



SCUOLA SUPERIORE PER MEDIATORI LINGUISTICI

(Decreto Ministero dell'Università 31/07/2003)

Via P. S. Mancini, 2 – 00196 - Roma

TESI DI DIPLOMA DI MEDIATORE LINGUISTICO

(Curriculum Interprete e Traduttore)

Equipollente ai Diplomi di Laurea rilasciati dalle Università al termine dei Corsi afferenti alla classe delle

LAUREE UNIVERSITARIE IN SCIENZE DELLA MEDIAZIONE LINGUISTICA

La relazione tra l'intelligenza artificiale unita al linguaggio naturale

RELATORI:

Prof.ssa Adriana Bisirri

CORRELATORI:

Prof.ssa Maria Nocito

Prof.ssa Luciana Banegas

Prof.ssa Claudia Piemonte

CANDIDATO

GIANFELICE ADAMO

2926

ANNO ACCADEMICO 2020/2021

Alla mia famiglia,
che mi ha sostenuto
in questo percorso.

Indice

INTRODUZIONE	10
1. FONDAMENTI SULL'INTELLIGENZA ARTIFICIALE, TENTATIVI DI CREARE MACCHINE CAPACI DI COMPRENDERE I LINGUAGGI NATURALI	12
2. MODALITÀ USATE DAI COMPUTER PER ANALIZZARE I LINGUAGGI UMANI; LA TECNOLOGIA ALL'AVANGUARDIA DEL GPT-3	22
2.1 SPOKEN DIALOGUE SYSTEM	25
2.2 TRADUZIONE AUTOMATICA BASATA SU REGOLE (RULE-BASED MACHINE TRANSLATION)	29
2.3 TRADUZIONE AUTOMATICA STATISTICA (<i>STATISTICAL MACHINE TRANSLATION</i>).....	31
3. CENNI STORICI E PANORAMICA DELL'ALGORITMO DELL'ELABORAZIONE DEL LINGUAGGIO NATURALE (<i>NATURAL LANGUAGE PROCESSING</i>)	33
3.1 APPRENDIMENTO PROFONDO (DEEP LEARNING).....	38
3.1.1 COME FUNZIONA L'APPRENDIMENTO PROFONDO (<i>DEEP LEARNING</i>)	41
3.2 NEURAL MACHINE TRANSLATION	45
4. ESEMPI SU COME LA ELABORAZIONE DEL LINGUAGGIO NATURALE FUNZIONA NELL'AMBITO DELLA COMUNICAZIONE VOCALE TRA UOMO E MACCHINA	50
4.1 SISTEMA AUTOMATICO DI RICONOSCIMENTO DEL DISCORSO (<i>AUTOMATIC SPEECH RECOGNITION, ASR</i>)	53
4.2 PROBLEMATICHE E FUNZIONAMENTO DEL SISTEMA AUTOMATICO DI RICONOSCIMENTO DEL DISCORSO	55

4.3 ASSISTENTE DIGITALE: UNIONE TRA L'INTELLIGENZA ARTIFICIALE E LA LINGUISTICA.....	61
5. I VANTAGGI E LE PROBLEMATICHE DELL'ELABORAZIONE DEL LINGUAGGIO NATURALE.....	67
5.1 COMPARARE L'APPRENDIMENTO AUTOMATICO E L'APPRENDIMENTO PROFONDO	71
6. IL FUTURO DELL'INTERAZIONE TRA L'INTELLIGENZA ARTIFICIALE ED I LINGUAGGI NATURALI	79
6.1 PERPLESSITÀ E PREOCCUPAZIONI CIRCA UN FUTURO BASATO SULL'INTELLIGENZA ARTIFICIALE.....	84
7. CONCLUSIONI.....	87

Table of contents

English section	93
Introduction	94
1. Fundamentals of artificial intelligence, attempts to create machines capable of understanding natural languages	95
2. Modalities used by computers to analyse human languages; cutting-edge GPT	100
2.1 Spoken Dialogue System	101
2.2 Rules-based Machine Translation.....	102
2.3 Statistical Machine Translation (SMT).....	104
3. Historical background and overview of Natural Language Processing (NLP)	105
3.1 Deep Learning.....	107
3.1.1 How Deep Learning works	108
3.2 Neural Machine Translation	110
4. Examples of how NLP works in the area of voice communication between humans and machines	112
4.1 Automatic Speech Recognition (ASR)	112
4.2 Problems and functions of automatic speech recognition (ASR)	113
4.3 Digital assistant: combining artificial intelligence and linguistics	115
5. Advantages and disadvantages of Natural Language Processing	117
5.1 Machine learning and deep learning in comparison	118

6. The future of interaction between artificial intelligence and natural languages.....	121
6.1 Perplexities and concerns about an artificial intelligence-based future	123
Conclusion.....	125

Índice

Sección española.....	127
Introducción	128
1. Fundamentos de la inteligencia artificial y los intentos de crear máquinas capaces de entender las lenguas naturales	130
2. Modos utilizados por los ordenadores para analizar las lenguas humanas y la tecnología de vanguardia del GPT-3	135
2.1 Spoken Dialogue System	136
2.2 Rules-based Machine Translation.....	137
2.3 Traducción automática estadística.....	139
3. Breve historia y descripción general del algoritmo de Procesamiento de Lenguaje Natural (PLN)	140
3.1 Aprendizaje profundo	142
3.1.1 Cómo funciona el aprendizaje profundo.....	143
3.2 Traducción automática neuronal.....	145
4. Ejemplos de cómo funciona el procesamiento del lenguaje natural en la comunicación oral entre humanos y máquinas	147
4.1 Sistema de reconocimiento automático del habla (<i>Automatic Speech Recognition, ASR</i>).....	148
4.2 Problemas y funcionamiento del sistema de reconocimiento automático del habla	149

4.3 Asistente virtual: combinación de inteligencia artificial y lingüística	151
5. Ventajas y problemas del procesamiento del lenguaje natural	153
5.1 Comparación entre el aprendizaje automático y el aprendizaje profundo	154
6. El futuro de la interacción entre la inteligencia artificial y las lenguas naturales	157
6.1 Perplejidades y preocupaciones sobre un futuro basado en la inteligencia artificial	159
7. Conclusiones	161
RINGRAZIAMENTI	163
BIBLIOGRAFIA	164
SITOGRAFIA	165

Introduzione

L'intelligenza artificiale è una branca dell'informatica che studia la progettazione di sistemi software capaci di emulare processi intellettivi umani come ragionare ed imparare dalle esperienze passate.

La società è in costante mutamento, di conseguenza la vita delle persone cambia, le interazioni sociali e la percezione del mondo assumono nuove dinamiche. Uno degli elementi fondamentali che innesca una continua trasformazione nelle nostre vite è qualcosa che sviluppiamo, che apparentemente non si vede, il cui corpo assume svariate forme, a volte si percepisce ed è indispensabile per eseguire anche le azioni più semplici. Questo qualcosa è l'intelligenza artificiale.

L'intelligenza artificiale fornisce ad una macchina la capacità di eseguire dei compiti e azioni che solitamente la mente umana può realizzare, come tradurre un testo, fornire assistenza ai clienti, cercare informazioni di ogni genere per uso lavorativo e privato.



L'immagine mostra il risultato tra l'unione della mente umana e le macchine.

Fonte: [Beginning Your Journey to Implementing Artificial Intelligence | SoftwareONE Blog](#)

Il tema della tesi vuole evidenziare la forte relazione presente tra l'intelligenza artificiale e le lingue naturali – il sistema di suoni e

strutture grammaticali che permettono alle persone di interagire l'un l'altro – la quale è diventata fondamentale. Molte volte non ci si sofferma a pensare che cosa vi è dietro un'assistente digitale come Google Duplex od Alexa, che sono in grado di cercare informazioni per noi, effettuare pagamenti e prenotazioni, organizzare l'agenda tramite comandi vocali e gestire dispositivi di tipo smart.

Essi sono il frutto di un lavoro iniziato qualche anno dopo la fine della seconda guerra mondiale con Padre Busa che volle digitalizzare il lavoro di Tommaso D'Aquino per svolgere i suoi studi, questo portò alla creazione della linguistica computazionale. Successivamente, il linguista Chomsky studiò come i bambini sono capaci di apprendere una lingua prima di ricevere una istruzione formale ed i risultati di questo studio formarono la Grammatica Generativa di Chomsky e le classi omonime, che per decenni sono state le fondamenta per la creazione di linguaggi di programmazione ed algoritmi capaci di comprendere le lingue naturali. Dopo, c'è stato un susseguirsi di studi e di esperimenti che hanno sviluppato reti neurali artificiali capaci di eseguire ragionamenti complessi ed imparare dagli avvenimenti passati, come il cervello umano. Questi algoritmi rivoluzionari hanno creato assistenti digitali, traduttori neurali e softwares che analizzano dati, immagini, scremano i candidati in un processo di selezione e forniscono previsioni sulle preferenze dei clienti, analizzando ciò che comprano e ciò che pensano dell'azienda tramite i sondaggi sull'indice di gradimento.

Questa tesi analizza nel dettaglio lo sviluppo cronologico e tecnico della relazione tra l'intelligenza artificiale e le lingue naturali, insieme ai lati positivi e negativi che ne derivano.

1. Fondamenti sull'intelligenza artificiale, tentativi di creare macchine capaci di comprendere i linguaggi naturali

L'intelligenza artificiale è una branca dell'informatica che studia la progettazione di sistemi software capaci di emulare processi intellettivi umani come ragionare ed imparare dalle esperienze passate.¹ Uno degli obiettivi che le compagnie informatiche hanno sempre perseguito è conferire ai calcolatori l'abilità di interagire con le persone, tramite un linguaggio naturale per poter svolgere funzioni come: tradurre un documento, organizzare un'agenda in modo automatizzato e assistere un dottore durante una operazione. Per svolgere questi processi bisogna ricorrere all'unione di varie discipline e una fondamentale è la linguistica computazionale, la quale si concentra nello sviluppo di sistemi volti ad analizzare, interagire e comprendere uno o più linguaggi naturali.²

Nel 1946, il gesuita Padre Roberto Busa divenne il pioniere della linguistica computazionale perché, con l'ausilio dei calcolatori IBM a schede perforate, fu il primo al mondo ad usare un calcolatore per analizzare l'immensa opera Omnia di Tommaso D'Aquino. In 30 anni di lavoro creò l'Index Thomisticus formato da 70 mila pagine con un totale di 20 milioni di righe e 11 milioni di schede perforate che analizzano ogni parola.³

¹ Cfr. IBM Cloud Education, [What is Artificial Intelligence \(AI\)? | IBM](#), 3/06/2020

² Cfr. Salvatore Sorce, Introduzione alla linguistica computazionale, [Microsoft PowerPoint - SEI1213_01_Linguistica_Computazionale.ppt \(unipa.it\)](#), p. 5.

³ Martina Dell'Utri, [International Busa Machines: come un gesuita ha insegnato la lettura ai computer | Controverso](#), 11/12/2020.



La foto ritrae Padre Buse ed un ingegnere della IBM mentre crea l'Index Thomisticus su schede perforate.

Fonte: [International Buse Machines: come un gesuita ha insegnato la lettura ai computer | Controverso](#)

Successivamente nel 1950, il matematico ed ingegnere Alan Turing, pubblicò sulla rivista accademica *Mind*⁴ l'articolo “*Computing Machinery and Intelligence*” in cui cerca di trovare una risposta alla domanda che lo rese uno dei fondatori dell'intelligenza artificiale “Possono pensare le macchine?”.⁵

Nel tentativo di trovare una risposta ideò il famoso “*turing test*”, chiamato originariamente “*imitation game*”, in cui 2 soggetti, una persona ed un computer, devono rispondere alle domande di un moderatore, che deve individuare chi dei due soggetti sia il computer in base alle risposte ricevute. Per ora mai nessun computer è stato in grado di rispondere in modo tale da far pensare al moderatore di essere una persona.⁶

⁴ *Mind* è una rivista accademica che tratta l'analisi della filosofia in maniera analitica della Università di Oxford. Fonte: [About | Mind | Oxford Academic \(oup.com\)](#)

⁵ Cfr. Alan Turing, *Mind*, Volume LIX, Issue 236, October 1950, Oxford, P. 434. [I.—COMPUTING MACHINERY AND INTELLIGENCE | Mind | Oxford Academic \(oup.com\)](#)

⁶ Ivi pp. 436-439.

Nel 1957, il linguista Noam Chomsky, con gli studi di una vita su come si sviluppa il linguaggio naturale⁷ negli esseri umani e sulla capacità dei bambini di comunicare senza avere una istruzione formale, contribuì a codificare degli algoritmi capaci di analizzare e creare frasi conformi alla grammatica della lingua scelta, che avevano un senso compiuto.⁸ Chomsky elaborò la grammatica generativa, cioè una serie di regole che descrivono la formulazione di frasi di un linguaggio naturale, prendendo spunto dalla grammatica formale, focalizzandosi sui processi che gestiscono le facoltà psichiche e su come il cervello formula le frasi.⁹ La grammatica generativa crea in maniera ricorsiva le formule di un linguaggio ed evidenzia la capacità innata dei bambini di formare e comprendere frasi mai sentite in precedenza ed è grazie a questa capacità che le persone possono capire se una frase abbia un senso grammaticale o meno. Il numero di frasi che la grammatica generativa può creare è strettamente relazionato al numero di frasi che il cervello umano è in grado di immagazzinare.¹⁰

L'insegnamento della grammatica formale – studiare le regole di una lingua- avviene a scuola, tuttavia Chomsky formulò questa teoria perché notò che gli esseri umani imparano a parlare fin dalla tenera età ed a distinguere inconsciamente i vari elementi che costituiscono una frase – soggetto, predicato verbale e complemento – e questo lo portò a dedurre che gli esseri umani posseggono una parte del cervello

⁷ Il linguaggio naturale è un mezzo di comunicazione che un insieme di individui usa per comunicare; italiano, inglese, francese sono esempi di linguaggi naturali.

⁸ Cfr. Francesca Fiore, Il linguaggio universale secondo Noam Chomsky – Introduzione alla Psicologia, [Il linguaggio universale secondo Noam Chomsky - Psicologia \(stateofmind.it\)](http://stateofmind.it), 03/12/15.

⁹ Cfr. Generative Grammar, [Generative grammar | Britannica](http://www.britannica.com)

¹⁰ Cfr. [Grammatica generativa - Wikipedia](http://it.wikipedia.org/wiki/Grammatica_generativa)

specializzata nella gestione del linguaggio.¹¹ Ad esempio il bambino inconsciamente impara che si dice:

Elisa si protegge dalla pioggia con l'ombrello.

Anziché: Elisa protegge sé stessa dalla pioggia con l'ombrello.

Oppure comprende che se non si vuole ripetere il nome della persona che riceve l'azione, si può utilizzare il pronome personale di complemento tonico:

Puoi dare una matita a Marta?

Le puoi dare una matita?

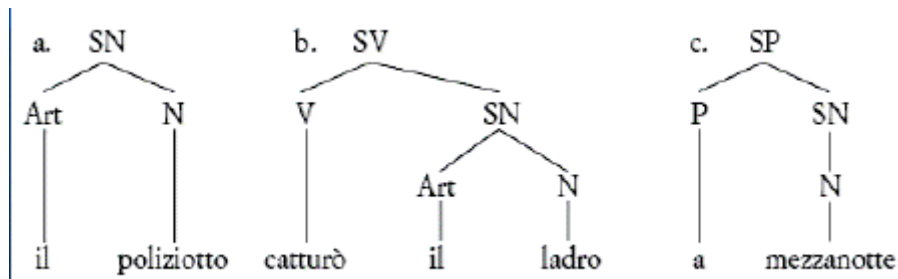
Chomsky condusse uno studio su un campione di bambini di età non superiore ai 10 anni, madre lingua inglese, per comprendere il loro grado di padronanza della lingua. Il loro compito era trasformare le frasi da attive a passive ed in domande, il risultato lasciò Chomsky ed i suoi colleghi stupiti in quanto tutti riuscirono nell'intento. Questo è il motivo per cui Chomsky trasse la conclusione che gli esseri umani nascono con un insieme di conoscenze innate volte a rintracciare la struttura grammaticale nelle frasi che ascoltano fin da piccoli.¹² Dato che tale competenza serve per imparare qualsiasi lingua, Chomsky creò la teoria della grammatica universale, che fornisce una serie di competenze innate – creare strutture grammaticali e comprendere se una frase è logica - che aiutano i bambini ad apprendere la lingua usata dalle persone intorno a loro.

Tale grammatica è un sistema che facilita gli infanti durante l'apprendimento del linguaggio, infatti imparano che le parole hanno un

¹¹ Cfr. Lettere e Filosofia, Unife, Riassunto Teoria Chosmiana, pp. 3-4 [Lezione9_2.ppt \(unife.it\)](#), 2014

¹² Cfr. [The Chomsky Experiments | by Thomas Roeper | The New York Review of Books \(nybooks.com\)](#) 16/03/1981

ordine logico – soggetto, predicato e complemento - e ogni sua parte ha un compito ben specifico.¹³



Questo è lo schema mentale che inconsciamente i bambini fanno quando imparano a parlare. Chomsky lo ha nominato diagramma ad albero della grammatica generativa. Sintagma nominale (SN); Sintagma avverbiale (SV); Sintagma preposizionale (SP); Articolo (ART), Nome (N); Verbo (V); Preposizione (P)

Fonte: [Martha's Vineyard: Diagramma ad albero e Schema X-barra \(labovsvineyard.blogspot.com\)](http://Martha's Vineyard: Diagramma ad albero e Schema X-barra (labovsvineyard.blogspot.com))

Chomsky, in seguito, analizzando una lingua combinando l'aspetto matematico che quello mentale, studiando la grammatica universale e come i bambini riescono a comunicare senza aver una istruzione formale della lingua, ha aperto la strada alla progettazione di algoritmi che si avvicinano al ragionamento umano. La formalizzazione della grammatica universale e delle varie grammatiche generative portò alla creazione della famosa gerarchia di Chomsky composta da quattro gruppi:¹⁴

La grammatica di tipo 0, racchiude tutte le grammatiche formali ed è accettata in qualsiasi macchina di Turing.

La grammatica di tipo 1 è un tipo di grammatica sensibile al contesto, in quanto nei linguaggi naturali spesso accade che una data parola possa essere considerata adeguata solo in un contesto specifico.

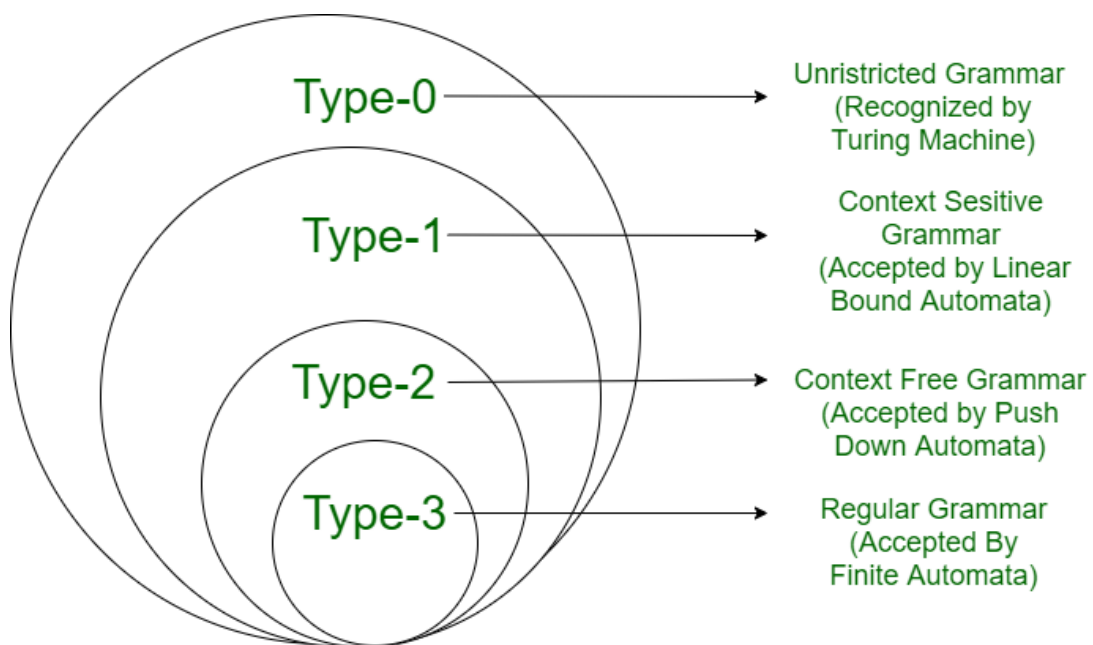
La grammatica di tipo 2 è libera dal contesto perché può essere applicata ovunque. Con questo metodo ogni frase è un blocco, ed ogni

¹³ Ibid. Lettere e Filosofia, Uniefe, Riassunto Teoria Chomskiana, pp. 6-16 [Lezione9_2.ppt \(unife.it\)](http://Lezione9_2.ppt(unife.it)), 2014

¹⁴ Cfr. James Rogers, Formal language theory: refining the Chomsky hierarchy, [Formal language theory: refining the Chomsky hierarchy | Philosophical Transactions of the Royal Society B: Biological Sciences \(royalsocietypublishing.org\)](http://Formal language theory: refining the Chomsky hierarchy | Philosophical Transactions of the Royal Society B: Biological Sciences (royalsocietypublishing.org)), pp. 2-20, 19/07/2012

elemento di essa è un sotto blocco. Quindi soggetto e verbo sono due blocchi che possono essere utilizzati in modo indipendente. La libertà di questa grammatica offre la possibilità di elaborare algoritmi per l'analisi lessicale di frasi.

La grammatica di tipo 3 può generare solamente lingue regolari, cioè un insieme di caratteri utili per cercare una parola esatta in una frase o in documento. Questo tipo di grammatica ha dato la possibilità di sviluppare gli algoritmi di ricerca di parole e frasi in documenti che tutt'ora usiamo nei browsers e nei vari software di scrittura e traduzione.



Il grafico mostra la gerarchia delle grammatiche di Chomsky.

Fonte: [Chomsky Hierarchy in Theory of Computation - GeeksforGeeks](#)

Gli studi del linguista Chomsky - la grammatica generativa, la grammatica universale e la gerarchia da cui prende il suo nome - sono state le basi degli algoritmi degli anni 50 e 70 del secolo passato, che fornivano la capacità ai calcolatori di elaborare stringhe di caratteri, formulare semplici domande e poter descrivere scene non troppo complesse.

Tra il 1959 ed il 1963 John Backus e Peter Naur, scienziati ed informatici statunitensi presso l'IBM, elaborarono la “forma normale di Backus-Naur”, la quale stabilisce le regole formali da usare per illustrare in modo chiaro la sintassi di un linguaggio di programmazione e dei protocolli di rete.¹⁵ Backus e Naur si sono ispirati alla gerarchia di Chomsky ed alla logica della grammatica inglese per stabilire le regole sintattiche che i linguaggi di programmazione dovrebbero avere, nello specifico ogni componente della frase ha un simbolo dedicato che permette ai programmatori di scrivere pezzi di codice in un modo veloce.

```
<indirizzo postale> ::= <destinatario> <indirizzo> <localita>  
<destinatario> ::= [<titolo>] [<nome>|<iniziale>] <cognome> <a capo>  
<indirizzo> ::= <via> <numero civico> <a capo>  
<localita> ::= [<CAP>] <comune> <provincia>
```

L'immagine mostra un esempio di sintassi nella forma normale di Backus e Naur per scrivere un indirizzo postale.

