ad ora è la “unsupervised machine translation” (Artetxe *et al.* 2017; Lample *et al.* 2017; Artetxe *et al.* 2018), cioè una tipologia di TA neurale che sfrutta corpora comparabili. Per corpora comparabili si intendono raccolte di testi appartenenti alla stessa tipologia testuale in ciascuna delle due lingue, dove i testi in una lingua non sono la traduzione dei testi nell’altra lingua. Partendo dai testi monolingue vengono create le rappresentazioni vettoriali (menzionate all’inizio della sezione) delle parole di ciascuna delle due lingue separatamente. Diverse tecniche sono poi disponibili per calcolare la similarità tra una rappresentazione monolingue di una parola nella lingua di partenza e una rappresentazione monolingue di una parola nella lingua di arrivo e creare un collegamento tra quelle effettivamente simili, ad esempio “sedia” e “chair” in un sistema che



traduce dall'italiano in inglese. Pur dimostrando la fattibilità di costruire modelli usando soli dati comparabili, la TA neurale non supervisionata non si è ancora avvicinata alle prestazioni dei modelli neurali stato dell'arte (Artetxe *et al.* 2017; Lample *et al.* 2017). Risultati decisamente migliori sono stati ottenuti utilizzando sistemi di TA statistica in cui il translation model (Sezione 4) viene sostituito dal modello contenente le rappresentazioni vettoriali bilingue con la loro similarità (Artetxe *et al.* 2018). Benché incoraggianti, va detto che i risultati degli esperimenti citati sono stati ottenuti su combinazioni linguistiche dove le due lingue sono relativamente simili tra loro (francese-inglese e tedesco-inglese). Rimane quindi da capire che impatto possa avere la TA non supervisionata sulle combinazioni linguistiche dove le due lingue sono molto distanti tra loro e pochi dati sono a disposizione.

## **6. Conclusioni**

Sono trascorsi 70 anni dalla nascita della TA. Durante questo periodo si sono verificati diversi cambiamenti nelle architetture utilizzate, che hanno ovviamente avuto un riflesso sui dati da utilizzare per addestrare i sistemi. Si è passati da sistemi basati su dati prodotti in modo completamente manuale (o in alcuni casi semiautomatico) e ricchi di informazioni linguistiche esplicite a sistemi in grado di sfruttare grandi quantità di dati per ottenere informazioni con cui tradurre tra due (o addirittura più di due) lingue.

Nonostante questi cambiamenti, l'obiettivo di fondo per chi ha lavorato allo sviluppo di sistemi è rimasto sempre quello: trovare dati necessari per far sì che un algoritmo possa produrre una traduzione il più possibile corretta, possibilmente facendo in modo che questo algoritmo possa essere applicabile a più combinazioni.

L'evoluzione dei sistemi ha dimostrato un avvicinamento progressivo a questo obiettivo. I primi anni della TA hanno messo a nudo i grandi limiti delle architetture basate su regole. Oltre a produrre output di qualità modesta, i sistemi RBMT richiedono la costruzione di dizionari e la scrittura di regole per

ogni combinazione linguistica, che è un lavoro molto dispendioso. Con i sistemi data-driven si sono sicuramente ridotti i costi per produrre i dati di training, e la qualità dell'output è migliorata. Tuttavia, la disponibilità di dati rimane ad oggi un ostacolo per la costruzione di motori di TA applicabili a un alto numero di combinazioni linguistiche e domini, nonostante i progressi portati dalle architetture neurali e nonostante le tecniche sviluppate e testate per sopperire alla mancanza di dati bilingue.

## **Bibliografia**

Arcan M., M. Turchi, S. Tonelli e P. Buitelaar (2014) "Enhancing Statistical Machine Translation with Bilingual Terminology in a CAT Environment", in Y. Al-Onaizan e M. Simard (eds.) *Proceedings of AMTA 2014*, Vancouver: BC, 54-68.

Artetxe, M., G. Labaka e E. Agirre. (2018) "Unsupervised Statistical Machine Translation", in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Bruxelles: Association for Computational Linguistics, 3632-3642, <https://doi.org/10.18653/v1/D18-1399>.

