

Modulo A

L'italiano fuori d'Europa

17. Gli sviluppi
possibili

11 dicembre 2019



Linguistica italiana II
Mirko Tavosanis
A. a. 2019-2020

Traduzione

- Google Traduttore, DeepL e Microsoft Translator offrono traduzioni di alto livello
- Nella pratica, la traduzione automatica è finalmente utile: per esempio, per molte persone adesso è più rapido rivedere la traduzione di un testo italiano in inglese piuttosto che scrivere in inglese... eccetera
- La traduzione sembra migliorare quanto più sono simili le lingue (spagnolo e italiano ottengono risultati migliori di inglese e italiano, per ora)
- Funziona sorprendentemente bene anche per il parlato, in tempo reale!

Il contesto delle attività di traduzione

- Nella pratica, la traduzione non è un evento in isolamento
- La traduzione si inserisce in una serie di pratiche, spesso con un valore commerciale
- Non tutte le culture hanno dato importanza alla traduzione: spesso si traduce poco (perché non c'è interesse, perché si legge direttamente nell'originale, ecc.)
- Il grosso della traduzione sembra diretto dalle lingue maggiori a molte lingue minori: non è detto che sia sempre così

L'Index translationum Unesco

(dati a volte vecchi di 30 anni)

	"TOP 50" Original language	
1	English	1266110
2	French	226123
3	German	208240
4	Russian	103624
5	Italian	69555
6	Spanish	54588
7	Swedish	39984
8	Japanese	29246
9	Danish	21252
10	Latin	19972
11	Dutch	19667
12	Greek, Ancient (to 1453)	18077
13	Czech	17161
14	Polish	14663
15	Norwegian	14276
16	Chinese	14071
17	Arabic	12410
18	Portuguese	11583
19	Hungarian	11297
20	Hebrew	10279

	"TOP 50" Target Language	
1	German	301935
2	French	240045
3	Spanish	228559
4	English	164509
5	Japanese	130649
6	Dutch	111270
7	Russian	100806
8	Portuguese	78904
9	Polish	76706
10	Swedish	71209
11	Czech	68921
12	Danish	64864
13	Chinese	63123
14	Italian	61087
15	Hungarian	55214
16	Finnish	48311
17	Norwegian	35161
18	Greek, Modern (1453-)	30459
19	Korean	28168
20	Bulgarian	27457

Fig. 1 Daily Translation Usage in Google Translate (May 2016)



Fig. 2 Translation Usage in Facebook (May 2016)



Grande importanza

- In alcuni periodi storici la traduzione è stata un'attività di grande importanza, per una cultura nel suo assieme o per un settore specifico
- La traduzione dal greco al latino è stata importantissima per la civiltà romana, e poi di nuovo nel Rinascimento
- La traduzione dall'olandese (*Rangaku*) è stata l'unico punto di contatto del Giappone con la cultura occidentale dagli inizi del Seicento alla metà del Settecento



Traduzione di tutto?

- È possibile che, per esempio, la traduzione di tutti i testi scritti formali diventi la regola?
- Ovviamente, molti conterrebbero imperfezioni, mentre le traduzioni umane oggi sono abbastanza revisionate...
- ... ma forse possiamo adattarci a questa mediocre qualità (oggi in alcune circostanze accettiamo già testi mediocri, purché utilizzabili)
- Effetti sull'italiano? Probabilmente queste tecniche non **diffonderebbero** la lingua, ma aiuterebbero a **conservarla** contro la pressione dell'inglese o di eventuali altre lingue

Traduzione del parlato

- Oggi possibile in tempo reale
- I risultati sono tutti da valutare
- Sono anche *molto* difficili da valutare, perché l'interazione ha senso tenendo conto delle competenze dell'interlocutore
- "Foreigner talk" facilitato dalla tecnologia?
- Vediamo innanzitutto il contesto

Traduzione e interpretariato

- La traduzione ha avuto spesso un alto prestigio culturale; l'interpretariato è stato considerato spesso un'attività meccanica
- La capacità di comunicare a voce indipendentemente dalle barriere linguistiche è al centro del discorso di Ostler sulle lingue franche
- Nella narrativa e nei film si è spesso immaginato di risolvere i problemi di comunicazione in tempo reale attraverso sistemi intuitivi e istantanei (incluso il "Babel fish")