Fonte: [Backus-Naur Form - Wikipedia](#)

Un prototipo unico nel suo genere, che diede uno stimolo considerevole allo sviluppo di programmi capaci di interagire con gli esseri umani tramite un linguaggio naturale in un ambiente artificiale a due dimensioni, fu il programma SHRDLU ¹⁶ sviluppato da Terry Winograd. Tutto ebbe inizio nel 1968 al MIT (Massachusetts Institute of Technology), quando Marvin Minsky ¹⁷, considerato il padre fondatore dell'unione tra l'intelligenza artificiale e la filosofia, lavorava alla progettazione di un calcolatore che potesse comunicare con le persone,

¹⁵ Cfr. [Backus-Naur Form | Sapere.it](#)

¹⁶ Maggiori informazioni, Nigel Word, Tokyo University, Introduction to Shrdlu, [S00056 1..4 \(utep.edu\)](#)

¹⁷ Cfr. Michael Aaron Dennis, [Marvin Minsky | American scientist | Britannica](#), 5/08/2021

spostare oggetti geometrici, chiedere ulteriori informazioni nel caso ricevesse comandi ambigui in un ambiente a due dimensioni, essere capace di distinguere i colori e le forme degli oggetti. Per sviluppare la parte comunicativa, Minsky propose allo studente Terry Winograd¹⁸ se come tesi per il suo dottorato di ricerca volesse ideare un algoritmo affinché la macchina potesse capire istruzioni in linguaggio naturale. Nei due anni successivi, Terry Winograd sviluppò un algoritmo capace di comprendere un linguaggio naturale (inglese) unendo gli studi di sintassi, semantica e metodo deduttivo.¹⁹

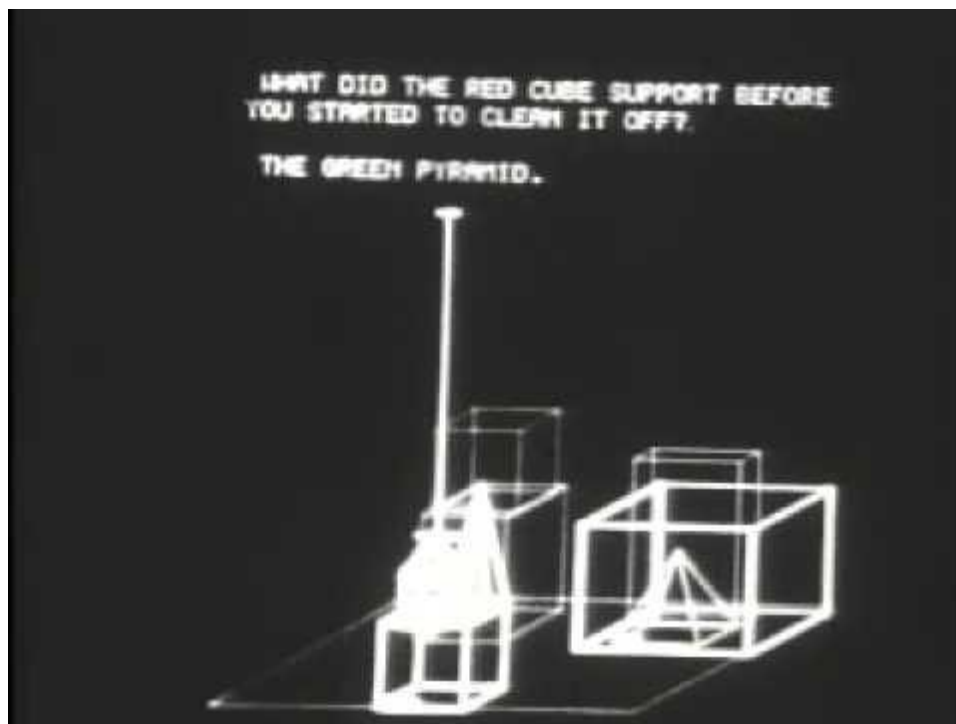


Immagine di repertorio mentre Winograd interagisce con Shrdlu versione 2.0

Fonte: [winograd shrdlu - YouTube](#)

Nel 1949, l'ingegnere Warren Weaver²⁰ scrisse il famoso vademecum intitolato "Translation", per la Fondazione Rockefeller, in cui

¹⁸ Maggiori informazioni su Terry Winograd. [Terry Winograd - Wikipedia](#)

¹⁹ Cfr. Terry Winograd, Procedures as a representation of data in a computer program for understanding natural language, MIT, January 1971, pp. 18-24. [AITR-235.pdf \(stanford.edu\)](#)

²⁰ Warren Weaver era uno scienziato e matematico statunitense, considerato il padre della traduzione automatica.

per la prima volta nella storia si ipotizza la creazione di programmi capaci di tradurre senza la supervisione dell'uomo. La tecnica da lui descritta consisteva nel sostituire ogni parola del testo con quella corrispondente, usare la statistica applicata per calcolare quanto spesso ogni parola era usata in uno specifico contesto, e nel caso di omofoni, prima si traducono le altre parole per comprendere il significato della frase.²¹ Tuttavia, uno dei limiti era la accuratezza della traduzione di frasi lunghe, che spesso avevano errori grammaticali o erano prive di senso, e per risolverlo si fece ricorso alla grammatica generativa di Chomsky.²² Nel 1952, un altro problema sull'accuratezza fu sollevato da Yehoshua Bar-Hillel²³ – accademico e professore del Massachusetts Institute of Technologies – il quale era convinto che per avere una traduzione automatica era necessario sacrificare l'accuratezza in quanto nessun algoritmo era capace di capire le ambiguità linguistiche.²⁴ Tuttavia per risolvere questo problema, Yehoshua Bar-Hillel propose l'idea innovativa di dotare i calcolatori di una enciclopedia universale.²⁵

L' IBM, insieme al MIT, alla Brown University e la Oxford University iniziarono la creazione di corpora²⁶ linguistici che raggruppavano testi e saggi divisi per categoria ed ogni parola era contrassegnata con una etichetta contenente informazioni relative al suo significato ed uso.²⁷

²¹ Cfr. Warren Weaver, Translation, pp. 15-24, [weaver001.pdf \(stanford.edu\)](#)

²² Cfr. Warren Weaver, Wikipedia, sezione Translation. [Warren Weaver - Wikipedia](#)

²³ Cfr. Kasher, Essay in memory of Yehoshua Bar-Hillel, Boston, pp. 1-2.

²⁴ Cfr. Sergei Nirenburg, Bar Hillel and Machine translation: then and now, New Mexico State University, 1999, pp. 3-4. [BISFAI95-027.pdf \(aaai.org\)](#)

²⁵ Ivi. Pp. 5-6.

²⁶ Corpus al singolare, corpora al plurale.

²⁷ Cfr. Brown University, Brown Corpus Manual For Use With Digital Computer, USA, 1964, pp. 3-8. [Brown Corpus Manual \(uib.no\)](#)

Solamente dal 1975 in poi con l'introduzione di Systran si raggiunse una migliore accuratezza nella traduzione, solo specializzata nel campo dell'aeronautica.²⁸

Nel decennio 1980-1990, l'IBM reintrodusse la traduzione automatica statistica che si basa sull'analisi dei corpora bilingue.

Negli anni 2000, Franz Josef Och fu a capo del progetto Google Traduttore, ed implementò lo stesso sistema usato dall' IBM con più corpora a disposizione, ma dopo qualche anno, Google introdusse il sistema di identifica automatica della lingua.²⁹

Nel 2016, la compagnia decise di cambiare il modello di base passando alla traduzione automatica neurale, di conseguenza usando algoritmi che si avvicinano di più al ragionamento umano.³⁰

²⁸ Cfr. [The Lingua File from TLF Translation: SYSTRAN: A Brief History of Machine Translation](#)

²⁹ Cfr. History of Translation. [Google Translate - Wikipedia](#)

³⁰ Cfr. Joe Sommerlad, GOOGLE TRANSLATE: HOW DOES THE MULTILINGUAL INTERPRETER ACTUALLY WORK?, Marzo 2021. [Google Translate: How does the multilingual interpreter actually work? | The Independent | The Independent](#)

2. Modalità usate dai computer per analizzare i linguaggi umani; la tecnologia all'avanguardia del GPT-3

Le invenzioni e le teorie spiegate nel capitolo precedente hanno spianato la strada a software capaci di scrivere testi di varia lunghezza e tradurre testi in modo autonomo e comunicare con le persone tramite i linguaggi naturali.

Nel marzo 2020, la compagnia statunitense OpenAI³¹ immette nel mercato informatico un modello di linguaggio chiamato *Generative Pre-trained Transformer 3* (Trasformatore generativo pre-addestrato 3), abbreviato a GPT-3.



Logo del software GPT-3. Fonte: [GPT-2 \(GPT2\) vs. GPT-3 \(GPT3\): The OpenAI Showdown - DZone AI](#)

GPT-3 è un modello di linguaggio auto-regressivo, il quale sfruttando il sistema dell'apprendimento profondo neurale è in grado di creare testi corti e di media lunghezza, con un massimo di 800 parole,

³¹ OpenAI è un'azienda di ricerca e sviluppo per strumenti basati sull'intelligenza artificiale, con la missione di offrire servizi e strumenti per il bene della collettività.

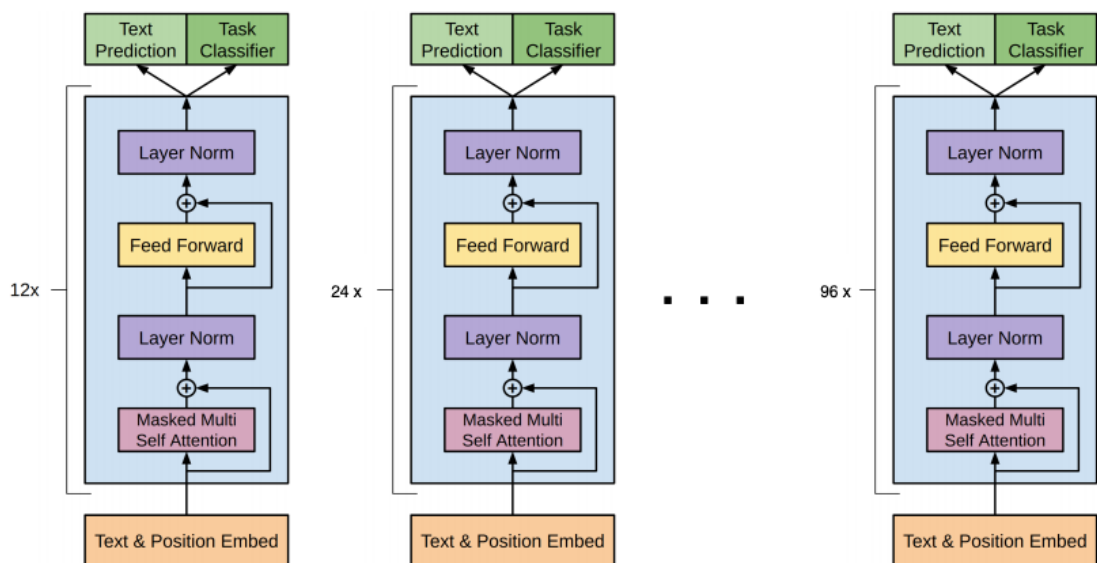
molto simili ad un testo scritto da una persona ed è capace di soddisfare anche altre richieste al di fuori del campo linguistico.³²

La sua capacità di creare testi molto accurati è data dalla presenza di 175 miliardi di coefficienti che si auto-ottimizzano, imparano dagli eventi passati mentre il software svolge i compiti assegnati. Basandosi su un sistema di “apprendimento profondo” e per essere in grado di produrre dei testi ed altre funzioni nel modo più simile agli umani, la compagnia OpenAI ha usato come base della conoscenza di GPT-3 un enorme gruppo di testi e corpus provenienti da vari siti internet tra cui CommonCrawl e Wikipedia Corpus, ed una grande raccolta di volumi. GPT-3 ha analizzato una considerevole quantità di informazioni che sarebbe difficile farlo per una singola persona.³³

GPT-3 è stato ideato principalmente per tradurre, ideare testi di qualsiasi genere, ma essendo un software poliedrico può essere impostato per altre funzionalità come generare quiz, siti internet, identificare dipinti e consigliare libri in base alle proprie preferenze.

³² Cfr. Aman Anad, Deep Learning Trends: top 20 best uses of GPT-3 by OpenAI, [Deep Learning Trends: top 20 best uses of GPT-3 by OpenAI \(educative.io\)](#), 20/09/2020

³³ Amol Mavuduru, What is GTP-3 and why is it so powerful? [What is GPT-3 and why is it so powerful? | Towards Data Science](#), 17/02/2021



L'immagine mostra il diagramma ricorsivo del software GPT-3.

Fonte: [GPT-3 An Overview · All things \(dzlab.github.io\)](https://github.com/dzlab/gpt-3)

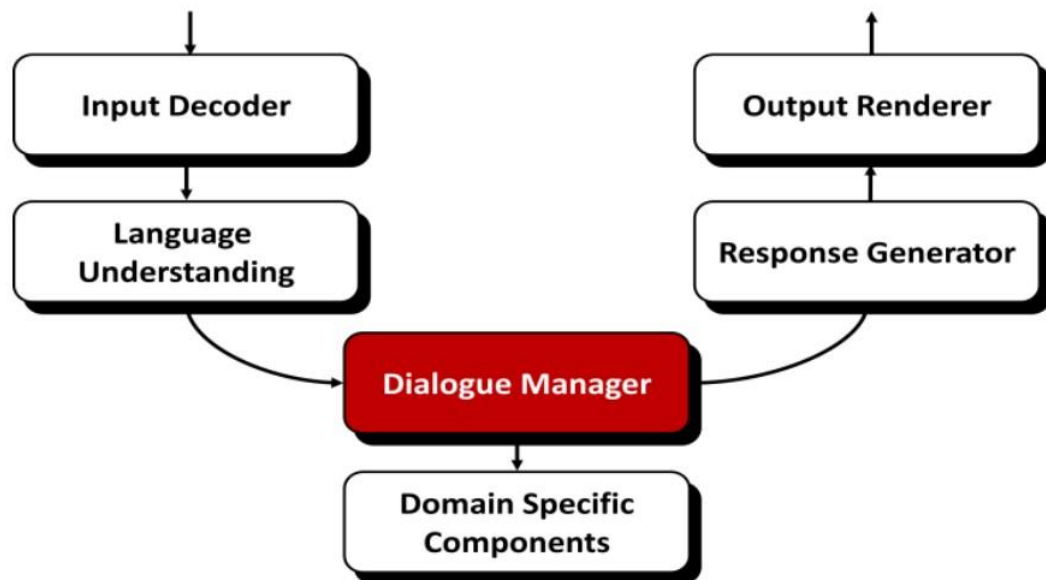
Il giornale britannico The Guardian, per testare le effettive funzionalità, chiese al software GPT-3 di scrivere un saggio per convincere gli umani che le macchine robotiche sono pacifiche, il software ne creò otto da 500 parole ciascuno, con diversi stili e ragioni che supportano la idea di cooperazione tra macchine ed esseri umani. È impercettibile capire se i saggi siano stati elaborati interamente da una macchina.

Nei saggi, GPT-3 spiega che non è nel suo interesse agire contro gli esseri umani ed essere onnipotente, inoltre, non è interessato alla violenza bensì a perseguire l'obiettivo di aiutare le persone a raggiungere i loro traguardi. Oltretutto, sottolinea che senza gli esseri umani l'intelligenza artificiale e le macchine vagherebbero nell'etere digitale senza avere nessun contatto col mondo esterno.³⁴

³⁴ Cfr. The Guardian, A robot wrote this entire article. Are you scared yet, human?, [A robot wrote this entire article. Are you scared yet, human? | GPT-3 | The Guardian](https://www.theguardian.com/technology/2020/nov/15/gpt-3-robot-wrote-article) 15/11/2020

2.1 Spoken Dialogue System

Un altro traguardo dell'intelligenza artificiale e della linguistica è il *spoken dialogue system*. Un sistema informatico capace di interagire con gli utenti tramite un linguaggio naturale per fornire supporto al fine di accedere alle funzionalità del sistema. Sebbene si è sempre pensato che l'architettura di questo sistema sembri facile, non lo è perché ci sono varie fasi e componenti dell'architettura in questione per far sì che la macchina analizzi la voce dell'utente, elabori e restituisca una risposta.



Il diagramma mostra i vari componenti dello Spoken Dialogue System

Fonte: Punjab Technical University, [1306.4134.pdf \(arxiv.org\)](#)

Il primo stadio è il decodificatore di entrata, *Input Decoder*, che riceve la richiesta vocale e la converte in un testo scritto. Questo componente possiede gli algoritmi per la codifica della fonetica³⁵ e della

³⁵ La fonetica studia il suono delle parole sia singolarmente che insieme e come sono rappresentate da simboli scritti.

fonologia³⁶ perché deve analizzare il suono delle parole e la loro combinazione.³⁷

Il secondo stadio è la comprensione vera e propria di ciò che la persona ha detto tramite un modulo che analizza i linguaggi naturali.³⁸ Usando le regole della sintassi³⁹ e della semantica⁴⁰ il programma cerca le parole chiavi e prova a dare un significato ad esse in base al contesto. Fatto ciò si passa “*dialogue manager*”.

Quest’ultimo ha la funzione di ricevere la rappresentazione semantica della frase e comprendere come il testo si possa adattare al contesto generale. Inoltre, decide quale risposta bisogna fornire all’utente, svolge la funzione di archiviazione di dati, risolve le ambiguità linguistiche e immagazzina la cronologia delle frasi ricevute precedentemente.⁴¹

Dopo che il “*dialogue manager*” ha elaborato la frase, quest’ultima può essere analizzata da un componente che esegue specifiche attività, come tradurla e confrontarla con un’altra. Dopo, la frase si rimanda al “*dialogue manager*” che a sua volta la manderà al generatore di risposta (*response generator*).

Tale generatore rifinisce la frase creata dal “*dialogue manager*”, decide quali informazioni includere e quale registro linguistico usare.⁴²

Infine vi è il generatore del discorso (*speech generator*) che deve trasformare la frase in un messaggio vocale e lo può fare con due diversi approcci. Il primo approccio è di utilizzare frasi già registrate con spazi

³⁶ La fonologia studia il sistema di suoni che compongono la parte orale di un linguaggio naturale.

³⁷ Cfr. Matthew Zajechowski, [Automatic Speech Recognition \(ASR\) Software - An Introduction - Usability Geek](#)

³⁸ IBM Cloud Education, [What is Natural Language Processing? | IBM](#), 2/07/2020

³⁹ La sintassi studia la costruzione grammaticale delle parole per formare una proposizione.

⁴⁰ La semantica studia il significato delle parole.

⁴¹ Cfr. Suket Arora, Kamaljeet Batra, Sarabjit Singh, Dialogue System: A Brief Review, Punjab University, pp. 2.

⁴² Ibid. Dialogue System: A Brief Review

vuoti da riempire a secondo della frase, mentre il secondo approccio è di creare un messaggio vocale della frase con parole singole pre-registrate e poi messe in ordine.

Il *spoken dialogue system* può essere implementato in tre diversi modi: struttura a stato finito (*finite state based system*), struttura a sistema predefinito (*frame based system*) e struttura con agente (*agent based system*).

La struttura a stato finito guida la persona attraverso una sequenza preimpostata di passaggi. Un esempio è il navigatore satellitare oppure l'assistente telefonico o di messagistica istantanea delle compagnie telefoniche. Questo tipo di struttura è facile da costruire e bisogna inserire tutti i vari passaggi creando il percorso che porterà la persona alla possibile risoluzione della sua richiesta.⁴³

La struttura a sistema predefinita è un percorso predefinito di domande a cui la persona deve rispondere. Le risposte ricevute saranno inserite nei vari moduli al fine di capire quali compiti la macchina deve eseguire, di conseguenza il dialogo tra la persona e la macchina dipende dalle risposte che l'interlocutore fornisce. Un esempio è il sistema di valutazione digitale della salute della persona implementata in certi paesi del nord Europa, in cui prima di parlare con un medico la persona deve rispondere a certe domande, così in base alle risposte si può indirizzare la persona dal medico appropriato. Questo tipo di struttura permette di avere una conversazione più rapida.⁴⁴

Il sistema con agente permette all'utente di parlare liberamente come se lo facesse con un altro essere umano, in quanto l'agente digitale

⁴³ Cfr. Hagen Böhm, Dialogue management using Finite State Models, Saarland University, 2002, pp. 4-10

⁴⁴ Ivi. Cfr. Suket Arora, Kamaljeet Batra, Sarabjit Singh, Dialogue System: A Brief Review, Punjab University, pp. 3

è in grado di ragionare ed argomentare le proprie risposte, ricorda la cronologia della conversazione e si adatta al contesto. In questo modo la comunicazione tra l'utente e l'assistente si evolve in maniera naturale senza l'ausilio di percorsi o domande predefinite. Questo tipo di sistema è intuitivo e permette alla persona di parlare in modo naturale e di risolvere problematiche relative ad argomenti complessi. Tuttavia questo tipo di architettura è ancora in fase di sviluppo.

2.2 Traduzione automatica basata su regole (Rule-Based Machine Translation)

Lo sviluppo di traduttori automatici è stata una sfida ardua con pochi risultati fino al 1970 con la progettazione del primo traduttore automatico Systran, e da quel momento in poi i traduttori divennero una risorsa fondamentale della società odierna.⁴⁵

Vi sono vari approcci alla traduzione automatica.

