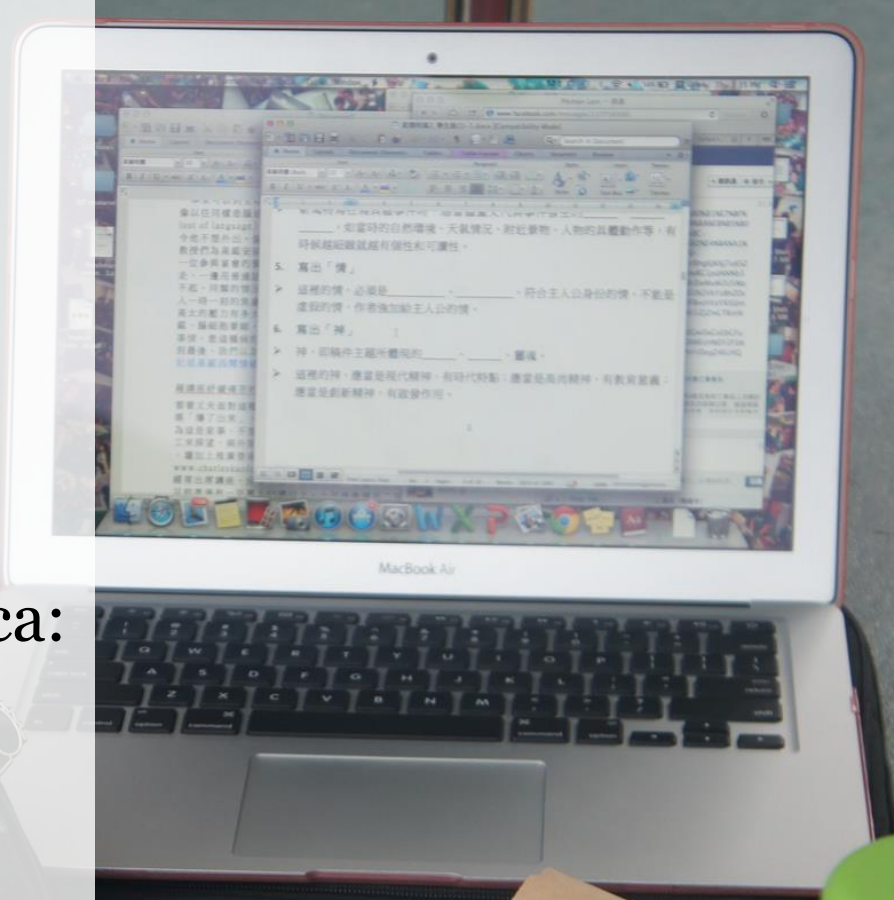


# Modulo B

## Valutare i sistemi automatici di interazione linguistica

### 2. La traduzione automatica: contesto e nozioni base

1 marzo 2017



Linguistica italiana II  
Mirko Tavosanis  
A. a. 2016-2017

# La traduzione automatica

Di che cosa stiamo parlando?

- Definizione (rapidissima)
- Storia
- Strumenti di uso comune

Purtroppo per la traduzione automatica non esiste una presentazione storica in volume equivalente a *The Voice in the Machine* di Pieraccini per il riconoscimento del parlato

# Non traduzione assistita

- Un importante settore di lavoro, oggi, è quello della **traduzione assistita** (*Computer Assisted Translation*, CAT)
  - Particolarmente importante per la traduzione specialistica (testi tecnici, professionali e simili)
  - Al centro della CAT: un traduttore umano
  - Il traduttore umano usa un sistema informatico per esempio per:
    - Allineare testi composti da sezioni ripetitive e assicurare uniformità nella traduzione
    - Tenere traccia del modo in cui un termine viene tradotto nel documento
    - A volte, integrare il prodotto di un sistema di traduzione automatica
  - Un programma di riferimento per fare questo lavoro: SDL Trados Studio, o semplicemente Trados
- **Non** ci occuperemo di questo, ma solo di traduzione fatta in modo completamente automatico

# Sistemi di traduzione automatica

- Funzionano? E se sì, come?
- In fin dei conti, anche un dizionario su carta è un sistema di «traduzione automatica»: contiene traduzioni di parole, traduzioni di frasi fatte con le parole

## traduzione

nf

**1** translation, rendering, version

◇ **traduzione a prima vista** unseen; **traduzione a senso** free translation; **una traduzione dal cinese all'inglese** a translation from Chinese into English; **traduzione fedele** close/faithful translation; **traduzione sbagliata** mistranslation; **teoria della traduzione** translation theory; **in traduzione** in translation; **fare esercizi di traduzione** to \*do translation; **fare una traduzione (di/in/dal)** to \*make/to \*do a translation (of/into/from); **leggere Milton in traduzione** to \*read Milton in translation; **sbagliare la traduzione di** to mistranslate

**2** (= *trasporto*) transfer, translation

◇ **traduzione di detenuti** transfer of prisoners

**3** FIN translation

◇ **traduzione valutaria** currency translation.