Traduzione e interpretariato

- Tipi di "traduzione" fatta da interpreti:
 - Simultanea: il discorso non viene interrotto
 - Consecutiva: il discorso viene interrotto per lasciare spazio alla traduzione
- Anche in questo caso piccole modifiche (di "interfaccia") possono fare la differenza
- Uno standard del passato: manuali di conversazione con frasi pronte all'uso

Monty Python - Il frasario ungherese
(in: *Adesso qualcosa di completamente diverso*, 1971)
https://it.wikipedia.org/wiki/Il_frasario_ungherese



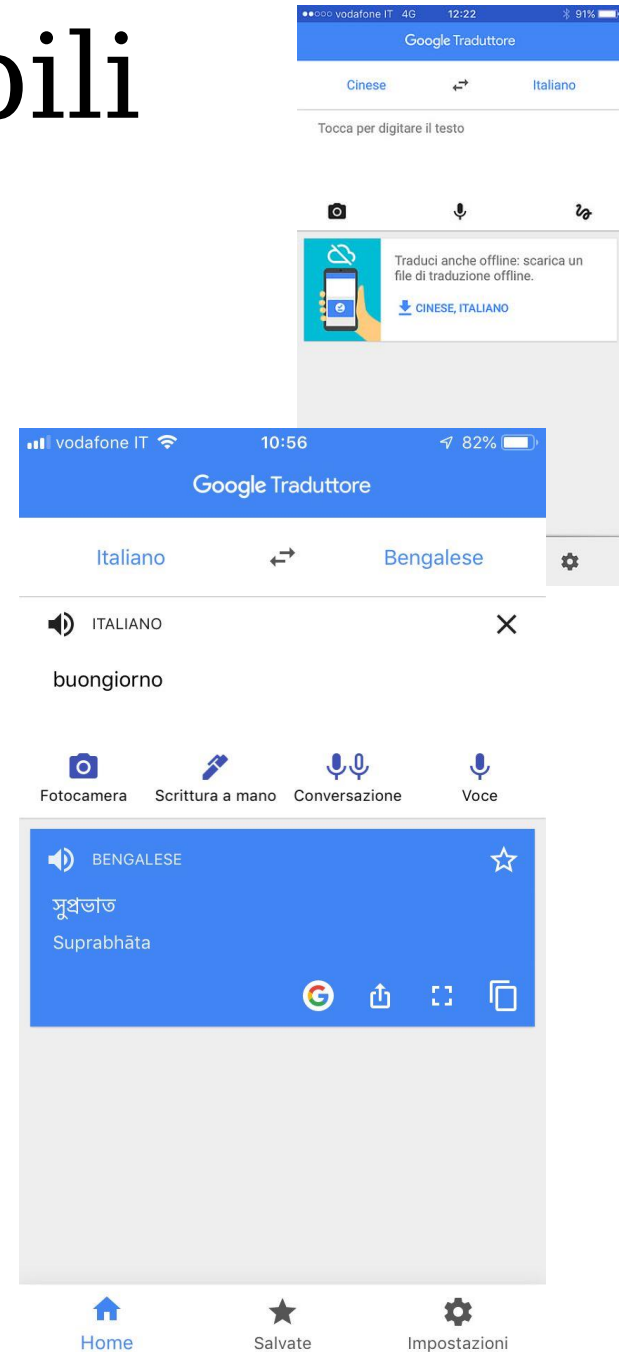
Traduzione automatica del parlato

- La differenza tra interprete e traduttore – e tra scritto e parlato – viene molto ridotta
- A monte, in molti casi un unico sistema si occupa di entrambe le funzioni (come avviene per Google Traduttore e Microsoft Translator)
- Le differenze più importanti diventano:
 - Prestazioni (capacità o meno di riconoscere anche il parlato, con tutti i problemi visti la scorsa settimana)
 - Interfaccia (adatta o meno a un lavoro in tempo reale)