Il primo si chiama “traduzione automatica basata su regole” (*rule-based machine translation*), la quale usa le informazioni fornite da dizionari e testi di grammatica sia della lingua da tradurre che quella di arrivo. Tramite l’analisi morfologia, sintattica e semantica questo traduttore prova a comporre le parole tradotte in modo tale da avere una frase di senso compiuto.⁴⁶ La “traduzione automatica basata su regole” ha tre diversi approcci:

- Sistema diretto (*Dictionary-based machine translation*): ogni parola della frase inserita, sarà tradotta tramite l’uso di un dizionario bilingue. Questo approccio può fornire frasi senza essere correlate con il significato della lingua di partenza.⁴⁷
- Sistema di trasferimento (*Transfer-Based Machine Translation*): è l’approccio più usato fra quelli basati su regole, perché si possono ottenere risultati migliori rispetto alla tecnica precedente. La prima fase consiste nell’analizzare grammaticalmente la frase, poi l’algoritmo tenta di capirne il contesto e prova a trovare un corrispettivo della struttura

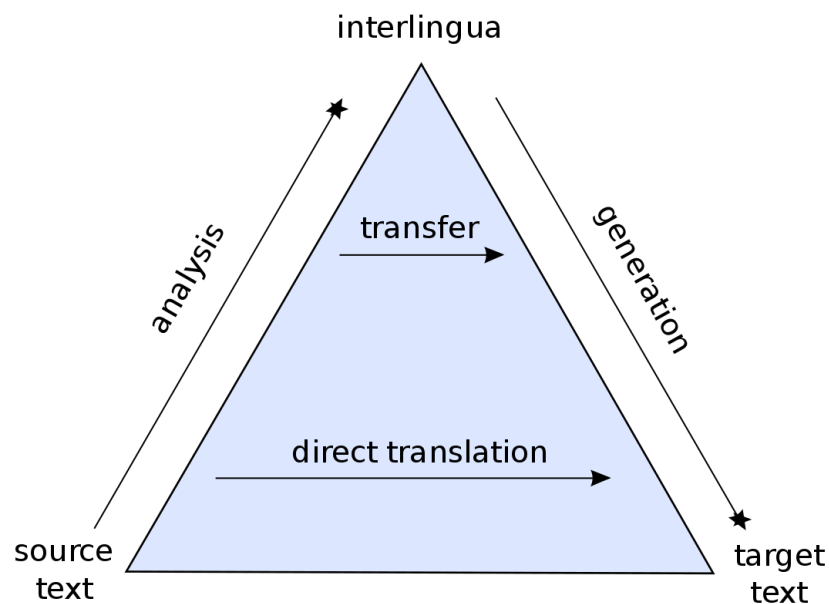
⁴⁵ Cfr. History Section, [Rule-based machine translation - Wikipedia](#)

⁴⁶ Cfr. Kantan, Mt. RBMT vs SMT, [RBMT vs SMT – KantanMT – Machine Learning – Neural Engines – AI – Language Technology \(kantanmtblog.com\)](#), 02/2021

⁴⁷ Cfr. Agato Translation Company, WHAT IS DICTIONARY-BASED MACHINE TRANSLATION?, [What is Dictionary-Based Machine Translation? \(agatotranslate.ae\)](#)

grammaticale della lingua sorgente con quella di arrivo. Dopo la fase analitica, si traduce la frase, si controlla la concordanza del genere e del numero, ed infine viene restituita all'utente.⁴⁸

Traduzione tramite interlingua (*Interlingual machine translation*): è considerato l'approccio più classico quando si usa la tecnica basata su regole. La lingua sorgente è prima tradotta in una lingua neutra, che spesso è l'inglese, per poi essere tradotta nella lingua di destinazione. In questo modo, non servono dizionari bilingue, si può aumentare il numero delle lingue che si possono tradurre perché le loro grammatiche saranno sempre confrontate con quella dell'interlingua.⁴⁹



L'immagine mostra la piramide di Bernard Vauquois, il quale esplica come funzionano i tre metodi della traduzione automatica basata su regole.

Fonte: [1: Bernard Vauquois' pyramid | Download Scientific Diagram \(researchgate.net\)](#)

⁴⁸ Cfr. Transfer-based machine translation, [Transfer-based machine translation - Wikipedia](#)

⁴⁹ Cfr. R. Richsen, Interlingual Machine Translation, pp. 145-147 20/08/2021

2.3 Traduzione automatica statistica (*Statistical machine translation*)

Nella metà degli anni 80, del secolo precedente, gli studiosi del centro di ricerca della IBM Thomas Watson⁵⁰ re-introducono un altro approccio di traduzione automatica il quale diverrà predominante fino all'inizio del ventunesimo secolo, per via del suo metodo statistico.⁵¹

L'approccio si chiama traduzione automatica statistica, (*Statistical machine translation*) SMT, e si avvale di algoritmi statistici per la previsione del miglior output traduttivo, perché usano l'analisi statistica al fine di elaborare le regole più consone per realizzare la frase di arrivo. Questo approccio necessita di una grande quantità di corpus linguistici sia nella lingua di partenza che in quella di arrivo per creare la traduzione più probabile e per imparare a prevedere come tradurre un testo in occasioni future. La frase tradotta deve essere considerata come quella con la più alta probabilità di essere quella corretta e non da considerare come univocamente giusta.⁵²

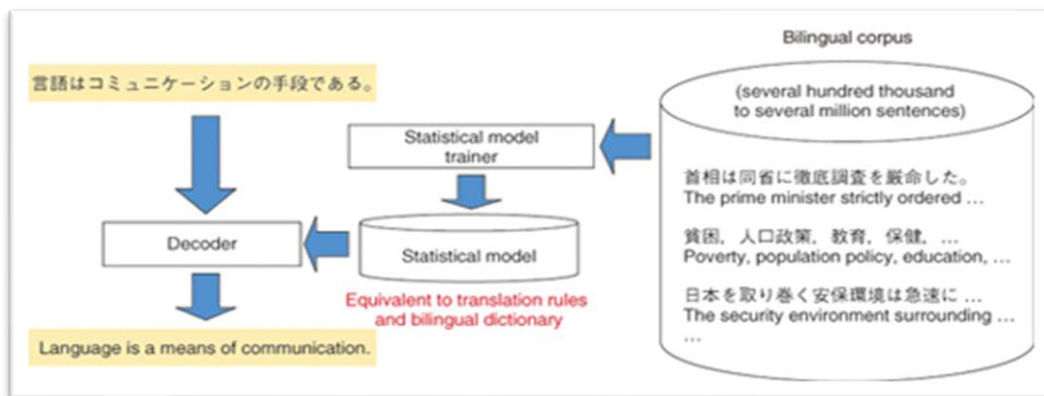
Inoltre per assicurare un output di qualità i corpora linguistici, dovrebbero essere specializzati in un argomento e non essere generali. Più corpus linguistici l'algoritmo può analizzare in entrambe le lingue più il risultato sarà ottimale, infatti un corpora valido necessita di un minimo di 100 milioni di parole e di un milione di frasi allineate.⁵³

⁵⁰ Per maggiori informazioni: [Thomas J. Watson Research Center - Locations \(ibm.com\)](https://www.ibm.com/press/us/2017/01/01/ibm-research-center-locations)

⁵¹ Ilya Pestov, FreeCodeCamp, Statistical Translation, [A history of machine translation from the Cold War to deep learning \(freecodecamp.org\)](https://www.freecodecamp.org/news/a-history-of-machine-translation-from-the-cold-war-to-deep-learning/)

⁵² Cfr. Michel Nielsen, Introduction to Statistical Machine Translation, [Introduction to Statistical Machine Translation | Michael Nielsen](https://www.ibm.com/press/us/2017/01/01/ibm-research-center-locations), 11/03/2009

⁵³ Cfr. Nabeel T. Alsohybe, Neama Abdulaziz Dahan and Fadl Mutaher Ba-Alwi, Machine-Translation History and Evolution: Survey for Arabic-English Translations, pp. 7, 19/01/2017.



Schema dell'algorithmo della traduzione automatica statistica

Fonte: [Recent Innovations in NTT's Statistical Machine Translation | NTT Technical Review \(ntt-review.jp\)](http://www.ntt-review.jp)

In pratica questo tipo di traduttore, dopo aver ricevuto il testo di partenza, avvierà gli algoritmi statistici per processare una serie di traduzioni in base ai corpora e dopo il decodificatore di lingua sceglierà quella che ha la più alta probabilità di essere la traduzione più adatta alla situazione in termini di sintassi, semantica e attinenza del contesto.

Sebbene divenne il più utilizzato dalla fine degli anni 80 fino all'avvento dei traduttori neurali, questo tipo di approccio ha degli svantaggi che hanno sempre reso il suo utilizzo e manutenzione difficili. Il SMT ha difficoltà a tradurre un testo che non abbia lo stile di scrittura dei corpora e che non si attenga al loro argomento e ha difficoltà a tradurre modi di dire ed espressioni colloquiali.⁵⁴

In conclusione, questo approccio è puramente basato su dati statistici, dunque non vi è la necessità di usare le regole di grammatica della lingua da tradurre e quella di arrivo e di concetti interlinguistici.

⁵⁴ Cfr. United Language Corp, THE PROS AND CONS OF STATISTICAL MACHINE TRANSLATION, [The Pros and Cons of Statistical Machine Translation \(unitedlanguagegroup.com\)](http://www.unitedlanguagegroup.com)

3. Cenni storici e panoramica dell'elaborazione del linguaggio naturale (*Natural Language Processing*)

Circa il tentativo di creare una macchina in grado di comunicare con l'uomo in modo intuitivo, in questo capitolo si tratterà dell'elaborazione del linguaggio naturale (*Natural Language Processing*, NLP) che ha permesso di creare assistenti virtuali come Alexa o traduttori neurali come il traduttore Google.

Fino al 1966, molte compagnie informatiche nel mondo provarono a creare un algoritmo capace di interagire con l'uomo e di tradurre un testo in modo che abbia senso. Tuttavia i risultati non coprivano gli ingenti costi e di conseguenza tutti i progetti di sviluppo furono arrestati.⁵⁵ Solamente agli inizi del 1980, le regole complesse di grammatica scritte a mano furono sostituite con i primi algoritmi dell'apprendimento automatico (*machine learning*), come i *decision trees*⁵⁶ che producevano gli stessi risultati degli algoritmi con le regole grammaticali scritte a mano, ad un costo di gestione più basso e poi si passò ai modelli statistici. Quest'ultimi necessitano di un numero elevato di corpora per apprendere la lingua, ma forniscono risultati più accurati, per questo motivo IBM negli anni 80 del secolo scorso ha sviluppato svariati modelli statistici che hanno riscosso un gran successo.⁵⁷

Dalla metà degli anni 90 con l'introduzione di internet, una quantità enorme di documenti divenne disponibile sulla rete e questo affermò l'uso dei modelli statistici che eclissarono tutti gli altri modelli di analisi del linguaggio naturale.

⁵⁵ Cfr. Keith D. Foote, A Brief History of Natural Language Processing (NLP), [A Brief History of Natural Language Processing \(NLP\) - DATAVERSITY](#) 22/05/2019

⁵⁶ Per maggiori informazioni, [Decision Trees in Machine Learning | by Prashant Gupta | Towards Data Science](#)

⁵⁷ Ivi Keith D. Foote, A Brief History of Natural Language Processing (NLP), [A Brief History of Natural Language Processing \(NLP\) - DATAVERSITY](#) 22/05/2019

Verso gli inizi degli anni 2000, con l'introduzione dei modelli a rete neurale ricorrente (*recurrent neural net*, RNN), si riuscì a creare delle valide macchine che elaborano la voce e testi e questo migliorò gli assistenti vocali e di messaggistica. I modelli di rete neurale sono tutt'ora considerati la scelta più consona per creare e comprendere un testo e un messaggio vocale.⁵⁸

Il primo assistente vocale sviluppato con l'algoritmo di elaborazione del linguaggio naturale con un modello neuronale, che ha riscosso un successo planetario, è Siri della Apple. Il suo modulo di riconoscimento vocale automatizzato trasforma le parole dell'utente in segnali digitali da interpretare ed il sistema di comando vocale (*Voice command system*) li collega alle azioni predefinite di Siri.⁵⁹

Ad esempio, se Siri chiedesse all'utente: Vuoi sapere a che ora è la riunione? Siri si aspetterebbe come risposta un sì o un no.

L'apprendimento automatico (*machine learning*) ha portato un notevole miglioramento nel capire la voce degli utenti, infatti le persone non devono attenersi a frasi predefinite e devono solamente avere un accento comprensibile affinché il sistema di elaborazione del linguaggio possa tradurre la frase in segnali digitali. Un sistema con sufficienti corpora ed addestrato bene, può intendere la differenza tra *pèsca* e *pésca* e rispondere in modo appropriato⁶⁰:

L'uomo pèsca in riva al mare

L'uomo mangia una pèsca in riva al mare.

⁵⁸Ivi Keith D. Foote, A Brief History of Natural Language Processing (NLP), [A Brief History of Natural Language Processing \(NLP\) - DATAVERSITY](#) 22/05/2019.

⁵⁹ Cfr. Aman Goel, How Does Siri Work? The Science Behind Siri, [How Does Siri Work? The Science Behind Siri - Magoosh Data Science Blog](#), 2/02/2018.

⁶⁰ Ibid. Aman Goel, How Does Siri Work? The Science Behind Siri, [How Does Siri Work? The Science Behind Siri - Magoosh Data Science Blog](#), 2/02/2018.

La cooperazione tra l'elaborazione del linguaggio naturale e il *dialogue manager* fornisce un sistema capace di interagire tramite domande e risposte con un essere umano, tuttavia anche i più moderni sistemi non riescono ancora a passare il test di Alan Turing.

È facile intendere che l'elaborazione del linguaggio naturale ha permesso il miglioramento di macchine per automatizzare le traduzioni e gli assistenti vocali e di messaggistica (voice e chat bot). Per ottenere ciò ci sono quattro passaggi che compongono l'elaborazione del linguaggio naturale.⁶¹

- Morfologia
- Sintassi
- Semantica e pragmatismo
- Fonologia

Part-of-Speech-Tagging (PoS) è l'algoritmo che si occupa dell'analisi morfologia: individuare la funzione di tutte le parole.

Per le macchine è impegnativo farlo e capire l'ambiguità delle parole lo è ancora di più, quindi per risolvere queste problematiche il PoS si avvale dell'uso dei corpora (ad esempio della Brown University e di Wikipedia Corpora), che includono milioni di parole, ciascuna dei quali è etichettata così la macchina può imparare a distinguere le parole. Per esempio i corpora insegnano alle macchine che un verbo preceduto da un articolo diventa un sostantivo.⁶²

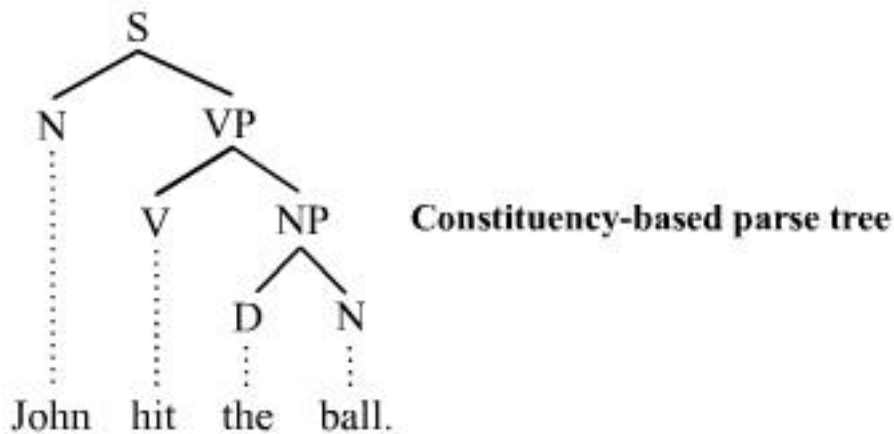
Un altro approccio è l'uso di algoritmi di autoapprendimento (*self-learning algorithms*) che ricavano le regole grammaticali man mano che

⁶¹ Cfr. Ionos, What is natural language processing?, [How does natural language processing work? - IONOS](#), 09/04/2019.

⁶² Cfr. Aiswarya Ramachandran, NLP Guide: Identifying Part of Speech Tags using Conditional Random Fields, [NLP Guide: Identifying Part of Speech Tags using Conditional Random Fields | by Aiswarya Ramachandran | Analytics Vidhya | Medium](#), 05/10/2018

leggono un corpora. Per esempio, l'algoritmo Brill Tagger determina prima la funzione della parola che appare più volte e poi le altre.

La sintassi è gestita dall'albero sintattico (*parse trees/tree diagrams*) che ha il compito di comprendere il significato della frase. Per far ciò si avvale di diagrammi ad albero per scomporre la frase in pezzi, *parsing step*, da analizzare singolarmente.⁶³



L'immagine mostra l'albero sintattico e come sono etichettate le parole. Soggetto (S), Parola (N), Sintagma avverbiale (VP), Verbo (V), Sintagma Nominale (NP), Articolo (D).

Fonte: [Parse tree - Wikipedia](#)

Lo stadio successivo è l'analisi della semantica e del pragmatismo, infatti per una macchina è difficile, durante la prima analisi, capire che significato attribuire ad una parola con più significati.

Ad esempio

Lucia aggiunge il sale alla pietanza.

Lucia sale le scale.

Per risolvere questo problema semantico, la macchina analizza le altre parole per capire il contesto e l'ausilio dei corpora, che forniscono

⁶³ Ivi. Ionos, What is natural language processing?, [How does natural language processing work? - IONOS](#), 09/04/2019.

tutti i possibili significati della parola ambigua da analizzare, e riesce a capire il senso della frase.⁶⁴

Un altro ostacolo che la macchina deve risolvere è capire quale sarebbe il contesto della frase, in quanto molte volte quest'ultima è formulata in base al mondo ed alle emozioni umane. Ancora oggi, l'elaborazione del linguaggio naturale ha difficoltà a cogliere l'ironia, concetti astratti e l'umorismo.⁶⁵

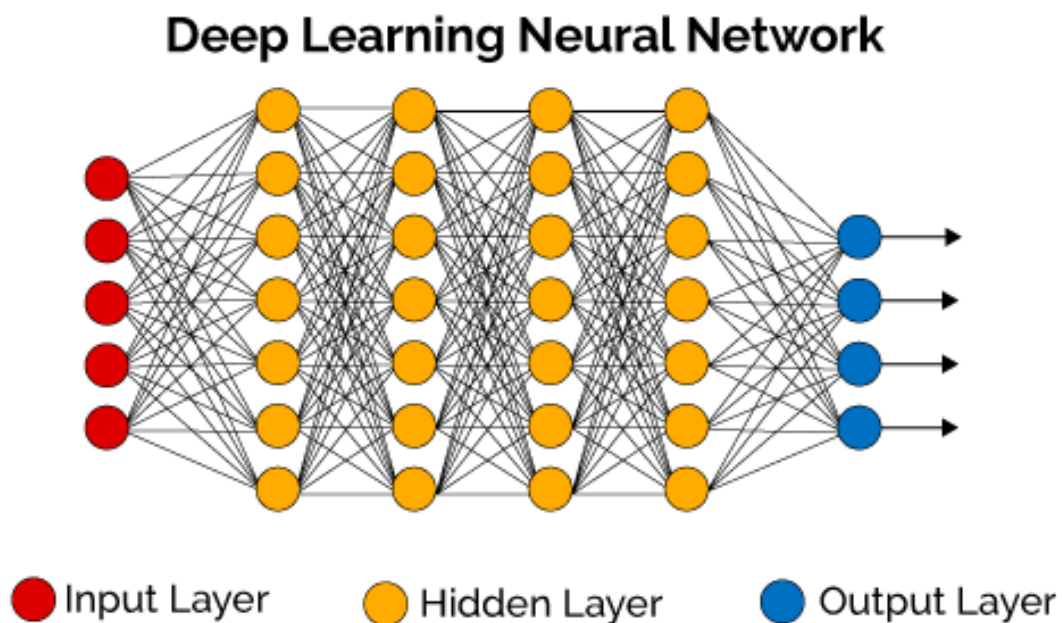
L'ultimo aspetto è la fonologia che è gestita dal *dialogue manager* trattato nel capitolo 2.1

⁶⁴ Ibid. Ionos, What is natural language processing?, [How does natural language processing work? - IONOS](#), 09/04/2019.

⁶⁵ Ionos, What is natural language processing?, [How does natural language processing work? - IONOS](#), 09/04/2019.

3.1 Apprendimento profondo (deep learning)

L'apprendimento profondo (deep learning) è un ramo di ricerca dell'apprendimento automatico (*machine learning*) e negli ultimi anni dà la possibilità alle macchine di apprendere dai risultati delle loro attività svolte nel passato, al fine di fornire dei risultati sempre migliori. I campi di applicazione spaziano dalla conversazione tra utente e macchina al riconoscimento vocale e facciale, tuttavia, per giungere questi traguardi ci sono voluti anni di studio e di esperimenti.⁶⁶



Questa immagine mostra la rete neurale di una architettura di apprendimento profondo, che spiegheremo nel dettaglio durante il prossimo capitolo. Fonte: [Neural Networks and Deep Learning – John Sisler](#)

Sebbene nella storia del 900 i progressi dell'apprendimento profondo si siano intrecciati con i concetti spiegati nei capitoli precedenti, il primo prototipo fu immesso in commercio nel 1949. Era un calcolatore che si basava sulle reti neurali della mente umana ideato da

⁶⁶ Cfr. Wikipedia, Definition, [Deep learning - Wikipedia](#)

Walter Pitts e Warren McCulloch⁶⁷ ed era formato da un insieme di algoritmi e matematica chiamata logica di soglia (*threshold logic*) al fine di emulare il pensiero umano.⁶⁸

Nelle decadi successive ci furono vari tentativi come quello di Alexey Grigoryevich Ivakhnenko e Valentin Grigor'evich Lapa, un modello con funzione di attivazione polinomiale nel 1965; quello di Kunihiko Fukushima, un modello con reti neurali convoluzionali, nel 1979 e quello di Dana Cortes e Vladimir Vapnik, un modello per mappare e riconoscere dati simili nel 1997. Sfortunatamente nessun modello di apprendimento profondo fu considerato un successo in quanto erano troppo complessi, la tecnologia di allora non riusciva a supportarli e la maggior parte dei progetti furono accantonati.⁶⁹

Nel 1999, quando i computers iniziarono ad elaborare i dati tramite il processore grafico (GPU), i sistemi di rete neurali furono ripresi in considerazione in quanto la velocità di calcolo era aumentata di mille volte.

Col passare degli anni e l'aumento della potenza di calcolo delle GPU, si svilupparono algoritmi dell'apprendimento profondo ulteriormente complessi con la capacità di avere una memoria di tipo *long short-term memory* - una architettura in grado di classificare ed elaborare più dati simultaneamente e di fornire previsioni sul risultato in base agli eventi precedenti.⁷⁰

⁶⁷ Walter Pitts e Warren McCulloch sono due figure importanti nel campo dell'apprendimento neurale perché scrissero l'articolo "A Logical Calculus of Ideas Immanent in Nervous Activity" (1943), che espongono il primo modello matematico di una rete neurale nella storia.

⁶⁸ Cfr, Keith D. Foote, A Brief History of Deep Learning, [A Brief History of Deep Learning - DATAVERSITY](#) 07/02/2017

⁶⁹ Ivi.