- Un essere umano con un dizionario può lavorare in modo molto algoritmico... ma con i computer siamo comunque in un altro mondo

# Traduzione automatica: definizione

- Il concetto è intuitivo: «Traduzione eseguita con speciali programmi per calcolatori elettronici che, opportunamente regolata, garantisce la corrispondenza tra testi di lingue diverse» (De Mauro)
- In inglese si parla di *Machine translation* (MT)
- Le tecnologie alla base sono molto differenziate, ma il principio è lo stesso: input in una lingua, output considerato «corrispondente» in un'altra lingua
- Vedremo la prossima settimana che il problema è più complicato di quel che sembra

# Traduzione automatica: storia

- L'idea di usare i computer per fare traduzione automatica nasce già negli anni Quaranta (ovviamente, solo per i testi scritti)
- Un «esperimento» condotto dalla Georgetown University e dalla IBM il 7 gennaio 1954 (traduzione automatica di frasi russe in inglese) diede l'idea che la traduzione automatica fosse a portata di mano – tempo cinque anni
- In realtà, naturalmente, la traduzione automatica non funzionava davvero: nella dimostrazione Georgetown-IBM gli esempi erano stati scelti in rapporto a quel che si poteva fare con il programma
- Le ragioni per il fraintendimento sono le solite: la complessità del linguaggio naturale (e delle altre cose che non si imparano a scuola) veniva molto sottovalutata
- In questo caso comunque i linguisti avevano avvisato immediatamente dei problemi: Yehoshua Bar-Hillel aveva notato per esempio che senza una conoscenza enciclopedica universale non è possibile disambiguare il significato di molte parole