Strumenti disponibili

- L'app Traduttore di Google sembra di gran lunga il miglior prodotto disponibile (permette anche di scaricare pacchetti di lingue per la traduzione offline)
- iTranslate viene spesso citato come seconda migliore alternativa
- Traduttore permette l'input a voce, digitando, disegnando o inquadrando immagini
- L'output è formato da testo o voce
- Una prova:
<https://youtu.be/kKLITKo8HM4>



Testimonianze

- Si trovano pochissime testimonianze di uso reale, soprattutto da parte di italiani – una ricerca potrebbe essere molto utile
- Non mi sembra esistano ancora statistiche
- (Google Traduttore ha avuto a lungo problemi in Cina e potrebbe averli ancora adesso: il concorrente cinese è Baidu, <http://www.baidu.com/>)
- Per il mobile, l'uso più documentato mi sembra la traduzione di scritte; meno la traduzione del parlato in tempo reale
- Viceversa, si trovano numerose testimonianze dell'uso di Google Traduttore per pagine web, e-mail e simili
- Molte testimonianze relative al mobile mostrano disagio e riluttanza a usare il sistema

Esempi

.Another essential for me was GOOGLE TRANSLATE because it allowed me not only to communicate with the voice option , but even to translate the kanji characters via the mobile phone camera, simply by framing the message that I wanted to translate.

Esempi

Before I left, I downloaded the Android app and the Japanese language pack for offline translation. I anticipated using this any time someone wanted to talk to me.

As it turns out, this was a terrible idea. Sticking a phone in someone's face every time I had a question was not intuitive and probably offensive. However, I did use it as a **phrasebook**. I saved phrases like "Thank you", "Where is the bathroom?", and "My hovercraft is full of eels" for quick reference. Eventually I was able to memorize them, but having them in an app was convenient. Of course, the usual disclaimers about accuracy still apply: for anything more complex than the absolute basics, be sure to double-check with either an actual phrase book or a native speaker. *In my experience, though, if you could translate a phrase to Japanese, and then use the reverse translate feature on that translation and get the correct sentence back, it tended to be close enough for native speakers to understand.*



Sarah Hashim-Waris
Los Angeles Times



Problema di interfaccia

- Parlare a una persona presente usando un telefono è ancora ritenuto scortese
- D'altra parte, ci siamo abituati rapidamente all'uso di smartphone (che hanno poco più di dieci anni)
- Qualche innovazione, anche non connessa – per esempio, auricolari senza fili e comodi da usare, in grado di accettare una trasmissione dall'esterno – potrebbe rendere di colpo normale questa situazione
- Un esempio di soluzione intermedia (che non ho provato): ili
<https://iamili.com/>
 - per inglese, cinese e giapponese
 - senza bisogno di connessione Internet
 - soprattutto, unidirezionale: dall'inglese al cinese, al giapponese o allo spagnolo (ogni strumento supporta una sola lingua)

Poco usata: scrittura controllata

- Scrittura controllata = scrivere seguendo alcuni vincoli con l'obiettivo di permettere qualche tipo di elaborazione sul testo
- Molto usata nella tradizione italiana per la *semplificazione* linguistica (Maria Emanuela Piemontese)
- Per la traduzione automatica, significa scrivere eliminando le ambiguità e i punti problematici
- Tuttavia, non permette mai di risolvere tutti i problemi per tutte le lingue

Futuro o passato?



[Download](#) [Business](#) [Tariffe](#) [Aiuto](#)

[Accedi](#) [Iscriviti](#)

Skype Translator

Abbatti le barriere linguistiche con gli amici, la famiglia e i colleghi.

Il nostro traduttore online può aiutarti a effettuare videochiamate in 8 lingue e inviare e ricevere messaggi istantanei in oltre 50 lingue.

Skype Translator sfrutta l'apprendimento automatico. Quindi più lo usi, più migliora. Ti ringraziamo per la tua pazienza durante il passaggio dalla modalità di antepima.