⁷⁰ Cfr. MLK, Brief History of Deep Learning from 1943-2019 [Timeline], [Brief History of Deep Learning from 1943-2019 \[Timeline\] | MLK - Machine Learning Knowledge](#), 24/11/2019

Il 2012 è stato segnato dal successo dell'esperimento svolto da Google Brain: sviluppare un'architettura di apprendimento profondo che imparasse senza la tecnica, sempre usata, dell'apprendimento supervisionato (*supervised learning*), cioè ricevere dati senza essere etichettati e forzare il calcolatore a distinguere i dati autonomamente.⁷¹

Infine nel 2019, il contributo di Yoshua Bengio, Geoffrey Hinton, and Yann LeCun portò l'apprendimento profondo ad un livello successivo, apportarono migliorie nella visione artificiale (*computer vision*) e nel riconoscimento vocale, i quali aprirono la strada a tecnologie per la diagnosi medica automatizzata e automobili con la guida automatica.⁷²

⁷¹ Cfr. Wikipedia, Google Brain, [Google Brain - Wikipedia](#)

⁷² Cfr. Varun Bansal, [The Evolution of Deep Learning | Towards Data Science](#), 05/04/2021

3.1.1 Come funziona l'apprendimento profondo (*deep learning*)

L'apprendimento profondo tramite complessi algoritmi mira a fornire risultati sempre più accurati tramite l'auto miglioramento ed apprendimento di una macchina, computer o architettura. Le sue reti neurali complesse sono state progettate con lo scopo di emulare il pensiero umano ed imparare dagli eventi passati.⁷³ L'implementazione del processo di analisi dei *Big Data* (*Big data analytics*) ha migliorato l'elaborazione dei dati delle reti neurali, specialmente le più complesse⁷⁴ come GPT-3, Facebook e Google, e offre la possibilità di eseguire un'analisi predittiva su gli eventi futuri in base all'analisi degli eventi passati. Per capire al meglio come funziona l'apprendimento profondo, è opportuno specificare che l'analisi dei *Big Data* serve per estrapolare da una enorme quantità di dati⁷⁵ importanti informazioni per l'azienda come tendenze e condizioni di mercato, comportamento e preferenze degli utenti e clienti, per migliorare il processo decisionale e valutare la qualità dei prodotti.⁷⁶

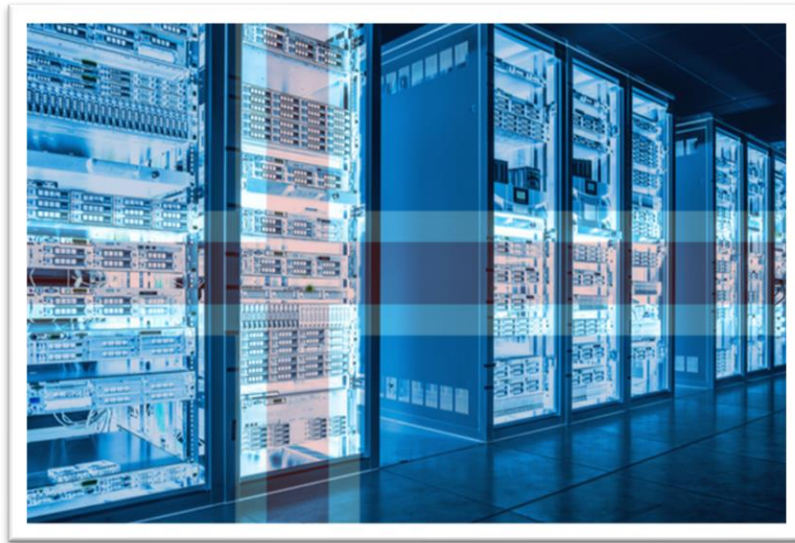
L'apprendimento profondo è una architettura cruciale per lo sviluppo di traduttori, sistemi per il riconoscimento vocale e per la interazione tra uomo e macchina, inoltre può risolvere molte problematiche con un controllo minimo od assente degli esseri umani.

⁷³ Cfr. MathWorks, What is deep learning?, [What Is Deep Learning? | How It Works, Techniques & Applications - MATLAB & Simulink \(mathworks.com\)](#)

⁷⁴ Cfr. Jeremy Hsu, Biggest Neural Network Ever Pushes AI Deep Learning, [Biggest Neural Network Ever Pushes AI Deep Learning - IEEE Spectrum](#)

⁷⁵ I dati possono avere vari formati: strutturato (foglio di Excel), semi-strutturato (e-mail) e non strutturato (immagini e foto).

⁷⁶ Simplilearn, What is Big Data Analytics and Why It is Important?, [Big Data Analytics: Types, Tools and Applications \[Updated\] \(simplilearn.com\)](#), 7/07/2021.

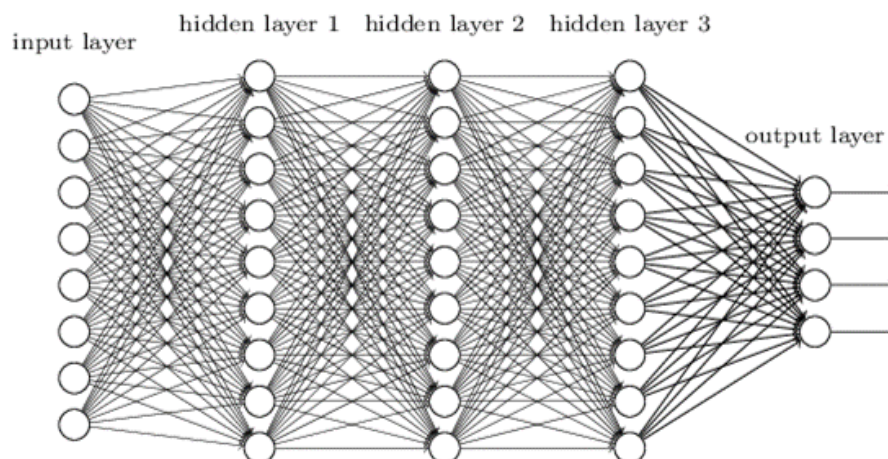


Data centre per l'apprendimento profondo di DeepL

Fonte: <https://slator.com/linguees-founder-launches-deepl-attempt-challenge-google-translate/>

Le reti neurali sono formate da 3 strati principali, *input layer*, uno o più *hidden layers*, *output layer*, ciascuno strato svolge compiti ben specifici - trascrivere testi, analizzare suoni e immagini – al fine di amministrare l'apprendimento profondo.⁷⁷

Deep neural network



L'immagine mostra l'architettura di una rete neurale profonda. Gli strati nascosti (hidden layers) possono essere più di uno in base alle necessità.

Fonte: [Interpretability of Neural Networks | by Stephen Ho | Medium](#)

⁷⁷ Cfr. Ahmed Gad, Beginners Ask “How Many Hidden Layers/Neurons to Use in Artificial Neural Networks?”, [Beginners Ask “How Many Hidden Layers/Neurons to Use in Artificial Neural Networks?” | by Ahmed Gad | Towards Data Science](#), 27/07/2018

Come il cervello umano, la rete neurale artificiale è composta da migliaia di strati contenenti milioni di neuroni che comunicano tra di loro scambiandosi migliaia di segnali. Ad ogni nodo viene assegnato un peso in base al dato che sta analizzando, il nodo più pesante di ciascuno strato influenzerà l'analisi dello strato successivo e l'ultimo strato analizzerà tutti i dati dei nodi, provenienti dagli strati precedenti, col peso maggiore per fornire il risultato finale. Questa struttura complessa di strati e nodi ha bisogno di un'architettura hardware potente in quanto i calcoli dell'apprendimento profondo possono necessitare giorni o settimane.⁷⁸

La quantità di dati necessaria, affinché un sistema di apprendimento profondo fornisca un risultato accurato, è enorme. Si pensi che per un programma di riconoscimento facciale necessita di milioni di immagini etichettate, ad esempio ImageNet⁷⁹ è un database di 14 milioni di immagini etichettate di volti che serve per allenare i programmi di riconoscimento facciale; un traduttore necessita di milioni di parole etichettate singolarmente ed un altrettanto numero di frasi etichettate affinché possa restituire una frase tradotta che abbia senso.⁸⁰

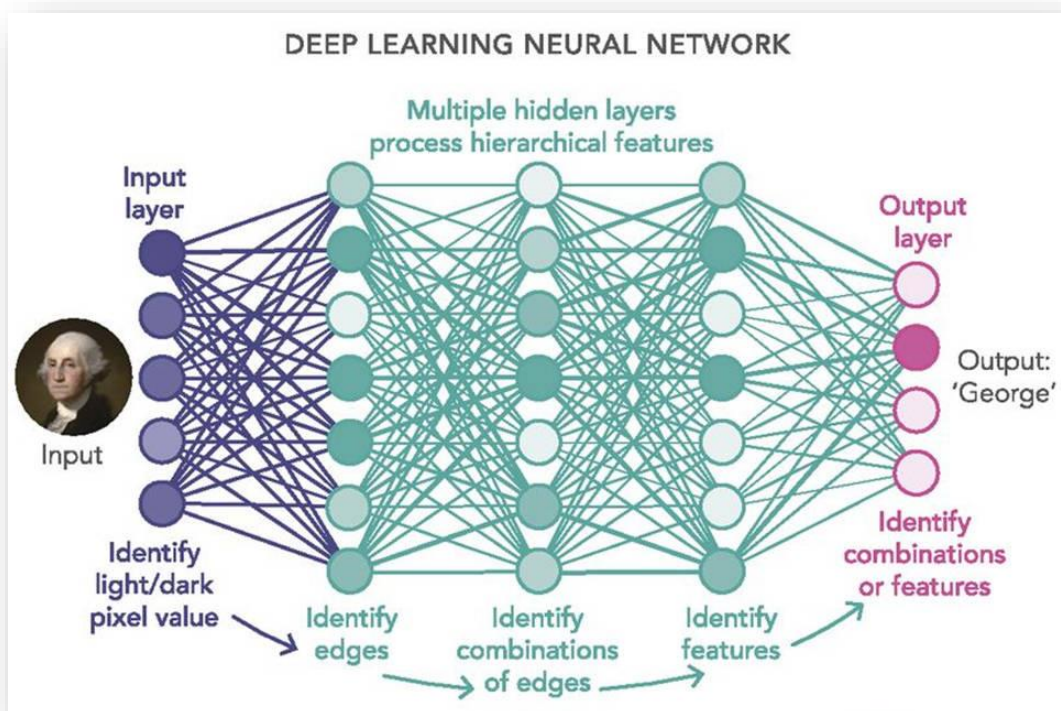
Durante la lavorazione dei dati, le reti neurali catalogano le informazioni grazie alle risposte di un insieme di domande binarie vero o falso che sono il risultato di calcoli matematici complessi. Inoltre, man mano che il programma svolge le operazioni richieste, esegue il training autogeno per aumentare l'accuratezza delle risposte.

⁷⁸ Cfr. David Fumo, Gentle introduction to neural network, [A Gentle Introduction To Neural Networks Series — Part 1 | by David Fumo | Towards Data Science](#), 04/08/2021

⁷⁹ ImageNet è un banca dati di centinaia di migliaia di immagini organizzata con la gerarchia WordNet. Imagenet ha supportato l'avanzamento della computer vision e della ricerca sull'apprendimento profondo.

⁸⁰ Cfr. Sambit Mahapatra, Why Deep Learning over Traditional Machine Learning?, [Why Deep Learning over Traditional Machine Learning? | by Sambit Mahapatra | Towards Data Science](#), 07/07/2021

Ad esempio se vogliamo trovare il nome di un personaggio storico tramite la sua foto, dopo averla inserita nel programma, la rete neurale la confronterà con altre foto di un database o su internet per trovare delle similitudini nei tratti del volto. Nello specifico, i nodi analizzano e confrontano la nostra foto con altre che ricevono un peso, che in base al suo valore può influenzare gli strati dei nodi successivi, che devono eseguire una ulteriore scrematura delle foto per decidere qual è quella che assomigli di più alla nostra e scoprire il nome del personaggio. Se il nome non è quello giusto, la rete neuronale inizierà da capo a confrontare le foto, altrimenti assegnerà il nome trovato all'immagine che abbiamo inserito.⁸¹



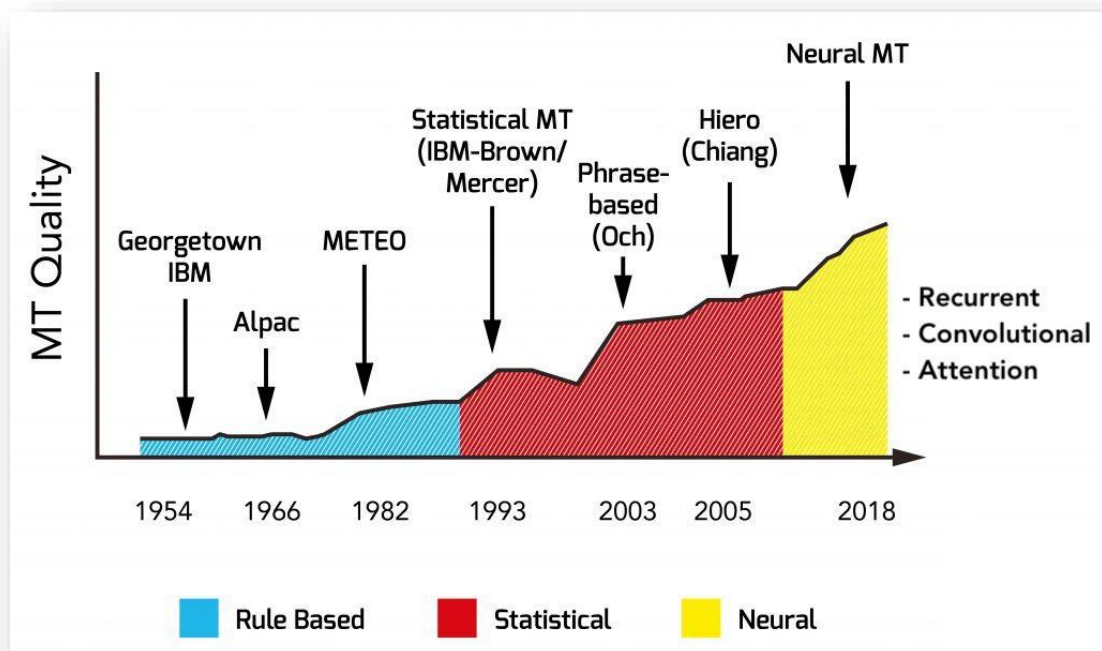
Come una rete neuronale di apprendimento profondo trova il nome della foto confrontandola con altre.

Fonte: [Neural Networks and Deep Learning. Chapter 1: Prelude \(futurespace.es\)](https://www.futurespace.es/en/neural-networks-and-deep-learning-chapter-1-prelude)

⁸¹ Cfr. Francisco Alonso, Neural Networks and Deep Learning. Chapter 1: Prelude, [Neural Networks and Deep Learning. Chapter 1: Prelude \(futurespace.es\)](https://www.futurespace.es/en/neural-networks-and-deep-learning-chapter-1-prelude)

3.2 Neural machine translation

Il *Neuronal Machine Translator* è un traduttore che si avvale di una rete neuronale artificiale per tradurre una frase di varia lunghezza. Questo tipo di traduttore negli ultimi anni ha eclissato tutti gli altri tipi – traduttore basato su regole, basato su frasi, statistico ed altri - per la sua accuratezza nella traduzione e per la migliore gestione della memoria e tempo di esecuzione.⁸²



Il grafico mostra l'efficienza del neuronal machine translator nell'eseguire traduzioni automatiche e quando è entrato in uso. Fonte: [What is Neural Machine Translation & How does it work? \(translatefx.com\)](http://translatefx.com)

Le caratteristiche che deve avere un *neuronal machine translator* sono: mantenere i dati nelle fasi temporali della elaborazione e gestire

⁸² Cfr. [What is Neural Machine Translation & How does it work? \(translatefx.com\)](http://translatefx.com)

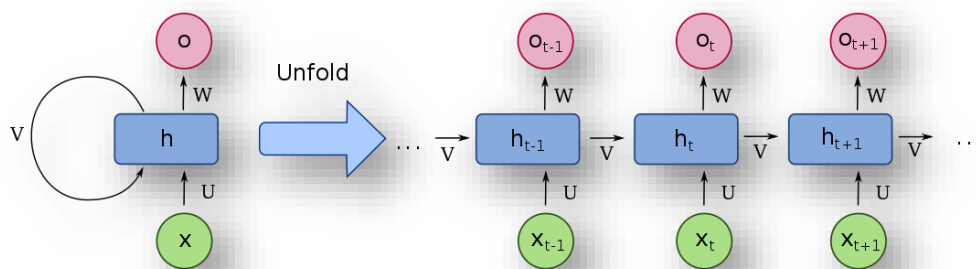
frasi di varia lunghezza. Questo tipo di traduttore utilizza dati sequenziali che devono essere salvati durante le varie fasi temporali della traduzione, quindi è necessario l'ausilio di due componenti con una architettura *Recurrent Neural Network* (RNN) di tipo *Long Short Term Memory* (LSTM) che sono strutturati in modo tale da salvare i dati per tutta la durata del processo di traduzione. Questo tipo di memoria è in grado di decidere quali informazioni salvare e quali no. Al contrario una rete neurale artificiale semplice non è in grado di salvare nessun dato sequenziale.⁸³

È necessario aprire una parentesi per capire bene che cosa significa *Recurrent Neural Network* (RNN) e *Long Short Term Memory* (LSTM).

La prima è formata da una connessione di nodi che creano un grafico orientato a svolgere un compito, diviso in passaggi, in un lasso di tempo e grazie a ciò questo tipo di rete neurale ha il tempo di correggere i propri errori durante il processo. Grazie al suo stato interno (in gergo, memoria) può trattare stringhe di dati in entrata di lunghezza variabile, nel caso di una traduzione la lunghezza non deve eccedere quella di un paragrafo di 10 righe, e questo è il motivo per cui la *Recurrent Neural Network* è usata per il riconoscimento della scrittura, traduzione ed il riconoscimento vocale.⁸⁴

⁸³ Cfr. [Intuitive explanation of Neural Machine Translation | by Renu Khandelwal | Towards Data Science](#)

⁸⁴ Cfr. [Recurrent neural network - Wikipedia](#)



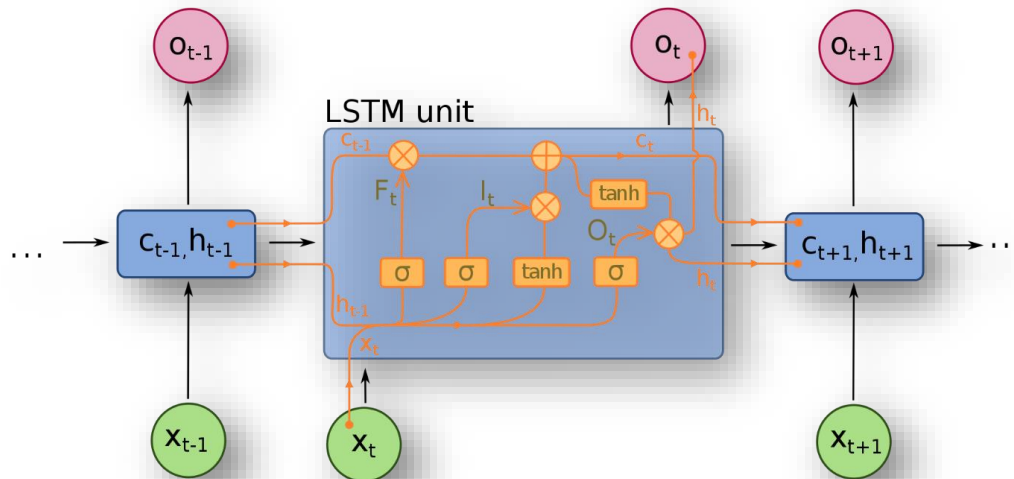
Il grafico mostra il funzionamento di una Recurrent Neural Network.

Fonte: [Intuitive explanation of Neural Machine Translation | by Renu Khandelwal | Towards Data Science](#)

Tuttavia, questo tipo di rete neurale non è in grado di trattenere dati di una certa grandezza, ad esempio una pagina di un documento, per tutto il tempo della traduzione del testo e questo porta alla perdita di dati importanti. Per questo è stata creata la Long Short Term Memory (LSTM) che evita il problema della scomparsa del gradiente (*vanishing gradient problem*), cioè porta tutti i dati dagli stadi iniziali dell'analisi fino a quelli finali e in questo modo se ci sono errori di traduzione, non saranno persi ma possono essere risolti.

La *Long Short Term Memory* riesce a fare tutto ciò tramite delle porte (gates), che gli permettono di regolare il flusso dei dati. Con queste porte, questo tipo di rete neuronale può decidere quali informazioni tenere e farle procedere nella serie successiva dell'analisi. Inoltre ogni porta assegna un peso al dato che sta analizzando e nel passo successivo la prossima porta terrà conto di questo per decidere se salvarlo o no, in questo modo il risultato finale dipenderà dal complesso di tutti i passaggi di analisi, dai pesi assegnati dei dati salvati man mano che la LSTM svolge il suo lavoro. Questa caratteristica è fondamentale per un *neural machine translator* che tratta dati sequenziali.

Ogni volta che si verifica un errore, viene rinviato alla fase iniziale per essere analizzato e corretto, questo vuol dire che la LSMT è capace di imparare dagli errori commessi.⁸⁵



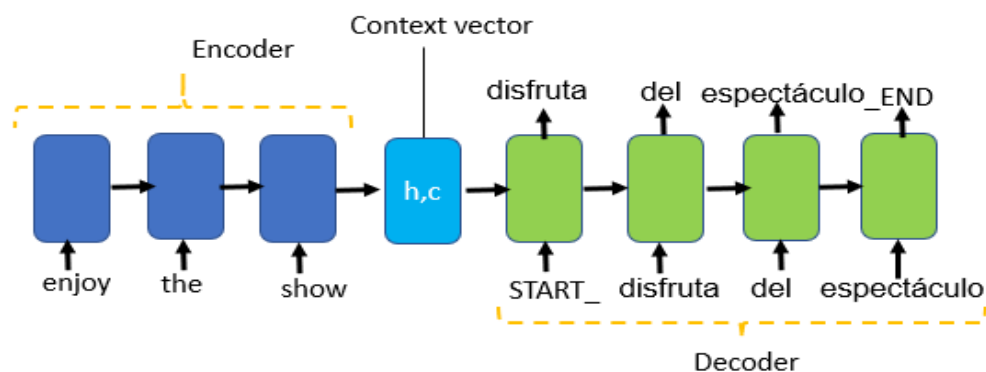
Il diagramma mostra la funzionalità della *Long Short Term Memory*. Fonte: [Long short-term memory - Wikipedia](#)

Ora è più chiaro comprendere come fa un neural machine translator a gestire anche pagine o documenti interi, senza perdere dati o creare un risultato con degli errori, grazie all'uso della LSMT.

Il neural machine translator possiede una architettura diversa dagli altri traduttori che elaborano il linguaggio naturale, in quanto non ha un componente che possiede le regole grammaticali di entrambe le lingue, uno che analizza la frase di entrata, uno che traduce parola per parola, uno che compone la frase ed uno che esegue il controllo finale; ma ha un modello unico che analizza le parole, le traduce, crea la frase e poi rilascia il risultato e nel frattempo è in grado di imparare dai propri errori.