Artetxe, M., G. Labaka, E. Agirre e K. Cho (2017) "Unsupervised Neural Machine Translation", ArXiv, CoRR abs/1710.11041, <http://arxiv.org/abs/1710.11041>.

Bahdanau, D., K. Cho e Y. Bengio (2014) "Neural Machine Translation by Jointly Learning to Align and Translate", ArXiv, CoRR abs/1409.0473, <http://arxiv.org/abs/1409.0473>.

Berger, A.L., P.F. Brown, S.A. Della Pietra, V.J. Della Pietra, J.R. Gillett, J.D. Lafferty, R.L. Mercer, H. Printz e L. Ures (1994) "The Candide System for Machine Translation", in *HUMAN LANGUAGE TECHNOLOGY: Proceedings of a Workshop held at Plainsboro* (New Jersey, 8-11 Marzo 1994), 157-162, <https://www.aclweb.org/anthology/H94-1028>.

Bertoldi, N. e M. Federico (2009) "Domain Adaptation for Statistical Machine

Translation with Monolingual Resources” in *Proceedings of the Fourth Workshop on Statistical Machine Translation*, Atene: Association for Computational Linguistics, 182-189, <https://www.aclweb.org/anthology/W09-0432>.

Bojar, O., R. Chatterjee, C. Federmann, Y. Graham, B. Haddow, M. Huck, A. Jimeno Yepes, P. Koehn, V. Logacheva, C. Monz, M. Negri, A. Neveol, M. Neves, M. Popel, M. Post, R. Rubino, C. Scarton, L. Specia, M. Turchi, K. Verspoor e M. Zampieri (2016) “Findings of the 2016 Conference on Machine Translation”, in *Proceedings of the First Conference on Machine Translation*, Berlino: Association for Computational Linguistics, 131-198, <http://www.aclweb.org/anthology/W/W16/W16-2301>.

Brown, P.F., S.A. Della Pietra, V.J. Della Pietra e R.L. Mercer (1993) “The mathematics of statistical machine translation: Parameter estimation”, in *Computational Linguistics* 19(2): 263-311, <https://www.aclweb.org/anthology/J93-2003>.

Caswell, I., C. Chelba e D. Grangier (2019) “Tagged Back-translation”, in *Proceedings of the Fourth Conference on Machine Translation, Volume 1*, Firenze: Association for Computational Linguistics, 53-63, <https://www.aclweb.org/anthology/W19-5206>.

Chandioux, J. (1976) “METEO : Un Système Opérationnel pour la Traduction Automatique des Bulletins Météorologiques Destinés au Grand Public”, *Meta* 21: 127-133.

Chandioux, J. e M.F. Guéraud (1981) “MÉTÉO : Un Système à l'Épreuve du Temps”, *Meta* 26: 18-22.

Chen, B., C. Cherry, G. Foster e S. Larkin (2017) “Cost Weighting for Neural Machine Translation Domain Adaptation”, in *Proceedings of the First Workshop on Neural Machine Translation*, Vancouver: Association for Computational Linguistics, 40-46, <https://doi.org/10.18653/v1/W17-3205>.

Chu, C. e R. Wang (2018) “A survey of domain adaptation for neural machine

translation”, in *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico: Association for Computational Linguistics, 1304-1319, <https://www.aclweb.org/anthology/C18-1111>.

Dalvi, F., N. Durrani, H. Sajjad, Y. Belinkov, D.A. Bau e J. Glass (2019) “What is One Grain of Sand in the Desert? Analyzing Individual Neurons in Deep Nlp Models”, in *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*, Honolulu, HA, 6309-6317.

Eck, M., S. Vogel e A. Waibel (2004) “Language Model Adaptation for Statistical Machine Translation Based on Information Retrieval”, in *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC’04)*, Lisbona: European Language Resources Association (ELRA), 327-330, <http://www.lrec-conf.org/proceedings/lrec2004/pdf/374.pdf>.