Scarica Skype per
Windows desktop



Microsoft Skype Translator

- Una collocazione impeccabile: l'interfaccia di comunicazione viene inclusa nello strumento normalmente usato per comunicare
- Nella pratica, lo strumento funziona sorprendentemente bene - anche se gli errori nella trascrizione si combinano con gli errori nella traduzione
- In fin dei conti, siamo abituati a comunicare così, con gli stranieri

Microsoft Skype Translator

The screenshot shows a Skype chat interface. At the top left, there is a profile card for 'MLA Unipi' with a star icon and a green 'In linea' status indicator. To the right of the profile are icons for video call, voice call, and a plus sign. Below these is a small globe icon representing the translation feature, which is highlighted by a red arrow. The chat history shows four messages: two 'Stai chiamando' (calling) and two 'Chiamata terminata' (call ended) messages, all dated 'Oggi' (today). At the bottom, there is a text input field with the placeholder 'invia tramite Skype' and a row of icons for attachments, documents, messages, video, voice, and a send button.

☆ MLA Unipi
In linea

Video call, Voice call, Add contact

Stai chiamando MLA Unipi Oggi 15:22

Chiamata terminata, durata 00:15 15:23

Stai chiamando **MLA Unipi** 15:24

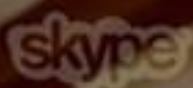
Chiamata terminata, durata 01:05 15:25

invia tramite Skype

Attachments, Documents, Messages, Video, Voice, Send

Funziona sorprendentemente bene! Per esempio, per italiano e cinese (verificato con l'Istituto Confucio di Pisa)

This call may be recorded to improve translation.



Hi grandma. Happy birthday. Are you having a great day?



Hola abuela. Feliz cumpleaños. ¿Tienes un buen día?



Hi Dylan. Yes, a great day thanks. And the flowers that you have sent are precious.

Hola Dylan. Sí, un día estupendo gracias. Y las flores que has enviado son preciosas.



I'm glad they arrived.

Me alegra que llegaron.



They arrived first thing in the morning.

Llegaron a primera hora de la mañana.

Type a message here



Lingue supportate

Arabo

Arabo (levantino)

Cinese (mandarino)

Inglese

Francese

Tedesco

Hindi

Italiano

Giapponese

Portoghese (Brasile)

Russo

Spagnolo

Conclusioni?

- Si dà per scontato che l'interpretariato sia imperfetto
- Più che con limiti invalicabili di qualità e comprensione, abbiamo a che fare con limiti dell'interfaccia
- Molti di questi limiti sembrano elegantemente superabili
- In fin dei conti, viviamo in un mondo in cui oggi molta gente gira già con gli auricolari





Oltre i numeri

- Ma se la traduzione automatica inizia a funzionare abbastanza bene...
- ... quali scenari si aprono?
- Sono ancora tutti da esaminare per quanto riguarda le applicazioni per la didattica
- Possiamo però vedere anche un contesto più ampio, tenendo presente i vincoli realistici
- Un punto importante: la capacità di conversazione in più lingue

Infiltrazione in casa

- Assistenti come Siri e Cortana permettono di fare poche cose
- Amazon Alexa e Google Home sono sistemi di ascolto sempre attivi per l'uso domestico
- Apprezzati in cucina, per la domotica, per la gestione di audio o gli ordini via Internet...
- Soprattutto, sono usati come «altoparlanti intelligenti»
- Un limite è dato dall'incapacità di comprendere davvero i contenuti della comunicazione