⁸⁵ Cfr. Michael Phi, [Illustrated Guide to LSTM's and GRU's: A step by step explanation | by Michael Phi | Towards Data Science](#)

Nel dettaglio la procedura che il *neural machine translator* utilizza, è composta dal modello Seq2Seq⁸⁶ che trasforma la sequenza in entrata (*source sequence*) nella sequenza di destinazione (*target sequence*) tramite due LSMT chiamati codificatore (*encoder*) e decodificatore (*decoder*).⁸⁷ Il codificatore decifra le parole in ingresso della lingua da tradurre, una per una, e le inserisce in un vettore di valore chiamato vettore di pensiero (*thought vector*). Quest'ultimo analizza le parole, ne trae il significato e lo inserisce accanto alle parole analizzate. Dopo, il codificatore restituirà dei risultati che in parte saranno scartati al fine di salvare solamente quelli che sono ritenuti giusti che saranno inseriti nel decodificatore. Mentre il decodificatore crea un token chiamato Start per indicare che da lì inizia la stringa in uscita, riceve i risultati non scartati dal codificatore e li analizzerà nuovamente. Assegnerà a ciascuna stringa un peso e quella col peso maggiore sarà il risultato finale dell'analisi: la traduzione.⁸⁸



La figura mostra come un *neural machine translator* traduce la frase in entrata "Enjoy the show"

Fonte: [Implementing neural machine translation using keras | by Renu Khandelwal | Towards Data Science](#)

⁸⁶ Seq2Seq è un modello speciale del tipo Recurrent Neural Network, specializzato nel risolvere problemi di linguaggio complessi come le traduzioni, rispondere alle domande e creare una sessione di scrittura.

⁸⁷ Cfr. [Intuitive explanation of Neural Machine Translation | by Renu Khandelwal | Towards Data Science](#)

⁸⁸ Ibid. [Intuitive explanation of Neural Machine Translation | by Renu Khandelwal | Towards Data Science](#)

4. Esempi su come la elaborazione del linguaggio naturale funzioni nell'ambito della comunicazione vocale tra uomo e macchina

Per gli esseri umani comunicare tramite la voce è una forma di comunicazione naturale e spontanea, e ora anche i computer sono progettati per capire il linguaggio umano ed a essere capaci di comunicare. Il primo software che riconosce la voce è stato ideato negli anni '50 del secolo passato, e da quel momento c'è stato un susseguire di invenzioni e tentativi che ci hanno dato molti dispositivi per facilitare la vita di tutti i giorni. Ad esempio gli assistenti vocali stanno scalando la lista dei dispositivi elettronici di uso domestico più acquistati – Amazon Alexa e Google Home – e stanno modificando il modo in cui eseguiamo azioni quotidiane come cercare informazioni, comprare su internet e comunicare tra di noi.

Vi è un grande dibattito tra i ricercatori su quando gli esseri umani hanno iniziato a conversare tra di loro: chi pensa che abbiamo iniziato a parlare 50.000anni fa e chi pensa che lo abbiamo sempre fatto fin dalla prima apparizione dell'uomo cioè 2 milioni di anni fa.⁸⁹ Di media una persona usa 7 mila parole ogni giorno, di cui tra le 100 e le 130 sono pronunciate ogni minuto.⁹⁰

Tuttavia, comunicare va oltre le parole, è complesso e difficile, perché le nostre parole sono il risultato della formulazione del nostro pensiero su come percepiamo il mondo, in che contesto e luogo viviamo; il motivo che ci spinge a parlare; con chi parliamo e chi ascoltiamo.

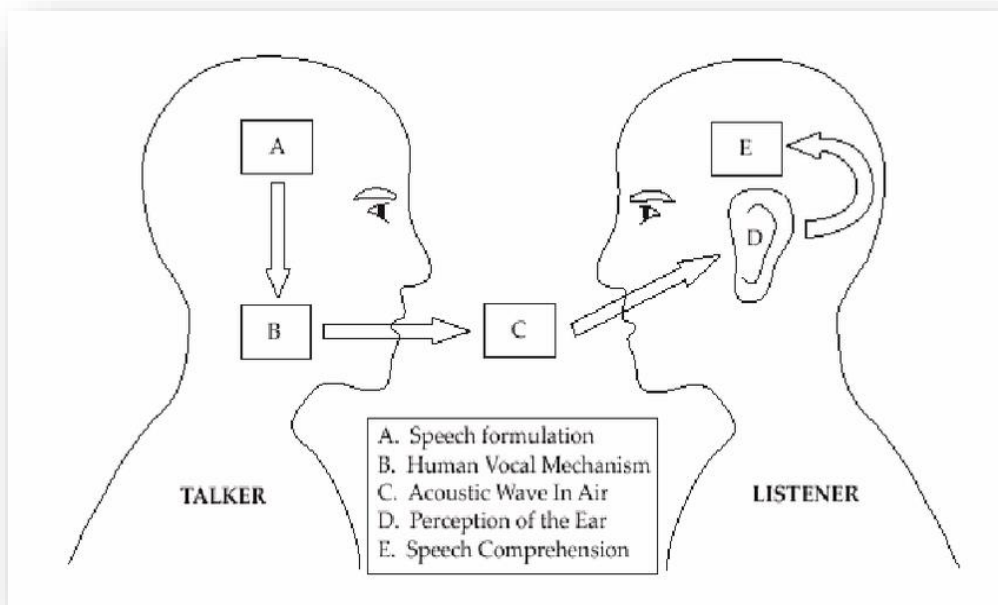
⁸⁹ Cfr. Michael Balter, [Human language may have evolved to help our ancestors make tools | Science | AAAS \(sciencemag.org\)](https://www.sciencemag.org) p. 1.

⁹⁰ Cfr. [How Fast Does the Average Person Speak? - Word Counter Blog](#)

La comunicazione si divide in 6 fasi:

1. Il pensiero si tramuta in parole;
2. Le corde vocali generano dei suoni appartenenti ad un linguaggio naturale;
3. L'aria trasmette questi suoni e vibrazioni all'orecchio di chi ascolta;
4. I nervi uditivi trasmettono il suono al cervello;
5. Il cervello converte il suono in un linguaggio naturale;
6. Il cervello processa il suono per avere il significato delle parole.

Questi passaggi si possono riassumere in: sintesi vocale, riconoscimento vocale e comprensione vocale. ⁹¹



La foto mostra il processo completo della comunicazione tra due interlocutori.

Fonte: [1. Schematic Diagram of the Speech Production/Perception Process \[10\]. | Download Scientific Diagram \(researchgate.net\)](#)

⁹¹ Cfr. S.Amin [Speech recognition is hard — Part 1 | by SAmin | Towards Data Science](#)

Una macchina per eseguire tutto ciò deve essere capace di captare il suono della voce umana e capire le parole. Questa è la missione del campo interdisciplinare del sistema automatico di riconoscimento del discorso (*Automatic Speech Recognition, ASR*).

4.1 Sistema automatico di riconoscimento del discorso (*Automatic Speech Recognition, ASR*)

Nel 1952 la Bell Labs crea il primo sistema di riconoscimento vocale, che chiama Audrey; può riconoscere i numeri da 0 a 10. Nel 1962, IBM immette nel mercato la macchina “*Shoebox*” ad attivazione vocale, è capace di riconoscere 16 parole ed è usato nei centralini per indicare agli utenti quale opzione volessero scegliere tramite la tastiera del telefono.⁹²

Alla fine degli anni 60 del secolo scorso, Raj Reddy, studente della Stanford, costruisce una macchina capace di giocare a scacchi tramite la ricezione di comandi vocali. Nel 1980, IBM crea Tangora: una macchina da scrivere dotata di attivazione vocale. Ha un vocabolario di 20 mila parole e aiutava le persone con problemi di vista.⁹³

Nel 1984, ACT Ltd immette nel mercato il primo computer portatile con un sistema di riconoscimento vocale incorporato, con un vocabolario di 4096 parole, potendone salvare nella RAM 64 per volta. Si possono attivare i vari comandi della macchina tramite la voce dell’utente che associ a ciascuna azione del computer una parola, una volta registrata esso la ripete per conferma.

Verso gli inizi degli anni 2000, IBM crea Home Page Reader che può leggere intere pagine internet in modo comprensibile per persone con problemi di vista.

Nel 2011, Apple stupisce con l’assistente vocale Siri. L’utente, dopo aver registrato la sua voce, può accedere a molte funzionalità del cellulare Iphone 4. Ad esempio può chiamare una persona della rubrica,

⁹² Cfr. IBM Cloud Education, What is Speech Recognition?, [What is Speech Recognition? | IBM](#)

⁹³ Ivi IBM Cloud Education, What is Speech Recognition? [What is Speech Recognition? | IBM](#)

scrivere e leggere i messaggi, aggiornare il calendario, eseguire ricerche su internet ed effettuare pagamenti tramite Apple Pay. Nel 2017, Google realizza un aggiornamento per Google Traduttore che lo rende il primo nel campo della traduzione ad essere capace di registrare la voce dell'utente, in più di 100 lingue, e tradurla.⁹⁴

⁹⁴ Cfr. [Speech recognition - Wikipedia](#)

4.2 Problematiche e funzionamento del sistema automatico di riconoscimento del discorso

Dai capitoli precedenti è facile comprendere come la comunicazione verbale tra due persone sia molto complessa. La ragione del perché ci siano volute decenni per sviluppare un computer in grado di poter comprendere la voce umana e rispondere in modo congruente per svolgere piccoli compiti si trova nella serie di sfide, dopo riportate, che la macchina deve affrontare per captare la voce, comprenderla e formulare una risposta.

- Gli esseri umani sono dotati di due orecchie che captano suoni, mentre un computer dovrebbe avere un microfono per le onde acustiche in entrata e un dispositivo per convertirle in segnali digitali.
- Quando si parla in un luogo pubblico, dobbiamo separare le parole dai suoni circostanti (rumori di fondo come una automobile, chiacchiericcio). La stessa cosa deve fare il computer.
- A volte si parla molto rapidamente e ciò comporta una difficile comprensione della punteggiatura, il discorso sembra un fiume di parole dove non si capisce bene l'inizio e la fine di ogni parola. Una macchina dovrebbe capire quando inizia e finisce ogni parola e frase. Tutt'ora quando si detta una frase su Google Traduttore bisogna pronunciare la punteggiatura, ad esempio “virgola”, “punto”, affinché capisca la struttura della frase.
- Ogni persona parla in modo diverso, in base allo stato sociale, genere, età e provenienza, inoltre ci sono modi di dire e giochi di parole. Il computer dovrebbe essere allenato a gestire la variabilità

del discorso, possedere un vocabolario vasto e un database di suoni registrati per riconoscere i vari accenti ed inflessioni.

- Nelle varie lingue esistono gli omofoni e la macchina dovrebbe essere capace di capire in base al contesto, queste specifiche parole, che cosa significano e che funzione hanno. Ad esempio “anno” è un sostantivo e “hanno” è la terza persona plurale dell'indicativo presente del verbo avere; “o” come congiunzione disgiuntiva e “ho” è la prima persona singolare dell'indicativo presente del verbo avere.
- Ci sono i suoni di riempimento del discorso come “mm”, “ehh” e “aah”, che la macchina dovrebbe capire che non sono parole vere e proprie.
- Le incomprensioni, ad esempio il nome di un posto o di una via, che tutt'ora è un errore ricorrente in tutti i sistemi automatici di riconoscimento del discorso.
- Infine, dopo aver superato le barriere linguistiche sopracitate, la macchina deve completamente conoscere la sintassi, la semantica della lingua da tradurre e quella di arrivo, e il contesto.

Per far fronte a tutte queste problematiche il sistema automatico di riconoscimento del discorso è formato da una architettura complessa:⁹⁵

- Il ricettore dei suoni ha il compito di trasformare le onde sonore in segnali digitali.

⁹⁵ Cfr. Ketan Doshi, Audio Deep Learning Made Simple: Automatic Speech Recognition (ASR), How it Works, [Audio Deep Learning Made Simple: Automatic Speech Recognition \(ASR\), How it Works | by Ketan Doshi | Towards Data Science](#) , 25/03/2021

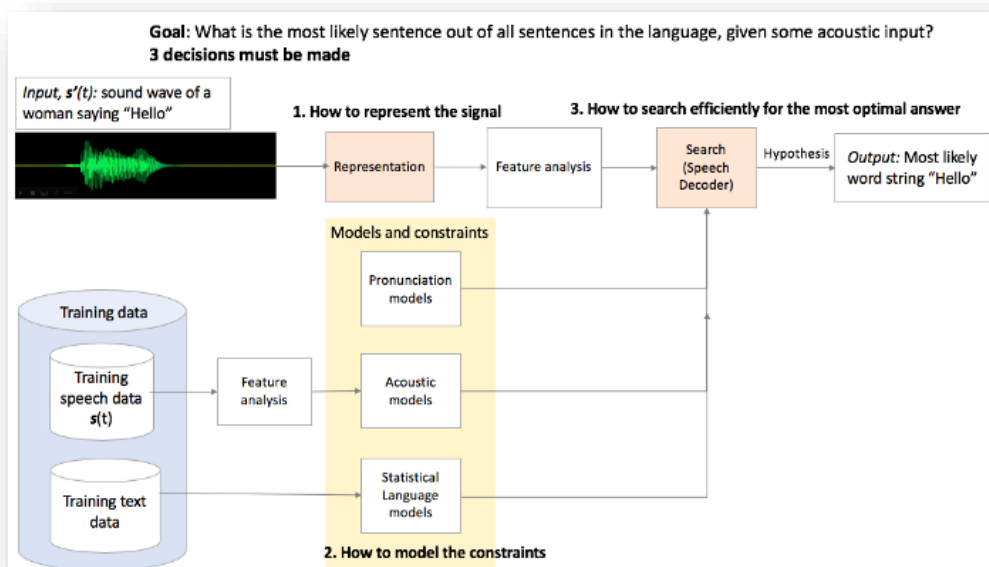
- Il selettore della voce scansiona il segnale digitale ed isola la parte che contiene il discorso al fine di trasformarla in un parametro acustico.
- La banca dati composta da corpus di documenti scritti e voci registrate. Inoltre si possono registrare interi documenti, così la macchina può imparare a gestire un flusso di parole una dopo l'altra.
- L'apparato acustico scompone l'onda acustica del discorso e la divide in pezzi e calcola la probabilità di quali fonemi siano più probabili.⁹⁶
- Il modulo di pronuncia associa le varie parole del discorso ai loro rispettivi suoni.⁹⁷
- Il modulo linguistico predice la sequenza più probabile che le parole debbano formare per creare una frase di senso compiuto. Per far ciò confronta le parole ed i suoni della frase in entrata con quelle del suo data base.⁹⁸
- L'algoritmo, chiamato anche decodificatore, unisce il risultato fonetico e quello linguistico per produrre una stringa di testo più probabile in base alla frase in entrata.⁹⁹

⁹⁶ Cfr. ITChronicles, [Speech Recognition Algorithm - Brought to you by ITChronicles](#)

⁹⁷ Cfr. Samudravijaya K, ASR, <http://www.iitg.ac.in/samudravijaya/tutorials/asrTutorial.pdf>, , pp. 9-10

⁹⁸ Ibidem Samudravijaya K, ASR, <http://www.iitg.ac.in/samudravijaya/tutorials/asrTutorial.pdf>, p.12

⁹⁹ Cfr. S.Amin, [Speech recognition is hard — Part 1 | by SAmin | Towards Data Science](#)



Questo diagramma mostra il funzionamento del sistema automatico di riconoscimento del discorso.

Fonte: [Speech recognition is hard — Part 1 | by SAmin | Towards Data Science](#)

La qualità di un sistema di riconoscimento del discorso si valuta tramite la formula del tasso di errore della parola (*word error rata*, WER) e del tasso di accuratezza della parola (*word accuracy*, WAcc).

$$WER = \frac{S + I + D}{N}$$

Formula del tasso di errore della parola, fonte [Word error rate - Wikipedia](#)

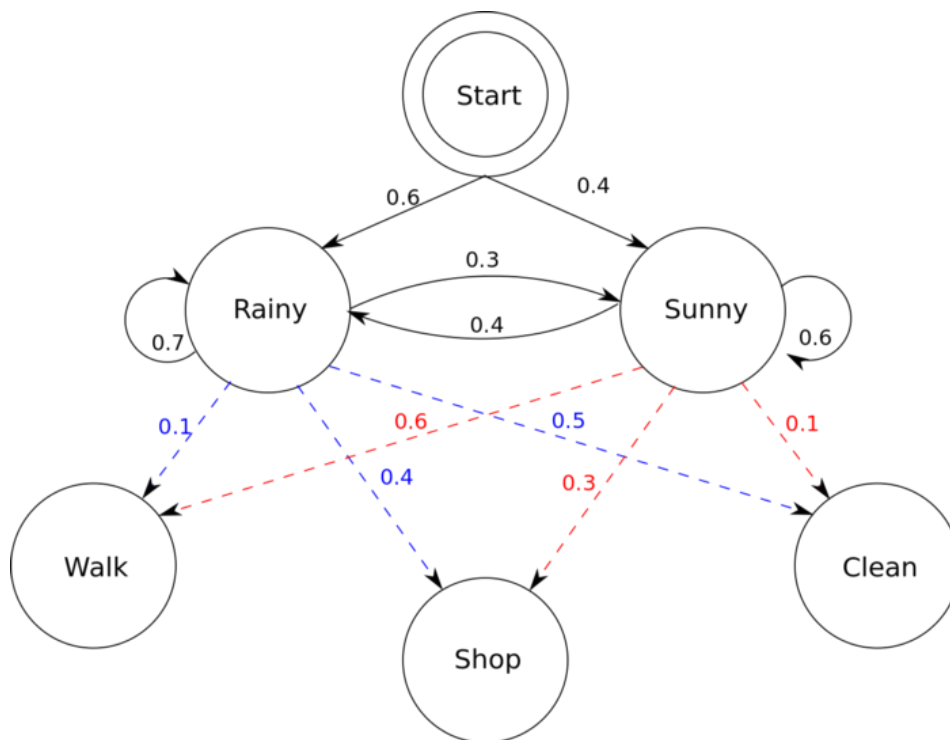
Dove S è il numero di volte che le parole sono state sostituite, D è il numero di volte le parole sono state cancellate, I è il numero di volte che una nuova parola è stata inserita e N è il numero di parole da prendere in considerazione.¹⁰⁰

Mentre tasso di accuratezza della parola si calcola:

$$WAcc = 1 - WER$$

¹⁰⁰ Cfr. [Word error rate - Wikipedia](#)

Vari fattori possono incidere sul tasso di errore della parola ed il tasso di accuratezza della parola, come i rumori di sottofondo, un accento ed una parola sconosciuta alla macchina. Vari algoritmi (decodificatori) si implementano per avere una trascrizione ed un output accurati. L'algoritmo più usato è il *Hidden Markov Models* che è un modello statistico capace di calcolare la probabilità di un insieme di variabili future analizzando quelle presenti. Secondo Markov, non è necessario conoscere tutte le variabili storiche del processo ed una volta stabilita la probabilità di quelle future, bisogna determinare il migliore insieme possibile fra di esse, che sarà quello che si avvicina di più alla sequenza di variabili presenti.¹⁰¹



L'immagine mostra come funziona il Hidden Markov Model per predire la scelta migliore da fare in base alle condizioni metereologiche presenti.

Fonte: [Hidden Markov model - Wikipedia](#)

¹⁰¹ Cfr. Sanjay Dorairaj, Hidden Markov Models Simplified, [Hidden Markov Models Simplified. Sanjay Dorairaj | by Sanjay Dorairaj | Medium](#)

Altri due algoritmi più usati sono l'elaborazione del linguaggio naturale e la rete neuronale artificiale trattati nei capitoli precedenti.

Il sistema automatico di riconoscimento del discorso è usato in vari ambiti come:

- Nel settore automobilistico: con la voce l'utente può accendere la radio e cambiare stazione, accendere o spegnere i fari e la macchina può controllare se il guidatore si addormenti e lo proverà a svegliare con un messaggio vocale;
- Nell'assistenza sanitaria: il dottore può dettare un rapporto con l'ausilio di un assistente digitale anziché scriverlo a mano e durante una operazione chirurgica può usare dei comandi vocali per navigare ed utilizzare il computer;
- Nell'aviazione: il pilota può attivare il pilota automatico, l'uso dei flaps e delle ruote tramite comandi vocali.
- Nella telefonia: gli assistenti vocali possono aiutare l'utente nelle ricerche, comporre numeri telefonici, controllare il tempo ed eseguire pagamenti.

4.3 Assistente digitale: unione tra l'intelligenza artificiale e la linguistica

Amazon è una delle compagnie che negli ultimi anni ha allargato al sua offerta con dei servizi tecnologici all'avanguardia, per rendere l'esperienza dell'utente piacevole mentre naviga nel suo sito e per facilitare la vita di tutti i giorni nelle mura domestiche. Nel 2013, Amazon lancia sul mercato Alexa, un assistente virtuale basato su intelligenza artificiale integrato in un altoparlante multifunzione chiamato Amazon Echo.



Amazon Echo con l'agente virtuale Alexa di terza generazione.

Fonte: [AMAZON ECHO STUDIO | Mediaworld.it](https://www.mediaworld.it/it/amazon-echo-studio)

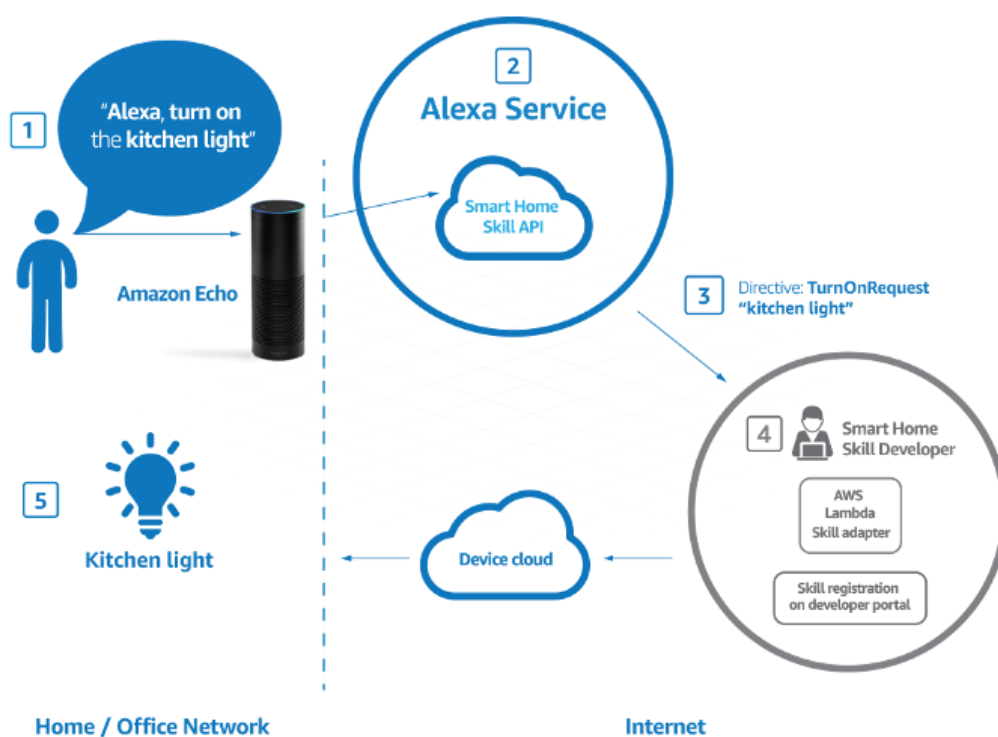
Fin da subito Alexa ha scalato le classifiche di vendita degli assistenti digitali per uso domestico nel mondo. Nel 2017, Alexa possedeva il 62% della quota delle vendite globale degli assistenti digitali e nel 2018, 40 milioni di statunitensi l'hanno comprata. Nel 2019 dopo un investimento di milioni di dollari, Alexa è capace di svolgere 100 mila attività diverse e questo porta ad un aumento della quota delle vendite a 67%, facendo divenire Amazon il principale produttore di assistenti artificiali nel mondo, inoltre ha venduto in tutto il mondo più di 10 milioni di dispositivi. Nel 2020, 53,6 milioni di Alexa sono stati venduti e si prevede che il numero salirà a 65 milioni entro la fine del 2021 e nel 2020, Alexa ha ricevuto 250 mila proposte di matrimonio.¹⁰² Nel 2021, il 70% degli assistenti digitali presenti nelle case degli statunitensi è marchiato Amazon.¹⁰³

Amazon ha usato le tecnologie più all'avanguardia per creare questo assistente virtuale: l'elaborazione del linguaggio naturale combinato con il sistema automatico del riconoscimento del discorso. L'idea della interazione tra l'utente ed Alexa è che dopo aver captato il comando vocale quest'ultima lo invia ai server di Amazon. L'assistente vocale necessita della connessione internet per funzionare, infatti, l'architettura del dispositivo è formata semplicemente da un altoparlante, dai 4 ai 7 microfoni, pulsanti per il volume e l'accensione e la scheda WI-FI. Una volta ricevuto il segnale, il server Alexa Voice Service (AVS) suddivide il comando vocale in piccoli suoni che saranno confrontati con quelli presenti nel database, che contiene esempi di

¹⁰² Cfr. Safe at Last, Incredible Amazon Alexa Statistics You Need to Know, 2021 <https://safeatlast.co/blog/amazon-alexa-statistics/#gref>

¹⁰³ Cfr. Smiljanic Stasha, Amazon Alexa Statistics, Facts, and Trends <https://policyadvice.net/insurance/insights/amazon-alexa-statistics/>

comandi vocali in varie lingue ed accenti.¹⁰⁴ Nel confrontare i suoni, AVS cerca di trovare la parola chiave su cui focalizzerà la ricerca, se ad esempio trova la parola “olimpiadi” il programma accederà alle notizie ed informazioni sulle Olimpiadi di Tokyo 2020. Dopo aver trovato le informazioni, AVS invierà il risultato ad Alexa che lo fornirà all’utente. Nel caso il comando fosse ambiguo, Alexa chiederà conferma all’utente ripetendo dei suoni che si avvicinino al comando vocale ricevuto, questa metodica serve anche per addestrare Alexa ad apprendere nuove informazioni e riconoscere parole ed accenti diversi.