Edunov, S., M. Ott, M. Auli, e D. Grangier (2018) “Understanding Back-Translation at Scale”, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Bruxelles: Association for Computational Linguistics, 489-500, <https://doi.org/10.18653/v1/D18-1045>.

Farajian, A.M., N. Bertoldi, M. Negri, M. Turchi e M. Federico (2018) “Evaluation of Terminology Translation in Instance-based Neural MT Adaptation”, in *Proceedings of the 21<sup>st</sup> Annual Conference of the European Association for Machine Translation*, Alicante, 149-158.

Foster, G. e R. Kuhn (2007) “Mixture-model Adaptation for SMT”, in *Proceedings of the Second Workshop on Statistical Machine Translation*, Stroudsburg, PA: Association for Computational Linguistics, StatMT ’07, 128-135, <http://dl.acm.org/citation.cfm?id=1626355.1626372>.

Garvin, P.L. (1968) *The Georgetown-IBM Experiment of 1954: An Evaluation in Retrospect*, <https://books.google.it/books?id=wq5bPgAACAAJ>.

Gaspari, F. e J. Hutchins (2007) “Online and Free! Ten Years of Machine Translation: Origins, Developments, Current Use and Future Prospects”, in

*Proceedings of MT Summit XI*, Copenhagen, Denmark, 199-206.

Gobbi, I. (2014) *Spagnolo Tecnico Semplificato*, tesi di dottorato, <http://amsdottorato.unibo.it/6681/>.

Gulcehre, C., O. Firat, K. Xu, K. Cho, L. Barrault, H.C. Lin, F. Bougares, H. Schwenk e Y. Bengio (2015) "On Using Monolingual Corpora in Neural Machine Translation", ArXiv, abs/1503.03535.

Hutchins, J.W. (1986) *Machine Translation: Past, Present, Future*, New York: John Wiley & Sons, Inc.

Hutchins, J.W. (1995) "Machine Translation: A Brief History", in *Concise History of the Language Sciences: From the Sumerians to the Cognitivists*, Amsterdam: Pergamon Press, 431-445.

Hutchins, J.W. e H. Somers (1992) *An Introduction to Machine Translation*, Londra: London Academic Press.

Johnson, M., M. Schuster, Q.V. Le, M. Krikun, Y. Wu, Z. Chen, N. Thorat, F. Viégas, M. Wattenberg, G. Corrado, M. Hughes e J. Dean (2017) "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation", *Transactions of the Association for Computational Linguistics*, 5: 339-351, [https://doi.org/10.1162/tacl\\_a\\_00065](https://doi.org/10.1162/tacl_a_00065).

Kalchbrenner, N. e P. Blunsom (2013) "Recurrent Continuous Translation Models", in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, WA: Association for Computational Linguistics, 1700-1709.

Koehn, P. (2010) *Statistical Machine Translation*, New York: Cambridge University Press, 1° edizione.

Koehn, P. (2019) *Neural Machine Translation*, Cambridge: Cambridge University Press, 3° edizione.

Koehn, P. e R. Knowles (2017) "Six Challenges for Neural Machine

Translation”, in *Proceedings of the First Workshop on Neural Machine Translation*, Vancouver: Association for Computational Linguistics, 28-39, <https://doi.org/10.18653/v1/W17-3204>.

Koehn, P., F.J. Och e D. Marcu (2003a) “Statistical Phrase-based Translation”, in *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 127-133, <https://www.aclweb.org/anthology/N03-1017>.

Koehn, P., F.J. Och e D. Marcu (2003b) “Statistical Phrase-based Translation”, in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Volume 1*, Stroudsburg, PA: Association for Computational Linguistics, NAACL '03, 48-54, <https://doi.org/10.3115/1073445.1073462>.

Koehn, P., H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin e E. Herbst (2007) “Moses: Open Source Toolkit for Statistical Machine Translation”, in *Proceedings of the ACL 2007 Demo and Poster Sessions*, Praga: Association for Computational Linguistics, 177-180.

Lample, G., L. Denoyer, e M. Ranzato (2017) “Unsupervised Machine Translation Using Monolingual Corpora Only”, ArXiv, CoRR abs/1711.00043, <http://arxiv.org/abs/1711.00043>.