Metodi di valutazione

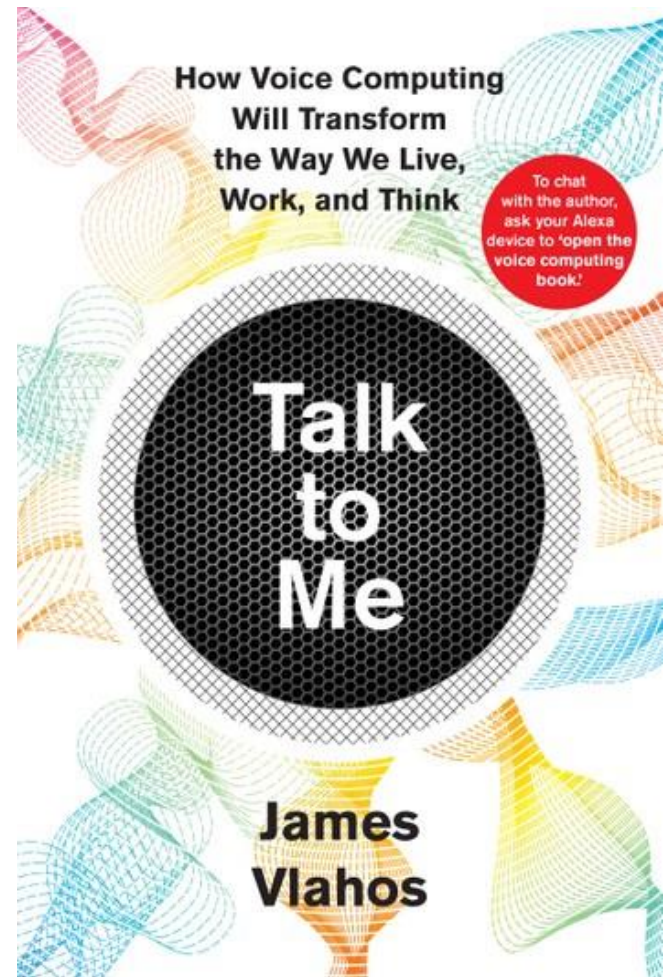
- Un sistema di dettatura funziona bene quando trascrive ciò che l'utente vuole che scriva (WER)
- Un altoparlante intelligente funziona bene quando fa quello che l'utente vuole che faccia: non occorre che trascriva tutte le parole, basta che individui quelle che servono a capire il comando
 - In effetti, il sistema usato per gli altoparlanti intelligenti è sostanzialmente identificazione di parole chiave (*pattern matching*)
 - Il principio base è l'identificazione dell'*intent*: capire qual è la funzione di cui si vuole servire l'utente
 - In sostanza: «Sveglia otto» ottiene gli stessi risultati di «Metti la sveglia alle otto»
- Con gli assistenti vocali compare anche una trascrizione delle richieste degli utenti, quindi volendo si può valutare la trascrizione... ma non è questo il centro dell'interesse

Voice Recognition Lift



Infiltrazione ovunque

- Il giornalista statunitense James Vlahos ha descritto in *Talk to Me* (2019) una serie di applicazioni dei sistemi conversazionali
- Alcune di queste vanno molto in profondità, inclusa la simulazione di conversazioni con il padre defunto
- Un inquadramento:
http://www.treccani.it/magazine/lingua_italiana/articoli/scritto_e_parlato/IA.html



Assistenti come docenti

- Un punto importante è l'uso di assistenti vocali per *insegnare* lingue straniere
- Alcuni corsi di lingua commerciali fanno già uso di sistemi di riconoscimento del parlato - ma l'uso è ancora piuttosto limitato (Rosetta Stone)
- Questa però è una sfida interessante per la didattica delle lingue: che cosa ci si può costruire attorno?
- E può essere un'estensione dell'uso dell'italiano?



Un'analisi recente

Catia Cucchiarini e Helmer Strik, *Second language learners' spoken discourse: practice and corrective feedback through automatic speech recognition*, cap. 8 di *Innovative methods and technologies for electronic discourse analysis*, a cura di Hwee Ling Lim e Fay Sudweeks, ISI, pp. 169-189.

- Nell'insegnamento della lingua ad adulti (L2) si riconosce l'importanza dei **giudizi** e delle **correzioni** – in aggiunta all'esposizione alla lingua target e ai tentativi basati sull'uso
- Tuttavia, giudizi e correzioni senza vincoli possono essere prodotti solo da esseri umani
- Si possono creare giudizi e correzioni «automatici» (= soluzione di esercizi scritti in un libro) solo in contesti vincolati
- Inoltre, giudizi e correzioni sono molto onerosi per gli insegnanti umani, quindi vengono forniti solo raramente
- Altro punto: in aula, per molti apprendenti è sgradevole esporsi a correzioni e commenti
- Conclusione: questa serie di fattori rende molto interessante lo sviluppo di sistemi automatici di correzione