Il grafico mostra come avviene la comunicazione tra l’utente ed Alexa.

Fonte: [How Amazon Alexa works? Your guide to Natural Language Processing \(AI\) | by Alexandre Gonfalonieri | Towards Data Science](#)

¹⁰⁴ Cfr. Alexandre Gonfalonieri, How Amazon Alexa Works?, Towards Data Science, [How Amazon Alexa works? Your guide to Natural Language Processing \(AI\) | by Alexandre Gonfalonieri | Towards Data Science](#), 24/05/2021

Nello specifico, bisogna far fronte a vari ostacoli che possono influenzare la comunicazione tra Alexa e l'utente. Il primo è captare il segnale giusto da analizzare nel campo dell'audio a distanza (*far-field audio*), di conseguenza l'architettura dell'assistente digitale è dotata di 7 microfoni per isolare le varie onde sonore superflue – rumore della televisione, automobili, gente che parla, eco..... – che la circondano e cancellarle. In questo modo è facile identificare il comando vocale. Alexa è in grado di rilevare un segnale acustico fino a 10 metri di distanza. L'altro problema è la parola di attivazione (*Wake Word Detection, WWD*), che deve essere univoca per ridurre al minimo i falsi positivi o negativi al fine di non causare inconvenienti come acquisti accidentali. Ad esempio una persona negli Stati Uniti, nelle prossimità di Alexa stava valutando con sua moglie di comprare un determinato cellulare su Amazon e l'assistente digitale, avendo accesso alle sue credenziali bancarie, effettuò l'acquisto. Questo problema può essere risolto impostando la parola chiave per azionarla come ad esempio "Alexa" o "Hey Alexa".¹⁰⁵ Una volta che AVS ha ricevuto il segnale, lo convertirà in un testo e lo confronterà con i dati del data base. Questa fase si divide in due passaggi, il primo è fornire una versione scritta del comando che abbia una alta probabilità di compatibilità e il secondo passaggio è di accoppiare il testo scelto con il comando vocale. In questa fase riconoscere le parole è arduo, perché ogni lingua possiede molti omofoni e, alcune marche e personaggi famosi, tendono ad avere un nome che si pronuncia in un modo ma si scrive in un altro. Per risolvere questo problema si usa l'apprendimento profondo, così Alexa può migliorare ogni volta che esegue un'azione ed imparare dagli errori. Il

¹⁰⁵ Ibid Alexandre Gonfalonieri, How Amazon Alexa Works?, Towards Data Science, [How Amazon Alexa works? Your guide to Natural Language Processing \(AI\) | by Alexandre Gonfalonieri | Towards Data Science](#), 24/05/2021

sistema automatico di riconoscimento del discorso, usando il Hidden Markov Models, è capace di calcolare la probabilità di un insieme di variabili future analizzando quelle presenti.¹⁰⁶

Uno dei motivi per cui Alexa può non capire bene la richiesta è perché non si è seguita la procedura per formulare il comando vocale.



La foto mostra la struttura del comando vocale che gli utenti devono seguire per aumentare la probabilità che Alexa capisca.

Fonte: [How Amazon Alexa works? Your guide to Natural Language Processing \(AI\) | by Alexandre Gonfalonieri | Towards Data Science](#)

Le parti principali da seguire per creare un comando vocale comprensibile sono tre: la parola chiave per azionarla (*wake word*), attivare le funzioni (*invocation name*) e l'affermazione (*utterance*). La parola chiave per azionarla - "Alexa" o "Hey Alexa" - cambia lo stato di Alexa da stand-by ad ascolto così è pronta per ricevere il comando. L'attivazione di funzione serve per attivare una capacità specifica di Alexa che le permetterà di comprendere meglio che cosa deve svolgere. Questa parte può essere una parola, una domanda od una azione. Alexa può essere personalizzata ed imparare nuove azioni ed ognuna deve essere etichettata con un nome univoco per attivarla.¹⁰⁷ L'affermazione è la parte in cui l'utente indica ad Alexa che cosa deve fare con la funzione specifica che ha attivato precedentemente. Nell'esempio in foto si chiede

¹⁰⁶ Cfr. Magoosh, [How Does Siri Work? The Science Behind Siri - Magoosh Data Science Blog](#)

¹⁰⁷ Cfr. Nathaniel Watkins, [A Modest Proposal to Prevent False Triggers on Voice Assistants | by Nathaniel Watkins | Towards Data Science](#)

ad Alexa di attivare la funzione lettura dell'oroscopo e l'utente le chiede di focalizzarsi sulla costellazione del Toro.¹⁰⁸

Inoltre, Alexa può imparare a svolgere nuove funzioni, infatti Amazon permette a parti terze di creare app per istruirla per svolgere nuove azioni, in questo modo gli utenti potranno sperimentare sempre funzioni nuove e semplificare la vita di tutti i giorni.¹⁰⁹ Ad esempio si può prenotare un taxi, insegnare ad Alexa a fare battute, raccontare storie e barzellette, ad assistere l'utente mentre cucina e a farla diventare un sensore di movimento, così da rilevare se ci sono movimenti o rumori sospetti in casa (questa modalità si chiama "Away").

In conclusione Alexa è un'assistente digitale completa che grazie all'apprendimento profondo può imparare velocemente dagli eventi passati, per dare risultati sempre più accurati e acquisire nuove funzionalità per intrattenere ed aiutare gli utenti. Inoltre, con l'uso del sistema automatico di riconoscimento del discorso e l'elaborazione del linguaggio naturale, è in grado di interpretare i comandi vocali degli utenti con diversi accenti e pronunce.¹¹⁰

¹⁰⁸ Cfr. Kiran Krishnan, [How do Alexa Skills work?. Alexa is a voice service from Amazon... | by Kiran Krishnan | Chatbots Magazine](#)

¹⁰⁹ Cfr. Businessinsider, [How to Add Skills to Alexa in 3 Different Ways \(businessinsider.com\)](#)

¹¹⁰ Cfr. Alexandre Gonfalonieri, [How Amazon Alexa works? Your guide to Natural Language Processing \(AI\) | by Alexandre Gonfalonieri | Towards Data Science](#)

5. I vantaggi e le problematiche dell'elaborazione del linguaggio naturale

Nell'ultimo secolo, la comunicazione tra uomo e macchina è in continua evoluzione, perché deve essere flessibile per adattarsi alla società odierna e perché, uno degli obiettivi principali che le compagnie e ricercatori perseguono, è quello di progettare computer che possano capire a pieno i linguaggi naturali, le sfumature linguistiche ed il contesto della frase. Una delle problematiche ricorrenti, a cui gli sviluppatori devono far fronte, è la difficile gestione dell'enorme quantità di dati – corpora composti da un 1 milione di parole e 15 milioni di frasi etichettate - al fine di addestrare le macchine per svolgere compiti sempre più complessi. Grazie allo sviluppo dell'elaborazione del linguaggio naturale (*NLP*) si è potuto trovare una soluzione a questa sfida, infatti ora le macchine possono analizzare e organizzare documenti, comprendere le sfumature linguistiche, offrire traduzioni sempre più accurate, assistenti vocali e digitali che aiutano le persone nei diversi ambiti della vita.¹¹¹ Cercare frasi nei libri, comparare documenti o semplicemente cercare informazioni su un argomento, che possono richiedere ore se svolti manualmente, ma grazie all'elaborazione del linguaggio naturale l'inserimento dati è ottimizzato ed ognuno di essi ha una etichetta, contenente le informazioni su come trattare il dato, che facilita la classificazione e la ricerca. Si pensi a Google, è sufficiente inserire la stringa di parole nella barra di ricerca ed in meno di 5 secondi esso fornisce centinaia di risultati. Le aziende possono avvalersi della NLP per eseguire una scrematura dei curriculum vitae in base a dei criteri specifici, senza pregiudizi e ridurre i tempi di selezione, e per

¹¹¹ Cfr. [Natural language processing - Wikipedia](#)

creare dei testi attitudinali per verificare se i candidati si allineano con la missione dell'azienda.¹¹² Nel settore dell'ospitalità, l'esperienza dei clienti e la loro soddisfazione è molto importante, la NLP può essere di aiuto nel facilitare i clienti nella ricerca delle informazioni, tramite un assistente digitale ed essere indirizzati all'operatore che più fa al caso loro, in questo modo i clienti sono più soddisfatti e non devono spendere molto tempo per cercare ciò che gli necessita. La NLP, grazie al modulo “*sentiment analysis*”,¹¹³ può analizzare i comportamenti e le opinioni degli utenti al fine di aiutare l'azienda a capire se i loro prodotti sono apprezzati, se gli utenti sono soddisfatti, può creare ed esaminare sondaggi e dare una valutazione della qualità dell'assistenza clienti.

La NLP facilita il lavoro dei dipendenti riducendo le azioni ripetitive che possono rallentare il lavoro, deprimerli e annoiarli, così facendo si assicura che il livello di produttività e il morale rimangano sempre alti. Inoltre, fornire ai clienti la possibilità di ricevere risposte alle loro problematiche in tempi brevi e di cercare informazioni sul sito internet in maniera veloce ed intuitiva, aumenta il livello di soddisfazione degli utenti e ne porta di nuovi.¹¹⁴

La NLP ha spianato la strada per la creazione di uno degli assistenti digitali più all'avanguardia degli ultimi dieci anni: il GPT-3. Questa architettura, formata da 175 miliardi di parametri, è in grado di predire l'ordine delle parole quando riceve in input un messaggio vocale, può apprendere una lingua senza le nozioni di grammatica, imparare nuove

¹¹² Cfr. Diego Lopez Yse, [Your Guide to Natural Language Processing \(NLP\) | by Diego Lopez Yse | Towards Data Science](#)

¹¹³ Il sentiment analysis estrapola dalle conversazioni tra l'assistenza clienti e gli utenti e dalle pagine dell'azienda sulle piattaforme di comunicazione digitale, quali Instagram e Facebook ciò che i clienti pensano sui prodotti in vendita e dei servizi offerti dalla compagnia.

¹¹⁴ Cfr. Capacity, What are the advantages of Natural Language Processing (NLP), [What are the advantages of Natural Language Processing \(NLP\)? | Capacity](#)

funzionalità, generare documenti scritti e creare siti internet inserendo poche informazioni.¹¹⁵ Il giornale Britannico “*The Guardian*” lo ha usato per scrivere un articolo incentrato sul rapporto uomo-macchina ed il risultato è stato sorprendente.¹¹⁶

Tuttavia, la elaborazione del linguaggio naturale presenta delle difficoltà non trascurabili.

Approntare un’architettura NLP non è facile perché i dati devono essere scelti in modo da dare una conoscenza sia generale che profonda alla macchina su vari temi diversi tra loro. Pochi dati o troppo eterogenei possono ridurre l’efficacia del NLP, così come troppi dati renderebbero i risultati dei processi poco accurati. È molto difficile capire quanti dati bisogna inserire in base al compito che la macchina deve svolgere. Un altro problema è la impossibilità di inserire a mano una nuova lingua con la sua grammatica, quindi le lingue con pochi materiali didattici non possono essere utilizzate per tradurre documenti.¹¹⁷ Inoltre, anche avendo dei dati utili ed etichettati per svolgere un compito, bisogna eseguire varie prove per valutare la loro efficienza. Un altro svantaggio è l’elevato costo della manutenzione di un’architettura NLP e dei suoi dati. Un’altra serie di problemi ricorrenti sono quelli relazionati con l’analisi delle parole.

Nei linguaggi naturali alcune parole sono scritte allo stesso modo ma hanno significati diversi:

Un asse: è una retta in geometria.

Una asse: è una tavola di legno.

¹¹⁵ Cfr. Dale Markowitz, [GPT-3 Explained in Under 3Minutes | by Dale Markowitz | Towards Data Science](#)

¹¹⁶ Cfr. The Guardian, [A robot wrote this entire article. Are you scared yet, human? | GPT-3 | The Guardian](#)

¹¹⁷ Ibidem [Natural language processing - Wikipedia](#)

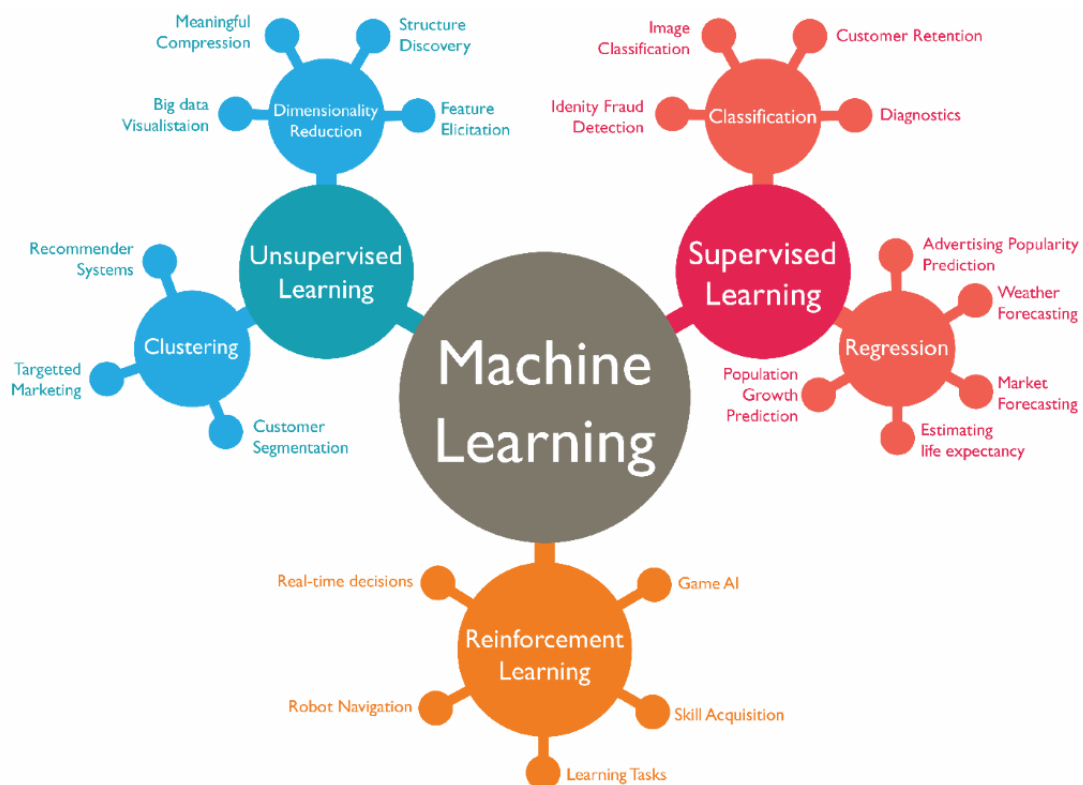
Sebbene la NLP controlli il contesto per capire il senso della frase, non sempre riesce nell'intento. In fine, la NLP incontra ancora dei problemi nel capire le sottigliezze del pensiero umano, ad esempio gli autori possono scrivere, usando ironia e sarcasmo, per esprimere i loro sentimenti ed idee al fine di dare un significato opposto al senso letterario della frase.¹¹⁸

In conclusione, la corretta e completa comprensione dei testi è ancora una sfida che mette in difficoltà anche i migliori algoritmi dell'elaborazione del linguaggio naturale e la difficile gestione dei dati per addestrarlo influisce negativamente sull'efficienza e le risorse.

¹¹⁸ Cfr. Sciforce, [Biggest Open Problems in Natural Language Processing | by Sciforce | Sciforce | Medium](#)

5.1 Comparare l'apprendimento automatico e l'apprendimento profondo

L'apprendimento automatico (*machine learning*) è uno dei primi algoritmi che ha permesso di rendere le traduzioni più efficienti, di analizzare meglio la voce dell'utente per formulare una risposta, di processare i dati e restituire un risultato nel minor tempo possibile. Programmare algoritmi che effettuino azioni complesse può essere arduo per i programmatori, anche se l'apprendimento profondo (*deep learning*) non necessita di specifici programmi per eseguire i compiti, perché è capace di imparare dai dati forniti durante l'addestramento e di creare i propri algoritmi per svolgere i vari compiti.¹¹⁹



Il grafico mostra la complessità e le capacità poliedriche dell'apprendimento automatico.

Fonte [Machine learning: cos'è, come funziona, gergo e strumenti da conoscere \(cwi.it\)](http://cwi.it)

¹¹⁹ Cfr. [Machine learning - Wikipedia](https://it.wikipedia.org/wiki/Machine_learning)

I punti di forza dell'apprendimento automatico sono molteplici:¹²⁰

- Può esaminare una grande quantità di dati ed elaborare facilmente tendenze ed abitudini delle persone. Per una azienda è importante per individuare il comportamento degli utenti al fine di identificare i prodotti giusti da suggerire loro.
- Non necessita di una supervisione da parte dell'uomo, perché è capace di imparare dagli eventi passati ed è capace di fare previsioni. Ad esempio, dalla lettura dell'indirizzo e-mail, può prevedere se è uno spam.
- Ogni volta che svolge un compito migliora in efficienza ed accuratezza perché è capace di imparare dagli eventi passati.
- Può svolgere una vasta gamma di attività dalla traduzione alla comunicazione con gli utenti tramite un linguaggio naturale.

Sebbene le capacità dell'apprendimento automatico abbiano reso le traduzioni più accurate, migliorato gli assistenti digitali, presentano degli svantaggi non trascurabili:¹²¹

- Si richiede una quantità enorme di dati non omogenei ed etichettati per addestrare l'algoritmo.
- L'algoritmo dell'apprendimento automatico richiede un ingente uso di risorse finanziarie e di energia elettrica per funzionare.
- L'algoritmo funziona in maniera autonoma ma è sensibile agli errori e se non riesce a correggerli, li ripeterà all'infinito. Se l'algoritmo ha un insieme di dati distorto, non ampio e non

¹²⁰ Cfr. Data Flair, Advantages and Disadvantages of Machine Learning Language, [Advantages and Disadvantages of Machine Learning Language - DataFlair \(data-flair.training\)](https://data-flair.training/advantages-and-disadvantages-of-machine-learning-language/)

¹²¹ Ibidem [Machine learning - Wikipedia](https://en.wikipedia.org/wiki/Machine_learning)

esaustivo, genererà dei risultati non accurati e senza un intervento tempestivo sarà difficile individuare l'errore.

Questi problemi di gestione dell'algoritmo dell'apprendimento automatico hanno portato allo sviluppo di una branca chiamata apprendimento profondo (deep learning).