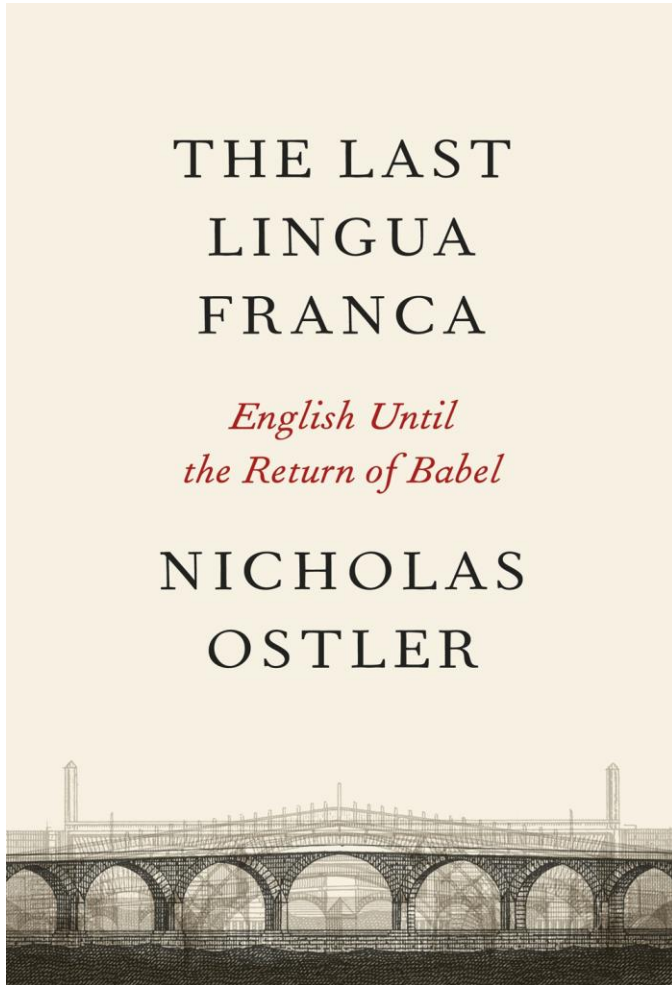
# Il momento di svolta

- Negli anni Novanta si diffonde la traduzione statistica, alimentata da grandi riserve di testi confrontabili sul web
- Per Google, inizialmente, le dimensioni di riferimento sono:
  - Corpora di 1 miliardo di parole per ogni lingua
  - Corpora paralleli di 200 milioni di parole per ogni coppia di lingue
- La traduzione «linguistica» (basata su regole grammaticali) viene abbandonata
- La traduzione automatica inizia ad avere applicazioni pratiche
- Potenzialmente può avere effetti rivoluzionari sulla comunicazione, in particolare nell'area delle lingue franche (LF)

# La traduzione automatica e le lingue franche



# 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
- I dati che vedremo adesso sono ripresi da questo libro, con rielaborazioni

# «Lingua franca» originale

- Il termine si riferiva, all'origine, a una lingua parlata per scambi commerciali nel Mediterraneo a partire dal Cinquecento e forse abbastanza stabile
- Per quel che se ne può ricostruire, era in sostanza una forma semplificata di italiano (con l'uso dell'infinito come forma base del verbo) simile a un pidgin
- Esempi: «Acosi, acosi, mirar como mi estar barbero bono, y saber curar, si estar malato, y ora correr bono»
- Come mai il nome? In buona parte del Medio Oriente (e più in là) già nel Medioevo tutti gli europei erano e sono chiamati «Franchi» (*firangi* in arabo, e poi in persiano)
- Per informazioni: Laura Minervini, voce *Italiano come lingua franca* nell'*Enciclopedia dell'italiano* (2010)  
([http://www.treccani.it/enciclopedia/lingua-franca-italiano-come\\_%28Enciclopedia\\_dell%27Italiano%29/](http://www.treccani.it/enciclopedia/lingua-franca-italiano-come_%28Enciclopedia_dell%27Italiano%29/))
- L'uso dell'infinito è diventato poi anche uno stereotipo letterario, usato ancora oggi per la parlata di stranieri

# «Lingua franca» secondo Ostler

- La definizione di Ostler è molto semplificata, ma coglie i tratti comuni a molte situazioni L2/LS
- In sostanza, le lingue franche sono: lingue straniere che vengono imparate e usate da persone che non le hanno come L1 perché **servono a qualcosa**
- «International English is a lingua franca, and by its nature, a lingua franca is a language of convenience. When it ceases to be convenient – however widespread it has been – it will be dropped, without ceremony, and with little emotion» (Ostler, p. XV)

# Ostler: utilità della storia

- «The only trustworthy guide we have to the long-term outcomes for real languages in real societies is the historical record» (p. 6)
- Anche per le evoluzioni future questa è la nostra unica base... accoppiata a qualche informazione sul modo in cui funzionano gli esseri umani
- Non ci siamo mai trovati in una situazione simile, ma possiamo fare qualche speculazione intelligente basata su ciò che sappiamo
- Vedere i precedenti storici ci può dare un'idea, con gli opportuni adattamenti, sul modo in cui saranno gestite le lingue franche in futuro

# Come si diffonde una lingua franca?

- Il processo può essere molto complicato e determinato da molti fattori storici
- Due note importanti:
  - **Non** è un processo determinato da **leggi meccaniche**
  - Le **tendenze** di oggi **non** dicono molto sulla situazione di domani
- Qualche previsione è però possibile sul medio periodo: le lingue «girano» solo sugli esseri umani, che le imparano nei primi vent'anni di vita e poi le usano per altri sessanta...
- ... quindi la demografia dice molto sull'uso futuro delle lingue!

# I numeri delle lingue (dipendono dai criteri...)

- Al momento si stima che siano in grado di parlare **inglese** 1143 milioni di persone: 331 milioni come madrelingua, 812 come seconda lingua  
(per esempio, nell'UE i madrelingua sono per ora il 13%, ma il 38% della popolazione conosce l'inglese come seconda lingua)
- Sono in grado di parlare **cinese mandarino** 1051 milioni di persone: 873 come madrelingua, 178 come seconda lingua... ma i parlanti L2 vivono in realtà quasi tutti in Cina (minoranze, cantonese...)
- L'inglese in sostanza è di gran lunga la lingua più parlata da *non madrelingua* a livello globale
- Molte lingue sono conosciute anche da molti non madrelingua, ma in aree geograficamente definite (francese, arabo, swahili, farsi...)