Controllo

- Un aspetto interessante (pp. 177-178): al momento ci sono poche certezze sul modo migliore per rivolgersi agli apprendenti
- Per esempio, è meglio fornire spiegazioni articolate o limitarsi a ripetere le parole sbagliate, senza fornire commenti?
- La scarsità di informazioni “is mainly due to the impossibility so far to create appropriate research conditions to offer feedback that is systematic, consistent, intensive, and clear enough to be perceived as such, and that provides opportunity for self-repair and modified output” (p. 178)
- L'uso di sistemi informatici aiuterebbe a verificare meglio le prestazioni delle diverse tecniche
- Su un altro piano: la scarsità di esempi documentati di correzione è ancora oggi un problema (p. 181)

La pronuncia

- Una delle aree di maggior interesse per i sistemi informatici è la pronuncia
- In un libro su carta non si può avere nessuna registrazione; in un corso su dischi, sì - ma manca naturalmente il fattore di correzione (le uniche attività che si possono avere sono il «ripeti»)
- Tra anni Novanta e primi Duemila si è visto che i sistemi sono in grado di fornire valutazioni «con un'alta correlazione» con le valutazioni umane
- Sono stati creati diversi sistemi di questo tipo, tra cui per esempio quelli di Rosetta Stone o Tell me More per l'inglese
- I risultati comunque non sono sempre soddisfacenti
- Per la conversazione, per esempio, Rosetta Stone si basa su una rete di tutori madrelingua

SPEECH RECOGNITION TALKING BACK REQUIRED.

One of the most important parts of learning a new language is speaking confidently. And whenever you do, it's important to know you're getting it right. That's why we've invested in speech recognition technology that listens closely to every syllable—and gives you immediate feedback on your pronunciation.

Our speech recognition technology is highly advanced. Robust enough to support the more than 30 languages we offer. And it's proprietary—which means we're the only language-learning method that offers it.

Want to know more? [Lets get granular.](#)

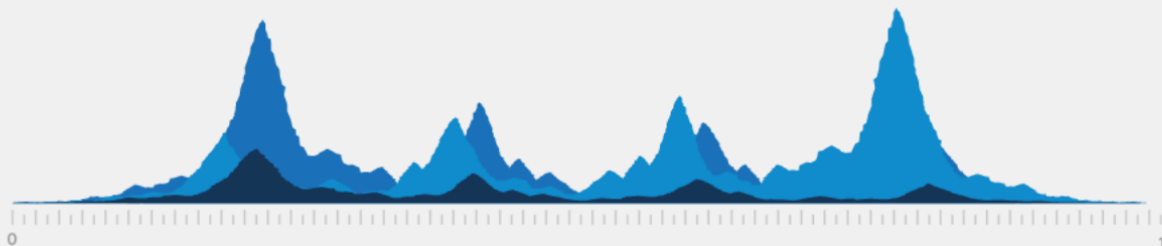
[TRY](#)[METHOD](#)[WHAT YOU'LL LEARN](#)[REVIEWS](#)[PRICING](#)

GETTING YOU TO SPEAK CONFIDENTLY IN YOUR NEW LANGUAGE IS OUR GOAL.

That's why you'll start speaking in the very first lesson. Instantaneously, our state-of-the-art speech recognition technology is ready to evaluate exactly how you sound.

Our speech recognition technology identifies the words you're pronouncing; while you're speaking them. (And provides real-time assessment.)

First, it verifies whether or not the words or phrases you've used are correct. Then, it compares your pronunciation to a native speaker model to see how you stack up.



Valutazioni di pronuncia

Per quantificare gli errori di pronuncia sono state proposte diverse metriche parziali (non c'è un indicatore di riferimento paragonabile a WER per la trascrizione o a BLEU per la traduzione automatica):

- Algoritmo GOP - «Goodness of pronunciation»
- Algoritmo wGOP - con pesi diversi per diversi suoni
- Controllo di tratti fonetici specifici con un classificatore
- Classificazione di errori noti
- Combinazione (con pesi variabili) dei sistemi precedenti

Valutazione del risultato

- Sono state fatte verifiche con gruppo di controllo
- Un gruppo di apprendenti riceve istruzioni linguistiche con feedback
- Un gruppo di controllo riceve le stesse istruzioni, ma senza feedback
- Alla fine, si misura la pronuncia dei due
- Per l'olandese, è risultato che la pronuncia di chi aveva ricevuto il feedback era effettivamente migliore