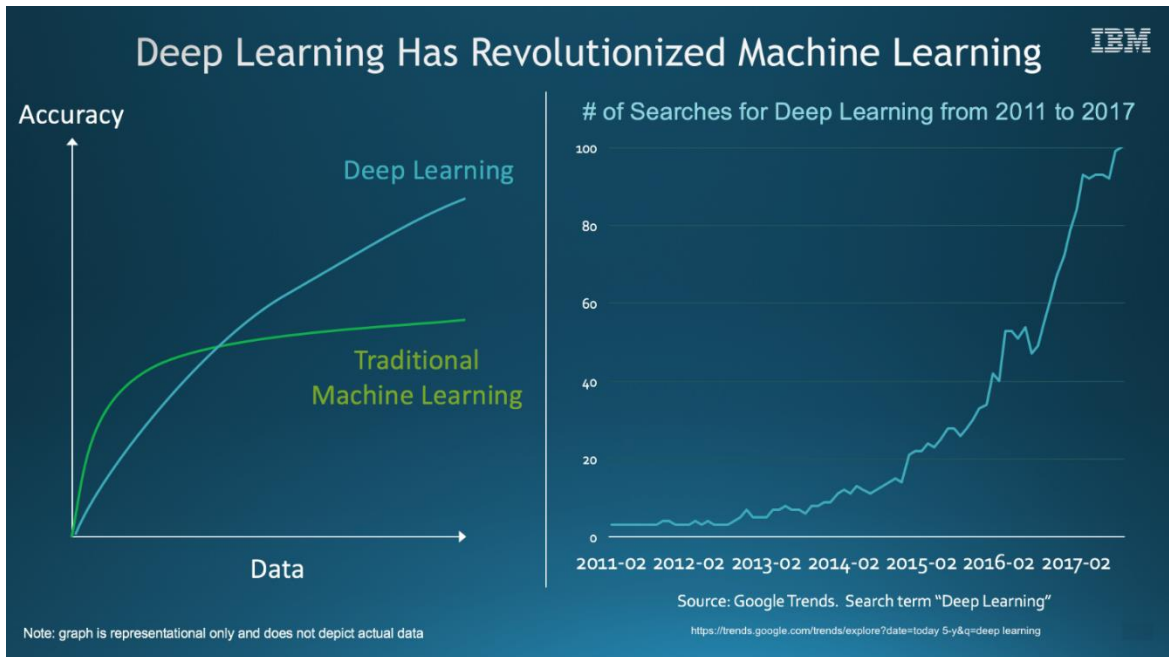
L'apprendimento profondo è composto da una gerarchia di strati – di solito dai 150 ai 200¹²²- i quali sono composti da nodi, che scompongono i dati in entrata in concetti più semplici. Quest'ultimi sono assegnati a ciascun nodo, dei vari strati, che dopo aver analizzato il concetto gli assegna un peso, più il peso è alto più quel dato andrà allo strato successivo fino a divenire parte del risultato. Ad esempio, un traduttore con l'algoritmo dell'apprendimento profondo scomporrà una frase in parole e dopo aver trovato la traduzione di ciascuna parola, le assembla per formare una frase di senso compiuto. Questo processo viene ripetuto più volte fino quando non si trova la traduzione che più rispecchia la frase in entrata- la frase deve avere un senso compiuto e il significato astratto che l'utente ha dato.¹²³

Due studi effettuati dall' IBM dimostrano l'efficacia e l'interesse verso l'apprendimento profondo. Nel grafico a sinistra si nota, che più si inseriscono i dati per addestrare l'algoritmo dell'apprendimento profondo e più accurati saranno i suoi risultati, invece nei tradizionali algoritmi di apprendimento automatico, dopo un forte miglioramento, l'accuratezza del risultato diventa stabile e non migliora più. Nel grafico

¹²² Cfr. RfWireless-World, Advantages of Deep Learning, [Advantages of Deep Learning | disadvantages of Deep Learning \(rfwireless-world.com\)](#) 20/07/2021

¹²³ Cfr. InData Labs, Deep Learning: Strengths and Challenges, [Deep Learning: Strengths and Challenges – InData Labs Blog](#)

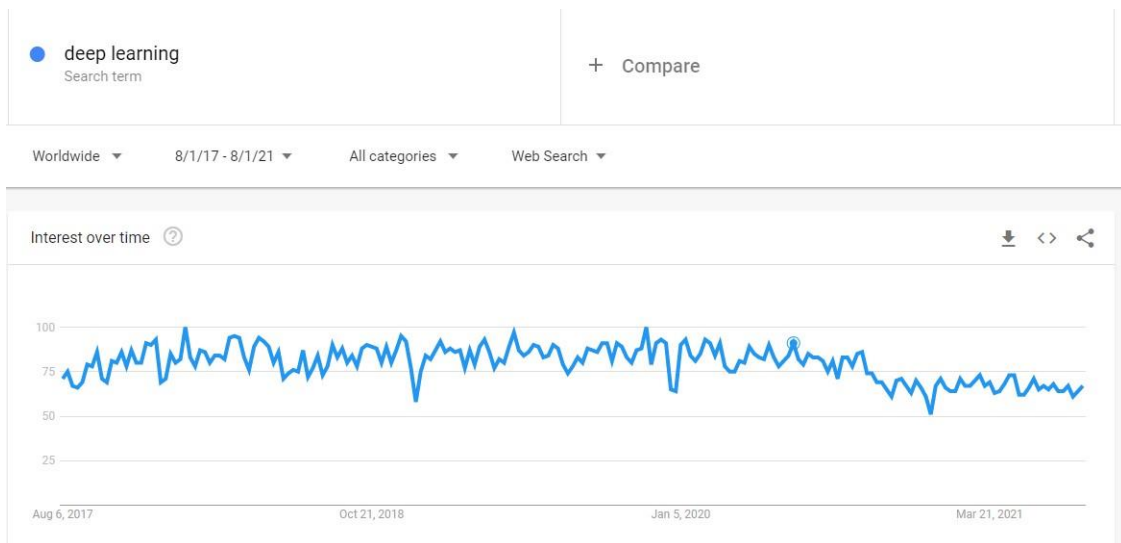
a destra si nota un crescente interesse delle persone verso l'apprendimento profondo tra il 2001 ed il 2017.



Il grafico a sinistra mostra come l'algoritmo dell'apprendimento profondo diventi più accurato man mano che si forniscano dati per addestrarlo; mentre il grafico a destra mostra il crescente interesse nelle persone sull'apprendimento profondo tra il 2011 ed il 2017

Fonte: [Deep learning performance breakthrough \(ibm.com\)](https://www.ibm.com/press/us/2015/04/20150420-deep-learning-breakthrough)

Ho svolto una ulteriore ricerca per controllare, dal 2017 al 2021, quale fosse la curva di interesse nella ricerca di articoli sull'apprendimento profondo su Google.



La curva di tendenza mostra come l'interesse verso l'apprendimento profondo è molto alto nelle ricerche su Google dal 1/08/2017 al 8/01/2021.

Fonte: [deep learning - Explore - Google Trends](#)

Sono molti i motivi per cui questo algoritmo sta entrando sempre di più nell'interesse dei singoli e delle compagnie come IBM e Google.

- La stessa rete neurale può svolgere diversi compiti ed analizzare diversi dati – immagini, testo, suoni, video- senza modificare la struttura interna, strati e nodi, come succede in una rete neurale con un algoritmo di apprendimento automatico.¹²⁴
- La maggior parte delle compagnie possiede dati in vario formato – immagini, video, documenti – e l'algoritmo dell'apprendimento automatico non è capace di analizzarli insieme per rivelare contenuti chiave per lo sviluppo aziendale. L'algoritmo di apprendimento profondo, invece, è

¹²⁴ Ivi. RfWireless-World, Advantages of Deep Learning, [Advantages of Deep Learning | disadvantages of Deep Learning \(rfwireless-world.com\)](#) 20/07/2021

stato progettato per incrociare ad analizzare dati di diversa natura.¹²⁵

- Questo algoritmo non ha bisogno di essere progettato ogni volta per imparare a svolgere un compito specifico, bensì i programmatori possono insegnarli ad eseguire un determinato compito tramite esempi, un insieme di dati o semplicemente gli chiedono di imparare da solo.¹²⁶
- L'algoritmo ha una struttura flessibile in cui si possono aggiungere nuovi strati e nodi per creare nuovi modelli di analisi ed apprendimento.¹²⁷
- L'algoritmo non ha bisogno di ricevere dati etichettati per capire di che natura sono e come analizzarli, perché non ha bisogno di linee guida per imparare. Gli altri algoritmi, come l'apprendimento automatico, hanno bisogno di ricevere dati con allegato delle informazioni che spiegano come gestirli, quindi una persona deve etichettare parole, immagini, video e questo richiede tempo e risorse. Un esempio: il radiologo deve etichettare tutte le radiografie che fa ogni giorno e deve specificare in ogni lastra che tipo di ossa sono state analizzate, se ci sono deformazioni, il significato di ogni parola e questo è un grande dispendio di tempo e risorse.¹²⁸

¹²⁵ Cfr. Valeryia Shchutskaya, Deep Learning: Strengths and Challenges, [Deep Learning: Strengths and Challenges – InData Labs Blog](#)

¹²⁶ Cfr. Alice Kint, 32 Advantages and Disadvantages of Deep Learning [32 Advantages and Disadvantages of Deep Learning | by Alice Kint | Medium](#)

¹²⁷ Ivi. Alice Kint, 32 Advantages and Disadvantages of Deep Learning [32 Advantages and Disadvantages of Deep Learning | by Alice Kint | Medium](#)

¹²⁸ Ibidem, Valeryia Shchutskaya, Deep Learning: Strengths and Challenges [Deep Learning: Strengths and Challenges – InData Labs Blog](#)

Sebbene l'apprendimento profondo presenti molte migliorie rispetto agli altri algoritmi con rete neurale artificiale, necessita di miglioramenti.

- Ha bisogno di molti dati per imparare a svolgere le funzioni date, tuttavia possono contenere informazioni sensibili come reddito, interessi e stato civile. Questo sta sollevando preoccupazioni sulla protezione e riservatezza dei dati perché c'è il rischio che qualcuno li venda o li visualizzi. Tuttavia negli ultimi anni, durante la fase di addestramento i dati sono anche cifrati per proteggerli.¹²⁹
- È difficile comprendere come questo algoritmo con la sua rete neurale giunga ad una particolare soluzione, in quanto è quasi impossibile studiare ogni passaggio dell'elaborazione di ciascun nodo e strato. Infatti, il pensiero della rete neurale è il frutto di migliaia di neuroni predisposti in strati interconnessi ed è difficile capire che ragionamento hanno fatto. Questa mancanza di trasparenza nel ragionamento dell'algoritmo non dà la possibilità di prevedere quando si possano verificare errori.¹³⁰
- L'algoritmo è capace di imparare da solo come svolgere le varie attività quindi è difficile capire come il suo ragionamento evolva nel tempo.¹³¹

¹²⁹ Cfr. Aysegul Takimoglu, What is Deep Learning?, [What is Deep Learning? Explore Use Cases / Examples, Benefits \(aimultiple.com\)](#)

¹³⁰ Ibid. Valeryia Shchutskaya, Deep Learning: Strengths and Challenges [Deep Learning: Strengths and Challenges – InData Labs Blog](#)

¹³¹ Ibidem Alice Kinth, 32 Advantages and Disadvantages of Deep Learning [32 Advantages and Disadvantages of Deep Learning | by Alice Kinth | Medium](#)

- Sebbene questo algoritmo sia più veloce degli altri, richiede una potenza di calcolo della scheda grafica molto elevata, durante la fase di addestramento, che è la fase più costosa in termini di tempo e risorse finanziarie.¹³²

In conclusione l'apprendimento profondo offre la possibilità di processare funzioni, di varia natura e complessità, in tempi brevi, migliorare la sua analisi apprendendo dai processi passati e restituire risultati molto accurati. In un futuro prossimo, compagnie che offrono servizi e necessitano di analisi sempre più accurate circa le preferenze degli utenti e molti ambiti lavorativi come la medicina, la traduzione e la ricerca ne gioveranno in efficienza e risparmio di risorse. Tuttavia ci sono ancora degli aspetti dell'apprendimento profondo che devono essere migliorati.

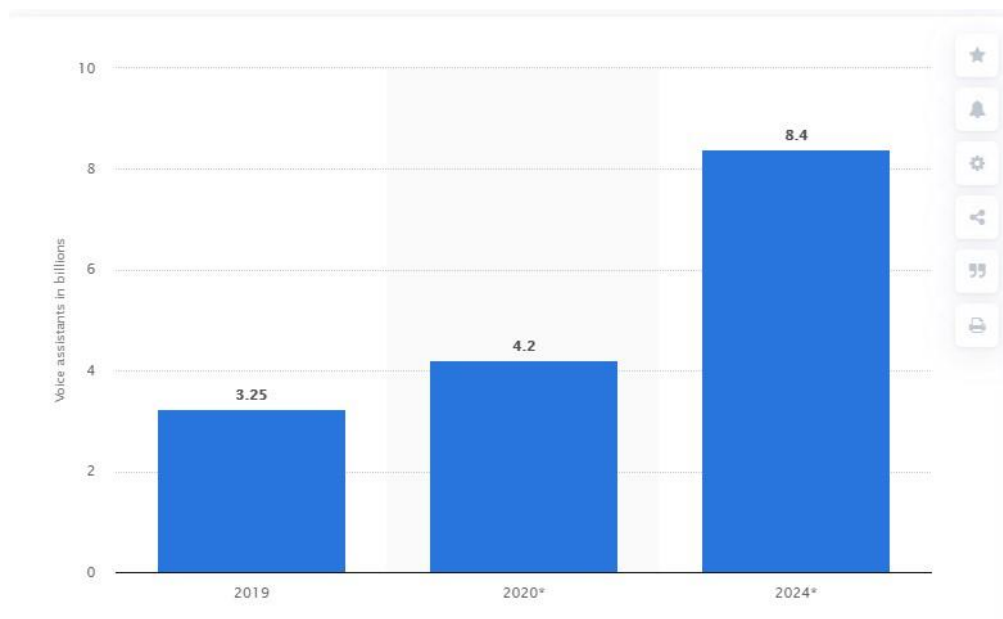
¹³² Cfr. Stanford University, Drawbacks of Deep Learning, [Drawbacks of Deep Learning | MS&E 238 Blog \(stanford.edu\)](#)

6. Il futuro dell'interazione tra l'intelligenza artificiale ed i linguaggi naturali

Il futuro dell'interazione tra l'intelligenza artificiale ed il linguaggio umano non è una via a senso unico bensì si dirama in due strade. Da una parte si pensa a come la vita sarà agevolata grazie ad assistenti digitali, che con comandi vocali potranno svolgere svariate funzioni ed aiutare le persone appartenenti a categorie svantaggiate. Dall'altro canto ci sono alcuni timori su come cambierà la maniera di comunicare tra persone e che impatto ci sarà sul mondo del lavoro.

La compagnia Adobe ha raccolto varie stime che indicano come gli assistenti vocali digitali faranno sempre più parte integrante della vita di tutti i giorni. Entro il 2026 quasi l'80% degli utenti con uno smartphone od una smart TV userà solamente comandi vocali per comunicare con l'assistente digitale, entro il 2023, 8 miliardi di assistenti vocali digitali saranno in uso e il 25% delle interazioni tra uomo e macchina nei posti di lavoro avverrà tramite comandi vocali, in confronto al 3% del 2019. Nel primo trimestre del 2021, negli Stati Uniti, il 16% dei consumatori preferisce un agente digitale vocale che li aiuti nell'organizzare l'agenda, chiamare e scrivere messaggi.¹³³ Inoltre il sito internet www.statista.com, uno fra i più accreditati nel fornire rapporti e statistiche circa il mercato di beni e servizi, finanza e tecnologia, ha elaborato un grafico di quanti assistenti vocali sono stati utilizzati nel 2019,2020 e predetto il numero per il 2024.

¹³³Cfr. Adobe Press, 15 mind-blowing stats about voice assistants, [15 mind-blowing stats about voice assistants \(adobe.com\)](https://www.adobe.com/press/insights/15-mind-blowing-stats-about-voice-assistants)



Il grafico mostra il numero di assistenti vocali, nell'ordine dei miliardi, usati nel 2019,2020 e predetto il numero per il 2024.

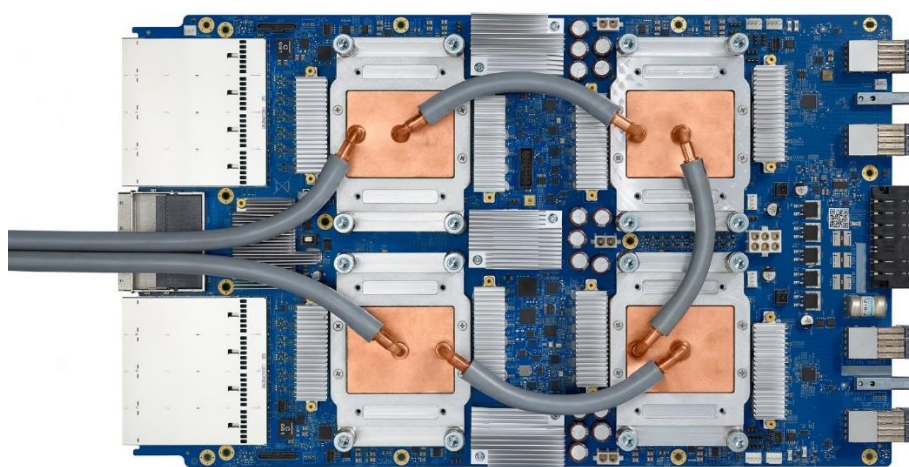
Fonte: • [Number of voice assistants in use worldwide 2019-2024 | Statista](#)

Inoltre, gli schermi degli smartphone si stanno rimpicciolendo e questo rende la navigazione tra le sue funzionalità, tramite il tocco, sempre più complicato, di conseguenza gli utenti preferiranno i comandi vocali come mezzo di comunicazione principale nei prossimi anni. Già il settore medico sta sperimentando l'uso di assistenti vocali per fornire supporto ai medici durante una operazione. *Amazon Alexa For Business* è capace di organizzare un calendario di riunioni e gestire le apparecchiature audio visive connesse alla rete WI-FI di una sala riunioni tramite comandi vocali.¹³⁴ Una fra le compagnie più all'avanguardia è Google in quanto sta per portare il futuro nelle nostre case con due prodotti innovativi che rivoluzioneranno per sempre il nostro approccio col cellulare e come comunicheremo con le persone.

Il primo è la nuova versione dell'assistente digitale che sarà incorporato nel nuovo smartphone, Google Pixel 6. Con solo la voce dell'utente, si possono scorrere le gallerie fotografiche, dettare messaggi

¹³⁴ Cfr. Eric Bieller, [What's The Future Of Human-Computer Interaction \(HCI\)? \(careerfoundry.com\)](#)

ed allegare file multimediali.¹³⁵ Google, che dal 2014 è diventato il porta bandiera dello sviluppo di reti neurali e dell'apprendimento profondo, ha progettato questo assistente usando l'apprendimento automatico.¹³⁶ Un documento della Cornell University, Stati Uniti, spiega come Google ha allenato per anni la sua rete neurale nello scoprire il nome delle strade in decine di migliaia di foto scattate per Google Map.¹³⁷ Nel 2016, la compagnia ha sviluppato il Tensor Processing Unit, un microchip capace di rendere le reti neurali più veloci.



Una scheda di Google contenente il Tensor Processing Unit.

Fonte: [Why Google believes machine learning is its future | Ars Technica](#)

Google ha deciso di equipaggiare i cellulari con questo tipo di microchip per dare più capacità di elaborazione ai cellulari, così facendo l'assistente digitale può funzionare anche senza internet a differenza degli altri assistenti digitali, come Alexa, che funzionano solamente con la connessione alla rete. Con questo assistente, Google mira ad aiutare le persone appartenenti a gruppi svantaggiati, che hanno difficoltà

¹³⁵ In questo video si può vedere il prototipo dell'assistente digitale all'opera. [Next Generation Google Assistant: Demo 2 at Google I/O 2019 - YouTube](#)

¹³⁶ Cfr. TIMOTHY B. LEE, [Why Google believes machine learning is its future | Ars Technica](#)

¹³⁷ Cfr. Ian J. Goodfellow, [\[1312.6082\] Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks \(arxiv.org\)](#)

linguistiche a comunicare liberamente.¹³⁸ Questo assistente può leggere qualsiasi testo in lingua originale o in una altra lingua, crea sottotitoli per i video e durante una chiamata trascrive in tempo reale su un documento, ciò che dice l'interlocutore per aiutare le persone con problemi d'udito.¹³⁹

La seconda invenzione è Google Duplex, una estensione dell'assistente digitale di Google. Quando l'utente vuole fare una chiamata, organizzare un evento o prenotare al ristorante, deve semplicemente inserire i parametri come numero da chiamare, a che ora si vuole prenotare e per quale motivo si sta chiamando e l'assistente di Google avvierà Duplex che chiamerà e parlerà con l'interlocutore. Al termine della chiamata l'utente riceverà un riepilogo della chiamata. Duplex riesce ad imitare la voce umana quasi perfettamente, ad usare intercalari come “cioè” e intersezioni come “um”.¹⁴⁰



Lo schema mostra il susseguire di passaggi per usare Duplex.

Fonte: [NLP in the Real World: Google Duplex | by Jerry Wei | Towards Data Science](#)

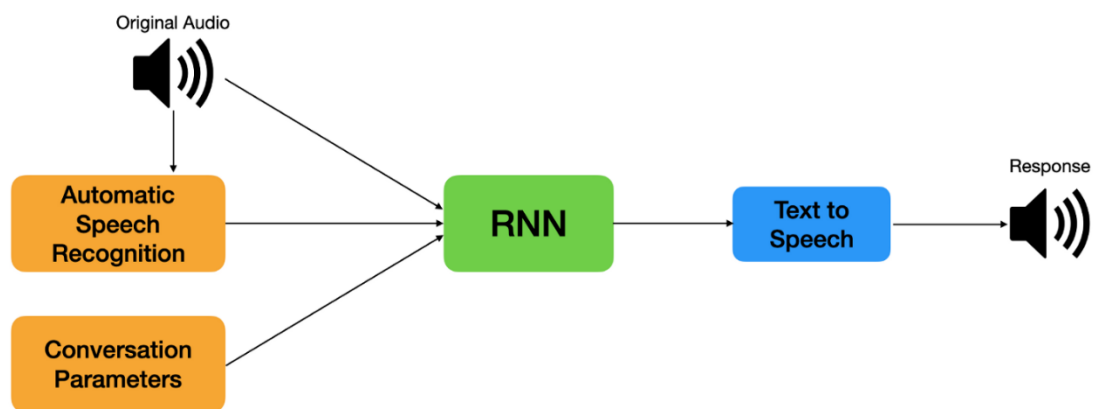
Duplex è la espressione massima dell'elaborazione del linguaggio naturale (NLP), in quanto si basa su una rete neurale ricorrente (RNN) per tradurre ciò che dice l'interlocutore e formulare le frasi di risposta. Incorporato in Duplex, c'è *Automatic Speech Recognition* (ASR) per riconoscere la voce e contestualizzare le frasi di chi parla ed il modulo

¹³⁸ Cfr. Il video mostra come l'assistente aiuti a comunicare persone appartenenti a gruppi svantaggiati, [Project Euphonia: Helping everyone be better understood - YouTube](#)

¹³⁹ Ibid. Timoty B. Lee, [Why Google believes machine learning is its future | Ars Technica](#)

¹⁴⁰ Cfr. Il video mostra le funzionalità di Duplex, [Google's AI Assistant Can Now Make Real Phone Calls - YouTube](#)

text-to-speech (TTS) per trasformare i messaggi scritti di Duplex in messaggi vocali.¹⁴¹



Lo schema mostra la complessa architettura interna di Duplex

Fonte: [NLP in the Real World: Google Duplex | by Jerry Wei | Towards Data Science](#)

¹⁴¹ Cfr. Jerry Wei, NPL in the Real World, Towards Data Science, [NLP in the Real World: Google Duplex | by Jerry Wei | Towards Data Science](#)

6.1 Perplexità e preoccupazioni circa un futuro basato sull'intelligenza artificiale

Sebbene l'intelligenza artificiale sia acclamata da molti perché porta benefici a tutta la popolazione e facilita la vita di persone appartenenti a fasce svantaggiate, cittadini e organismi nazionali stanno sollevando delle perplessità.



L'immagine mostra la mano dell'uomo e della macchina, guidata dall'intelligenza artificiale, intente a trovare un terreno comune per creare un futuro migliore per entrambi.