# Il cinese può rimpiazzare l'inglese, come lingua di comunicazione internazionale?

- *Molto* improbabile – nel prevedibile futuro, e nonostante la crescita economica della Cina
- Per quanto oggi sia di moda il cinese, i bambini delle scuole italiane (e di quelle di tutto il mondo) imparano l'inglese, e non si vede come la situazione possa cambiare in tempi brevi
- Oggi del resto i cinesi studiano inglese, ma gli americani non studiano cinese...
- Inoltre, il cinese è *molto* difficile per la maggior parte dei non-cinesi, e per chi parla in generale una lingua indoeuropea
- La difficoltà dell'apprendimento è relativa (per gli americani il Dipartimento di Stato calcola: 6 mesi per lo spagnolo, 24 per il cinese), ma al di fuori della Cina in pratica non esistono lingue simili al cinese

# La tesi di Ostler

- In futuro sarà meno importante avere una lingua di comunicazione internazionale, perché le macchine migliorano
- Tradurre tutto correttamente richiede forse una vera intelligenza artificiale (lontanissima...), ma in molte situazioni è sufficiente una traduzione “abbastanza buona”
- Già oggi i traduttori automatici tipo Google Translate permettono di farsi un’idea di testi anche in lingue sconosciute, e a volte questo basta
- Inoltre, questi strumenti **non fanno che migliorare**
- Prima o poi raggiungeremo il punto in cui **per alcune culture** (la Cina innanzitutto) non sarà più conveniente studiare una lingua straniera per fini pratici: forse entro pochi decenni?





Google™  
Translate



Benvenuto  
nell'anteprima di  
Skype Translator

Mai più barriere linguistiche: una  
conversazione alla volta.

Registrati subito



# Funzionamento e prodotti

# Scritto o parlato?

- Potenzialmente, un sistema di traduzione del parlato potrebbe funzionare così:

Riconoscimento vocale nella L1 > Traduzione > Sintesi vocale nella L2

- Nella pratica, i normali sistemi di riconoscimento del parlato spesso non forniscono informazioni importanti per la traduzione (Pieraccini 2012, pp. 274-275): per esempio, la punteggiatura
- Occorre quindi avere un sistema dedicato, che per esempio faccia ipotesi sul punto in cui iniziano e finiscono le frasi e comunichi questi dati al modulo di traduzione
- Oggi alcuni sistemi possono lavorare sia sullo scritto sia sul parlato... alcune funzionalità dipendono dall'interfaccia, più che da vincoli base

# Google Traduttore

- Nel 2003 viene lanciato Google Translate (in seguito Google Traduttore, in italiano), coprendo la traduzione di 4 lingue con l'inglese – ma non reciprocamente
- Le lingue coperte sono aumentate rapidamente e adesso sono 103, con qualità variabile
- Nel 2013 (ultimo dato disponibile) veniva usato da 200 milioni di persone al mese
- Il livello di qualità nelle lingue a me note è veramente molto alto... ma quanto alto?

# Tecnologia

- Fino al 2006 il servizio usava solo il software SYSTRAN
- Nel 2006 Google inizia a usare la Statistical Machine Translation partendo da inglese e arabo ed estendendo poi il sistema ad altre lingue
- L'addestramento iniziale ha usato il corpus dei documenti tradotti nelle sei lingue ufficiali delle Nazioni Unite:
  - Arabo
  - Cinese
  - Inglese
  - Francese
  - Russo
  - Spagnolo
- Dall'ottobre 2007 SYSTRAN è stato abbandonato per tutte le lingue ed è stato usato solo un sistema Google di Statistical Machine Translation (per alcune lingue le cose sono cambiate pochi mesi fa, come vedremo subito)
- Le traduzioni in lingue diverse dall'inglese vengono fatte usando l'inglese come lingua ponte