Estensione ad altri ambiti

- Dalla fonetica si può passare alla morfologia e alla sintassi
- Su questo è stato fatto lavoro anche per i testi scritti, ma con pochi risultati
- Alcune tecniche di aiuto possono essere anche semplici: per esempio, si chiede di combinare pezzi di frase precomposti e il riconoscimento vocale serve solo a determinare se è stata pronunciata la frase corretta
- Gli esempi forniti mi sembrano molto limitati



Valutazioni di Cucchiarini e Strik

- I sistemi disponibili non sono ancora adeguati a coprire tutti gli aspetti di una lingua
- Vincolare le risposte a parole determinate è in sostanza la soluzione più efficace (p. 179)
- Il resto è molto più incerto
- Del resto, non esiste ancora la conversazione libera!

Alexa Prize

- Un articolo di James Vlahos su «Wired» (27 febbraio 2018) descrive le procedure di giudizio per l'Alexa Prize 2017
<https://www.wired.com/story/inside-amazon-alexa-prize/>
- I bot che si sono classificati meglio sono basati soprattutto su regole; il vincitore è in parte regole in parte intelligenza artificiale
- Attualmente è in corso l'edizione 2019, partita con 10 partecipanti sponsorizzati
<https://developer.amazon.com/alexaprize>
- Obiettivo: «[to] create socialbots that can converse coherently and engagingly with humans on a range of current events and popular topics such as entertainment, sports, politics, technology, and fashion.»



The Alexa Prize

Over \$3.5 Million to Advance Conversational Artificial Intelligence

December 2017 - November 2018



Announcing the 2018 Alexa Prize Finalists

Amazon is excited to announce the finalists competing in the 2018 Alexa Prize. Congratulations to Alquist from the Czech Technical University in Prague, Alana from Heriot-Watt University, and Gunrock from University of California Davis!

Nel contesto

- I sistemi di interazione vocale saltano spesso la mediazione di testi umani preesistenti: a volte leggono una risposta «da Wikipedia», a volte forniscono una risposta partendo da dati grezzi (e potenzialmente possono rispondere in più lingue)
- Per alcuni settori, si intravede un futuro non remoto in cui, una volta inseriti dati da qualche parte, un assistente digitale può elaborarli e presentarli nella lingua dell'utente
- Anche qui: **conservazione** dell'italiano finché ci sarà gente che preferirà ottenere i risultati della ricerca in italiano
- Inoltre, questi sistemi permetteranno forse di fare un tipo soddisfacente di **conversazione** anche se non sono disponibili madrelingua: per alcuni entusiasti, questo potrebbe essere un canale di diffusione della lingua (motivazione «culturale»)
- Portarsi dietro una bolla di italiano?

Ian McDonald, *River of Gods* (2009)

Mr Nandha the Krishna Cop travels this morning by train, in a first class car. Mr Nandha is the only passenger in the first class car. The train is a Bharat Rail electric shatabdi express: it piles down the specially-constructed high-speed line at three hundred and fifty kilometres per hour, leaning into the gentle curves. Villages roads fields towns temples blur past in the dawn haze that clings knee-deep to the plain. Mr Nandha sees none of these. Behind his tinted window his attention is given over to the virtual pages of the *Bharat Times*. Articles and video reports float above the table as the lighthoek beats data into his visual lobes. In his auditory centre: Monteverdi, the Vespers of the Blessed Virgin performed by the Camerata of Venezia and the Choir of St Mark's.