Langlais, P. (2002) “Improving a General-purpose Statistical Translation Engine by Terminological Lexicons”, in *COLING-02 on COMPUTERM 2002: Second International Workshop on Computational Terminology, Volume 14*, Stroudsburg, PA: Association for Computational Linguistics, COMPUTERM '02, 1-7, <https://doi.org/10.3115/1118771.1118776>.

Lakew, M. S., Q.F. Lotito, M. Negri, M. Turchi e M. Federico (2018) “Improving Zero-Shot Translation of Low-Resource Languages”, Arxiv, CoRR abs/1811.01389, <http://arxiv.org/abs/1811.01389>.

Nagao, M. (1984) “A Framework of a Mechanical Translation Between

Japanese and English by Analogy Principle”, in A. Elithorn e R. Barnerji (eds.) *Artificial and Human Intelligence*, Amsterdam: North-Holland, 173-180.

Nyberg, E.H. III e T. Mitamura (1992) “The KANT System: Fast, Accurate, High-quality Translation in Practical Domains”, in *COLING 1992, Volume 3*, 15th International Conference on Computational Linguistics, 1069-1073, <https://www.aclweb.org/anthology/C92-3168>.

Pecina, P., A. Toral, V. Papavassiliou, P. Prokopidis, A. Tamchyna, A. Way e J. van Genabith (2015) “Domain Adaptation of Statistical Machine Translation with Domain-focused Web Crawling”, *Language Resources and Evaluation*, 49(1): 147-193, <https://doi.org/10.1007/s10579-014-9282-3>.

Pinnis, M. e R. Skadinš (2012) “MT Adaptation for Under-resourced Domains - What Works and What Not”, in A. Tavast, K. Muischnek, M. Koit (eds.) *Human Language Technologies – The Baltic Perspective – Proceedings of the Fifth International Conference Baltic HLT 2012*, Tartu, Estonia, 176-184, <https://doi.org/10.3233/978-1-61499-133-5-176>.

Reifler, E. (1961) “The Machine Translation Project at the University of Washington, Seattle, Washington: Outline of the Project”, in *Mechanical Translation, Volume 6*, New York: Technology Press of the Massachusetts Institute of Technology e Wiley (pubblicazione congiunta), 25-32.

Sennrich, R., B. Haddow e A. Birch (2015) “Improving Neural Machine Translation Models with Monolingual Data”, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Volume 1*, Berlino: Association for Computational Linguistics, 86-96, CoRR abs/1511.06709. <http://arxiv.org/abs/1511.06709>.

Sutskever, I., O. Vinyals e Q. V. Le (2014) “Sequence to Sequence Learning with Neural Networks”, in *Proceedings of the 27th International Conference on Neural Information Processing Systems, Volume 2*, Cambridge, MA: MIT Press, NIPS’14, 3104-3112, <http://dl.acm.org/citation.cfm?id=2969033.2969173>.

Tiedemann, J. (2012) “Parallel Data, Tools and Interfaces in OPUS”, in

*Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul: European Language Resources Association (ELRA), 2214-2218.

Toma, P., L.A. Kozlik e D.G. Perwin (1972) *Optimisation of SYSTRAN system (1)*, New York: Rome Air Development Center, 1-18.

Toma P., L.A. Kozlik e D.G. Perwin (1970) *Systran Machine Translation System*, Technical Report AD0875290, Defense Technical Information Center.

Tomita, M. e J.G. Carbonell (1987) "The Universal Parser Architecture for Knowledge-Based Machine Translation", in *Proceedings of the 10th International Joint Conference on Artificial Intelligence, Volume 2*, San Francisco, CA: Morgan Kaufmann Publishers Inc., IJCAI'87, 718-721, <http://dl.acm.org/citation.cfm?id=1625995.1626020>.

Zhao, B., M. Eck e S. Vogel (2004) "Language Model Adaptation for Statistical Machine Translation via Structured Query Models" in *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*. Ginevra: COLING, 411-417, <https://www.aclweb.org/anthology/C04-1059>.