# Google e le reti neurali

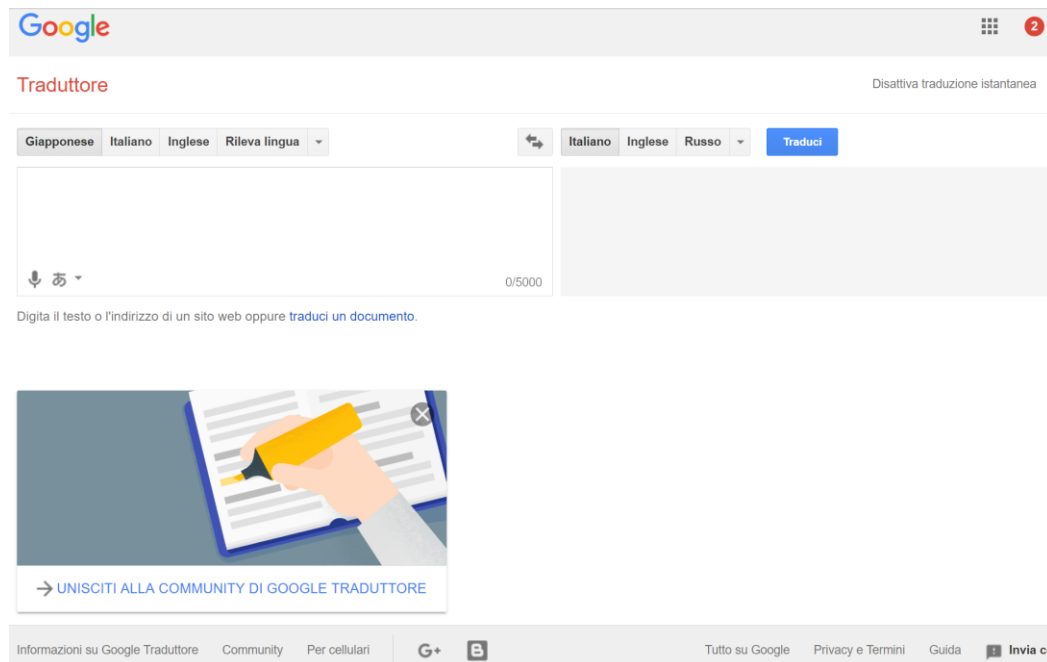
- Anche nella traduzione sono arrivati i sistemi di apprendimento automatico / reti neurali che abbiamo già visto per l'apprendimento del parlato
- Il 15 novembre 2016 Google ha annunciato il passaggio di una parte dei servizi di Google Traduttore a un sistema basato sull'apprendimento automatico: Google Neural Machine Translation (GNMT)
- Il sistema GNMT traduce “whole sentences at a time, rather than just piece by piece. It uses this broader context to help it figure out the most relevant translation, which it then rearranges and adjusts to be more like a human speaking with proper grammar”
- Il sistema GNMT finora è disponibile per traduzioni tra otto lingue: inglese, cinese, francese, tedesco, giapponese, coreano, portoghese, spagnolo e turco... per l'italiano non dovrebbe mancare molto
- Il Sistema GNMT non usa una lingua ponte ma traduce direttamente da una lingua all'altra

Fonte: Barak Turovsky, *Found in translation: More accurate, fluent sentences in Google Translate*, Google Blog, 15 novembre 2016

<https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/>

# Interfacce

- Google Traduttore è disponibile attraverso molti canali:
  - Sito web
  - App (importante per le funzioni di traduzione audio e video in tempo reale, simili a quelle di Microsoft Skype Translator)
  - Estensione del browser...
- Noi ci baseremo sul sito web, usando l'indirizzo <https://translate.google.com/>



# Un assaggio di futuro?

- “Mirko, thank you. I certainly do share your feeling that it was one of the benefits of the existence of the USSR that it forced capitalism, in order to compete, onto a more humane path than it would otherwise have followed; and that we now see various aspects of that softening or humanising being peeled away again, without the Soviet Union to make them seem necessary. I also tend to think that history has not yet necessarily finished with the \*idea\* of the planned economy, even if the Soviet version of it was impossible, clumsy and tyrannical. By the way, there's an Italian translation coming, from Bollati Boringhieri.”
- “Just a question from the linguistic point of view: do you read Italian? Or have you used a translation tool to read the post?”
- “I depend on Google Translate, I'm afraid. I can follow some Italian by analogy with French and school Latin, but not enough...”

<http://linguaggiodelweb.blogspot.it/search?q=spufford>



# Un assaggio di futuro?

- “I can't read Italian, but Google translate tells me this is a very nice review of *Complication*. Thanks, @RedazioneLH!”
- “Also, I read (with the help of Google translate) Jessica's piece on Italian stereotypes in film.”
- “I can't read Italian, but from what I got from Google translate, they are the cheapest option.”

# Ancora più futuro



[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 anteprima.

Scarica Skype per  
Windows desktop