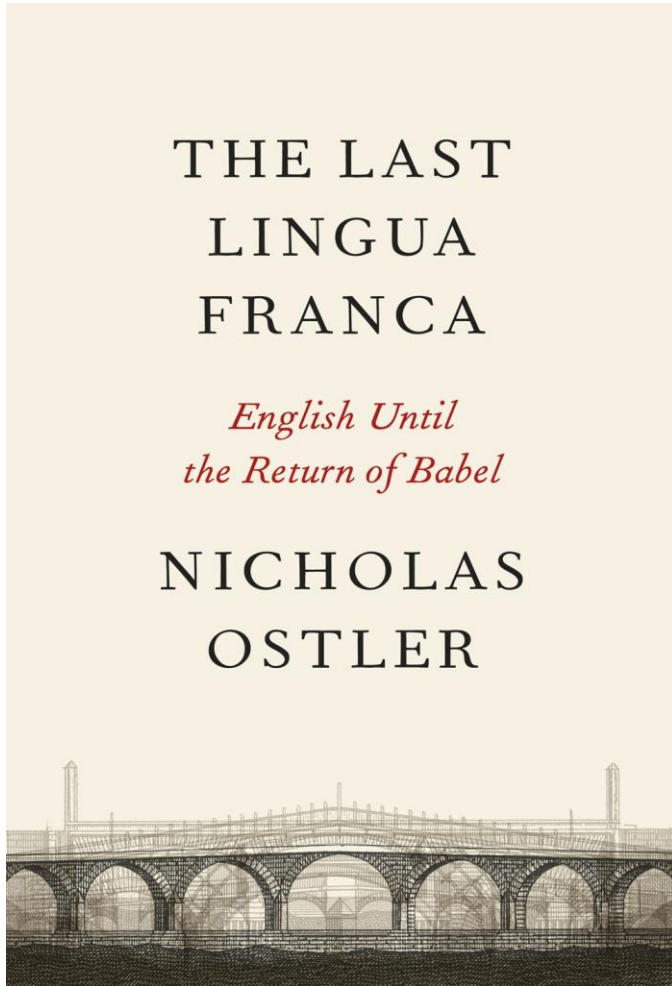
Mr Nandha loves very much the music of the Italian renaissance. Mr Nandha is deeply fascinated with all music of the European humanist tradition. Mr Nandha considers himself a Renaissance man. So he may read news of the water and the maybe-war and the demonstrations over the Hanuman statue and the proposed metro station at Sarkhand Roundabout and the scandal and the gossip and the sports reviews, but part of his visual cortex the lighthoek can never touch envisions the piazzas and campaniles of seventeenth century **Cremona**.

Mr Nandha has never been to **Cremona**. He has never visited Italy. His imaginings are Planet History Channel establishing shots cut with his own memories of Varanasi, the city of his birth, and Cambridge, the city of his intellectual rebirth.

The train slams past a rural brickworks; kiln smoke lying on top of the mist. The ranks of stacked bricks are like the ruins of an unborn civilization. Kids stand and stare, hands raised in greeting, dazed by the speed. After the train has passed, they scramble up on to the track and look for paisa coins they have wedged into the rail joints. The fast trains smear them flat into the rail. There's stuff you could buy with those coins but none would be as good as seeing them become stains on the high-speed express line.

The chai-wallah sways down the carriage.

Nicholas Ostler, *The last lingua franca*



- Allen Lane, Londra, 2010
- Libro divulgativo... utile perché fa il punto sulla situazione attuale e sull'evoluzione futura in modo intelligente e con diversi confronti storici
- In sintesi, la tesi: l'inglese sarà l'*ultima* lingua franca per la comunicazione internazionale, e la traduzione automatica avrà un ruolo importante negli sviluppi

Sviluppi improbabili?

- Molti degli sviluppi accennati sono decisamente improbabili
- Tuttavia, la pervasività dei sistemi automatici rende senz'altro più facile passare da una lingua all'altra per tutta una serie di scopi
- L'idea di base: il mondo basato su un monolinguisimo inglese, o su una divisione di ruoli tra l'inglese (come lingua franca) e le altre lingue (tra cui l'italiano) **non** è inevitabile - è una possibilità tra le altre, anche se più verosimile di altre
- Un cambiamento politico spettacolare può far cambiare la situazione in tempi brevi

Certezze

- Gli esseri umani hanno bisogno di una lingua per esprimersi
- La capacità di imparare una lingua dipende da molti fattori, ma di solito è richiesto uno sforzo che pochi fanno da adulti
- La socializzazione contemporanea fa sì che le lingue della società ospite si impongano sulle lingue immigrate
- Le lingue nazionali europee (tra cui l'italiano) sono molto resistenti nel contesto sociale moderno