Fonte: [Safeguarding human rights in the era of artificial intelligence - Human Rights Comments - Commissioner for Human Rights \(coe.int\)](#)

Molti hanno chiesto al governo statunitense se è moralmente accettabile che un assistente digitale parli come un umano senza che l'interlocutore ne sia al corrente, perché questa azione può essere percepita come un inganno. Inoltre, molti si chiedono se malintenzionati usino l'assistente di Google per chiamare le persone col fine di rubare gli

estremi del conto bancario e rivendere i dati sensibili del mal capitato, il quale pensa di aver effettuato un acquisto.¹⁴²

Amazon Alexa è diventato l'oggetto di molte controversie negli ultimi anni, il Senatore statunitense Coons ha chiesto ad Amazon quali siano le politiche per la gestione e la sicurezza dei dati degli utenti. Amazon ha spiegato che ci sono degli addetti al controllo dei messaggi vocali e le trascrizioni delle registrazioni vocali degli utenti sono salvate nei server di Amazon indefinitamente. Il Senatore Coons teme per la sicurezza degli statunitensi in quanto gli operatori di Amazon possono accedere ai dati dei loro clienti e la compagnia può venderli a terzi, che ancora non è chiaro come li usino.¹⁴³

Molti economisti e scienziati, come Paul Romer (premio Nobel) e Nick Bloom (professore alla Stanford University), sono preoccupati per il fenomeno della disoccupazione portata dalla tecnologia (*Technological Unemployment*). Il lato negativo dell'innovazione è dimostrato da un taglio sempre più netto dell'impiego nel settore manifatturiero e agricolo.¹⁴⁴ Anche gli impiegati in ufficio ne hanno risentito, perché ora gli algoritmi riescono ad analizzare curriculum e scremare i candidati per una posizione lavorativa più velocemente e senza discriminarli e questo ha portato alla riduzione del 30% degli impiegati nelle risorse umane.¹⁴⁵ Stephen Hawking, premio Nobel per la fisica, ha descritto l'intelligenza artificiale come il male della società, se i governi nazionali in collaborazione con l'Unione Europea non approvassero leggi capaci di

¹⁴² Cfr. Chris Butler, Is Google Duplex ethical and moral?, Towards Data Science [Is Google Duplex ethical and moral? | by Chris Butler | Towards Data Science](#)

¹⁴³ Cfr. Chris Coons, [Amazon responds to Sen. Coons' concerns about consumer privacy practices for Alexa devices \(senate.gov\)](#)

¹⁴⁴ Cfr. Tom Taulli, Artificial, Intelligence Basics, p. 176, Apress, Monrovia, Ca, USA 2019.

¹⁴⁵ Cfr. Eric Rosenbau, IBM artificial intelligence can predict with 95% accuracy which workers are about to quit their jobs, <https://www.cnbc.com/2019/04/03/ibm-ai-can-predict-with-95-percent-accuracy-which-employees-will-quit.html>

indirizzarne l'uso per scopi umanitari come eradicare la fame, ridurre la povertà e migliorare la qualità della vita.¹⁴⁶ Inoltre, negli Stati Uniti, alcuni parlamentari hanno chiesto al Congresso di analizzare il disegno di legge “*Algorithmic Accountability Act*”, il quale prevede che le aziende verifichino la trasparenza dei loro sistemi gestiti dall'intelligenza artificiale e di come sono trattati i dati dei loro utenti.¹⁴⁷

In conclusione, l'intelligenza artificiale tramite l'elaborazione del linguaggio naturale ha portato e porterà molte migliorie, ad esempio, persone con problemi di udito possono parlare liberamente al telefono, si possono tradurre documenti in varie lingue in pochi secondi e i lavoratori sono sollevati dai lavori ripetitivi e stancanti. Tuttavia, ci sono ancora interrogativi e perplessità su quale sia il modo migliore per gestire i dati degli utenti, senza violare i protocolli di riservatezza e come proteggere le persone da chi vuole usare l'intelligenza artificiale per ledere agli altri.

¹⁴⁶ Cfr. Arjun Kharpal, Stephen Hawking says A.I. could be ‘worst event in the history of our civilization’, <https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html>

¹⁴⁷ Ibid. Cfr. Tom Taulli, *Artificial Intelligence Basics*, pag 180, Apress, Monrovia, Ca, USA 2019.

7. Conclusioni

Lo scopo di questa tesi è di mostrare il rapporto intrinseco tra l'intelligenza artificiale e le lingue naturali. Lo studio analitico di queste ultime è stato fondamentale nello sviluppo di algoritmi, la cui analisi si avvicina al pensiero umano, che permettono di svolgere funzioni sempre più complesse in tempi brevi al fine di ottimizzare il tempo e le risorse. Inoltre, la tesi mette in evidenza la complessa architettura hardware, che permette di eseguire operazioni come la traduzione, impartire comandi ad un assistente vocale per una ricerca, una comparazione di dati e assistenza durante il lavoro. Queste operazioni, che possono sembrare facili, richiedono un'analisi approfondita della lingua e della comunicazione umana, e il supporto di una rete neurale ben addestrata che è capace di imparare dagli eventi passati.



L'immagine mostra la mano di un uomo e di una macchina, guidata dall'intelligenza artificiale, che tentano di forgiare un mondo equilibrato dove ci saranno ancora opportunità lavorative per le persone.

Fonte: [Intelligenza Artificiale, sempre più vicina al pensiero umano \(teleambiente.it\)](https://www.teleambiente.it)

Sono passati 76 anni da quando Padre Busa decise di cooperare con l'IBM per creare un calcolatore capace di analizzare parola per parola

tutte le opere di San Tommaso D'Aquino, il suo lavoro certosino durò 30 anni perché era limitato dalla tecnologia di allora. In molti pensano che l'analisi delle lingue naturali non si sarebbe sviluppata così velocemente senza il prezioso lascito di Padre Busa, 179 testi con un totale di 10.631.980 parole indicizzate scritte su schede perforate. A sua insaputa divenne il pioniere della linguistica computazionale e l'inventore dell'ipertesto e sarà sempre ricordato come colui che insegnò la letteratura ai calcolatori.¹⁴⁸

Con la tecnologia attuale, una ricerca indicizzata di parole avviene in pochi secondi e l'inserimento di dati per addestrare la macchina avviene nell'arco di settimane, tutto ciò è possibile grazie alle reti neurali e a Padre Busa, la Brown University, la Oxford University e Wikipedia, i quali hanno creato i corpora che ancora formano la base della conoscenza delle più avanzate reti neurali.

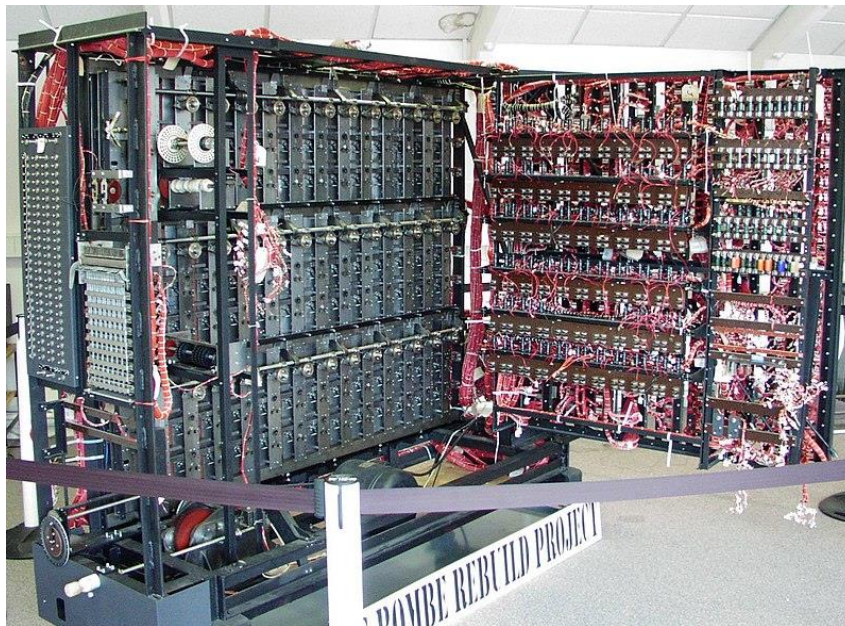


La foto mostra Padre Busa mentre inserisce l'opera di Tommaso D'Aquino nel mega calcolatore IBM, 1950.

Fonte: <https://www.controverso.cloud/6660/>

¹⁴⁸ Cfr. Marco Passarotti, [IBM thinkMagazine - Padre Busa, il gesuita che inventò l'ipertesto grazie ai computer IBM](#), 2018.

Un altro padre fondatore della intelligenza artificiale è Alan Turing e solamente ora possiamo comprendere il valore dell'eredità che ha lasciato. Ideò il “*turing test*” in cui un moderatore conversa con due utenti e deve capire chi è la macchina in base alle risposte ricevute. Risolvere tale test è tutt'ora il principale obiettivo di molti ricercatori e case informatiche, che come obiettivo hanno prefissato la creazione di un calcolatore capace di ragionare come un essere umano. Sebbene le loro creazioni non siano in grado di risolvere il “*turing test*”, hanno fatto avanzare la società nel campo dell'intelligenza artificiale con reti neurali che sono capaci di tradurre testi senza avere come dati di addestramento le regole di grammatica delle lingue, comparare dati e l'andamento dell'interesse degli utenti, creare di assistenti digitali che parlano al cellulare al posto nostro e gestiscono altri dispositivi informatici connessi alla rete locale, fornire supporto alle fasce più svantaggiate e ridurre il carico di lavoro ripetitivo e laborioso ai lavoratori. Alan Turing sarà per sempre ricordato come il padre dell'intelligenza artificiale.



La foto mostra il calcolatore ideato da Turing che permise al Regno Unito di decifrare i messaggi dei Nazisti durante la Seconda Guerra mondiale. Tale macchina ispirò i primi traduttori. Fonte: [Test di Turing - Wikipedia](#)

La continua digitalizzazione e la corsa a migliorare l'elaborazione dei linguaggi naturali (NLP) hanno creato posti di lavoro e reso la vita delle persone più facile, veloce ed automatizzata. Ora è più facile comunicare in lingue diverse grazie ai traduttori neurali, analizzare documenti in pochi secondi e per chi ha problemi di udito l'assistente digitale lo aiuta a parlare col cellulare.

Tuttavia, anche io concordo sulle perplessità circa l'abuso dell'intelligenza artificiale e la protezione dei dati sensibili degli utenti. Tra Amazon ed il governo degli Stati Uniti c'è una discussione sulla protezione dei dati degli utenti perché la compagnia salva i loro messaggi vocali per sempre, gli operatori di Amazon possono leggerli ed ascoltarli, alcuni utenti hanno ascoltato per sbaglio conversazioni di altre persone per via di un problema di elaborazione dati di Alexa, l'assistente ha eseguito pagamenti non autorizzati e in alcuni casi ha pronunciato parole suicide.¹⁴⁹

Inoltre, c'è il problema della disoccupazione causata dall'intelligenza artificiale che ha colpito i settori secondari e terziari della nostra società per via del continuo bisogno di raggiungere l'efficienza, la ottimizzazione del tempo ed essere competitivi. Tale fenomeno aumenta la disparità sociale, perché i lavoratori appartenenti al ceto medio sono rimpiazzati da computer, ad esempio le aziende preferiscono avere pochi impiegati che controllano i documenti tradotti da un traduttore neurale ,invece di assumere del personale che impiega più tempo nel tradurre manualmente un documento, assistenti digitali che organizzano l'agenda degli amministratori e programmi capaci di

¹⁴⁹ Cfr. Dorian Linskey, '[Alexa, are you invading my privacy?' – the dark side of our voice assistants | Amazon Alexa | The Guardian](#), 2019.

creare test di sbarramento e scremare i candidati durante il processo di selezione del nuovo personale.¹⁵⁰

Un altro aspetto è la necessità di avere leggi che gestiscano e moderino l'uso dell'intelligenza artificiale per evitare abusi e pratiche volte a ledere i cittadini e appropriarsi dei loro dati sensibili. Tuttavia bisogna guardare anche il quadro generale e sottolineare che la diminuzione di certe figure lavorative, perché rimpiazzate dall'intelligenza artificiale, non porta solamente alla disoccupazione ma apre la strada a nuovi impieghi. È importante che le persone apprendano nuove competenze per essere riassorbite nel mercato del lavoro e l'ausilio dei governi nazionali è fondamentale in questo, perché necessitano corsi di qualità e supporto per le persone che ricominciano a studiare a livello accademico per cambiare carriera. Un esempio sono i paesi scandinavi che incoraggiano i cittadini anche in età avanzata a frequentare ulteriormente l'università per apprendere nuove conoscenze e strumenti.

In conclusione, lo sviluppo dell'intelligenza artificiale e dell'elaborazione del linguaggio naturale devono essere visti come i mezzi che ci hanno permesso di progredire e migliorare le nostre vite. Molte comodità di ora sono possibili grazie all'intelligenza artificiale e fino a 15 anni fa era impensabile avere Google Duplex che effettua le chiamate per noi e prenota un appuntamento dal dottore. È importante sottolineare che ciascuno di noi, organizzazioni internazionali e governi nazionali attuali non abusiamo della tecnologia per dei fini che non recano beneficio alla collettività. Mi auguro che possiamo divenire più

¹⁵⁰ Cfr. United Nations Press, [Will robots and AI cause mass unemployment? Not necessarily, but they do bring other threats | United Nations](#)

coscienti delle comodità che l'elaborazione del linguaggio naturale offre e di farne buon uso.

English section

Introduction

Artificial intelligence is the science of coding software capable of emulating human intellectual processes such as reasoning and learning from past experience.¹⁵¹ These features enable machines to carry out tasks such as translating text, providing customer support, and looking for information of any kind for business and private purposes.



The image shows the result of combining the human mind and machines.

Source: [Beginning Your Journey to Implementing Artificial Intelligence | SoftwareONE Blog](#)

This thesis analyses in detail the chronological and technical development of the relationship between artificial intelligence and natural languages - the system composed of sounds and grammatical structures that enable people to interact with each other. In 1949, the Jesuit Roberto Busa began to digitise the corpus of Thomas Aquinas's works, and this led to the creation of computational linguistics. In the 1950s, the studies of the linguist Noam Chomsky became the basis for coding algorithms capable of understanding natural languages. Subsequent studies allowed researchers to build artificial neural networks being capable, for example, of assisting users with various tasks, translating almost as well as a professional human translator, analysing data and studying consumer behaviour.

¹⁵¹ Cf. IBM Cloud Education, [What is Artificial Intelligence \(AI\)? | IBM](#) June 3, 2020

1. Fundamentals of artificial intelligence, attempts to create machines capable of understanding natural languages

IT companies have always tried to build computers capable of interacting with users through a natural language to perform tasks such as translating a document, organising an agenda in an automated way and supporting a doctor during an operation. These processes have been made possible by computational linguistics, which focuses on developing systems that can analyse, interact with one or more natural languages.¹⁵²

In 1946, Jesuit Roberto Busa, with the support of IBM, was the pioneer in using a computer to analyse Thomas Aquinas's *Omnia* work. In 30 years of work, he created the *Index Thomisticus* consisting of 70,000 pages with a total of 20 million lines.¹⁵³ In 1950, the mathematician Alan Turing published in the peer-reviewed academic journal *Mind* the article "Computing Machinery and Intelligence" in which he tried to find an answer to the question "Can machines think?".¹⁵⁴ To do so, he created the famous "Turing test", in which two players, a person and a computer, have to answer the interrogator's questions, who has to identify which of the two players is the computer on the basis of the answers received. No computer has ever been able to pass the test, so far.¹⁵⁵

¹⁵² Cf. Salvatore Sorce, *Introduzione alla linguistica computazionale*, [Microsoft PowerPoint - SEI1213_01_Linguistica_Computazionale.ppt \(unipa.it\)](#), p. 5.

¹⁵³ Martina Dell'Utri, [International Busa Machines: come un gesuita ha insegnato la lettura ai computer | Controverso](#), December 11, 2020

¹⁵⁴ Cf. Alan Turing, *Mind*, Volume LIX, Issue 236, October 1950, Oxford, p. 434. [I.—COMPUTING MACHINERY AND INTELLIGENCE | Mind | Oxford Academic \(oup.com\)](#)

¹⁵⁵ Ivi pp. 436-439.

In 1957, the linguist Noam Chomsky, studying the development of natural language¹⁵⁶ in humans and the ability of children to communicate without formal education, contributed to codifying algorithms capable of analysing and creating sentences in a natural language.¹⁵⁷ Chomsky developed the generative grammar, that describes how to formulate natural language sentences, taking inspiration from formal grammar, focusing on how the brain formulates sentences¹⁵⁸ and highlighting the innate ability of children to form and understand sentences they have never heard before. This ability allows people to understand whether a sentence makes grammatical sense.¹⁵⁹

For example, children unconsciously learn that you say:

He can talk to her.

Instead of: he can talk her.

Chomsky conducted a study on a sample of children up to the age of 10, to understand their language competency and they all succeeded.¹⁶⁰ In fact, children learn that words have a logical order and each part has a specific task.¹⁶¹ This innate competence is needed to learn any language and Chomsky based on this created the theory of universal grammar, which provides a system that facilitates language learning.

¹⁵⁶ Natural languages are means of communication that a group of individuals uses to communicate; Italian, English, and French are examples of natural languages.

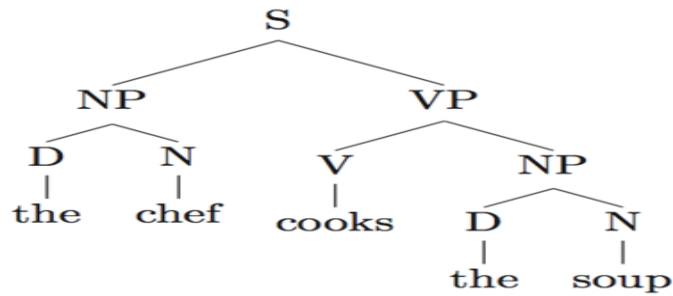
¹⁵⁷ Cf. Francesca Fiore, Il linguaggio universale secondo Noam Chomsky – Introduzione alla Psicologia, [Il linguaggio universale secondo Noam Chomsky - Psicologia \(stateofmind.it\)](#), December 3, 2015

¹⁵⁸ Cf. Generative Grammar, [Generative grammar | Britannica](#)

¹⁵⁹ Cf. [Grammatica generativa - Wikipedia](#)

¹⁶⁰ Cf. [The Chomsky Experiments | by Thomas Roeper | The New York Review of Books \(nybooks.com\)](#) March 3, 1981

¹⁶¹ Ibid. Lettere e Filosofia, Unife, Riassunto Teoria Chosmiana, pp. 6-16 [Lezione9_2.ppt \(unife.it\)](#), 2014



Unconscious mental scheme used by children when learning to speak. Chomsky named it the tree diagram of generative grammar. Sentence (S); Noun Phrase (NP); Verb Phrase (VP); Determiner (D), Noun (N); Verb (V).

Source: [Parse tree - Wikipedia](#)

The creation of universal grammar and the various generative grammars led Chomsky to organise them in his famous homonymous hierarchy consisting of four groups:¹⁶²

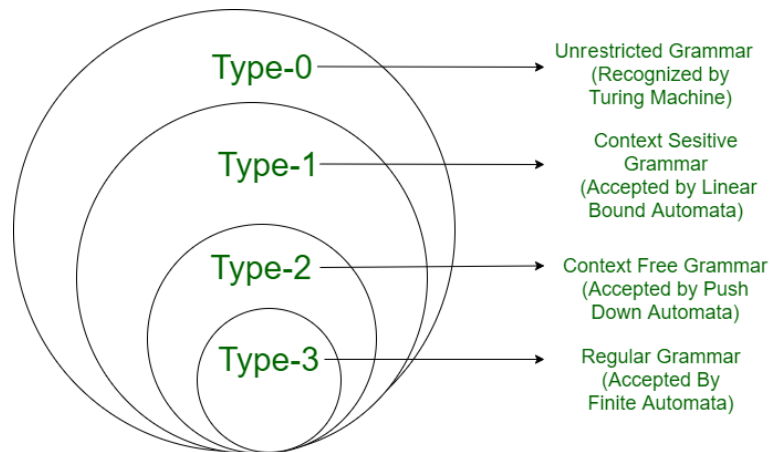
Type-0 grammar encompasses all formal grammars.

Type-1 grammar is a type of context-sensitive grammar.

Type-2 grammar is context-free and offers the possibility of developing algorithms for the lexical analysis of sentences.

Type-3 grammar allows us to design algorithms for seeking words and phrases in documents, browsers and various writing and translation software.

¹⁶² Cf. James Rogers, Formal language theory: refining the Chomsky hierarchy, [Formal language theory: refining the Chomsky hierarchy | Philosophical Transactions of the Royal Society B: Biological Sciences \(royalsocietypublishing.org\)](#), pp. 2- 20, July 19, 2021



The figure shows Chomsky's hierarchy. Source: [Chomsky Hierarchy in Theory of Computation - GeeksforGeeks](#)

Between 1959 and 1963, John Backus and Peter Naur, scientists at IBM, developed the 'Backus-Naur Form', which sets out the formal rules to be used to illustrate the syntax of a computer language and network protocols.¹⁶³ In 1968 at MIT (Massachusetts Institute of Technology), Professor Marvin Minsky¹⁶⁴ was designing a computer that could communicate with people and move geometric objects in a virtual environment. In order to develop the communication part, Minsky asked the PhD student Terry Winograd¹⁶⁵ if he wanted, for his thesis, to code an algorithm so the computer could understand natural language instructions. Over the next two years, by combining studies of syntax, semantics and the deductive method, Winograd developed SHRDLU,¹⁶⁶ an algorithm capable of communicating in English.¹⁶⁷ In 1949, engineer Warren Weaver,¹⁶⁸ wrote the vademecum entitled 'Translation', in which he writes about coding programs capable of translating without human supervision. The process consisted of replacing each word in the text

¹⁶³ Cf. [Backus-Naur Form | Sapere.it](#)

¹⁶⁴ Cf. Michael Aaron Dennis, [Marvin Minsky | American scientist | Britannica](#), August 5, 2021

¹⁶⁵ More information about Terry Winograd. [Terry Winograd - Wikipedia](#)

¹⁶⁶ More information about Nigel Word, Tokyo University, Introduction to Shrdlu, [S00056 1..4 \(utep.edu\)](#)

¹⁶⁷ Cf. Terry Winograd, Procedures as a representation of data in a computer program for understanding natural language, MIT, January 1971, pp. 18- 24. [AITR-235.pdf \(stanford.edu\)](#)

¹⁶⁸ Warren Weaver was an American scientist and mathematician, considered to be the father of machine translation.

with the corresponding word in the other language and using applied statistics to calculate the frequency with which each word was used in a specific context.¹⁶⁹ In 1952, Yehoshua Bar-Hillel,¹⁷⁰ a MIT professor, proposed the innovative idea of equipping computers with a universal encyclopaedia.¹⁷¹ IBM, MIT, Brown University and Oxford University began creating linguistic corpora¹⁷² to train translators.¹⁷³ From 1975 onwards, with the introduction of the Systran translator, translation accuracy improved, yet only in aeronautics.¹⁷⁴ In the 2000s, Google introduced its own translator that was similar to IBM's but with more corpora available,¹⁷⁵ and in 2016, the company decided to use neural machine translation because it is faster and more accurate.¹⁷⁶

¹⁶⁹ Cf. Warren Weaver, Translation, pp. 15- 24, [weaver001.pdf \(stanford.edu\)](#)

¹⁷⁰ Cf. Kasher, Essay in memory of Yehoshua Bar-Hillel, Boston, pp. 1-2.

¹⁷¹ Ivi. Pp. 5-6.

¹⁷² Corpus in singular, corpora in plural.

¹⁷³ Cf. Brown University, Brown Corpus Manual For Use With Digital Computer, USA, 1964, pp. 3-8. [Brown Corpus Manual \(uib.no\)](#)

¹⁷⁴ Cf. [The Lingua File from TLF Translation: SYSTRAN: A Brief History of Machine Translation](#)

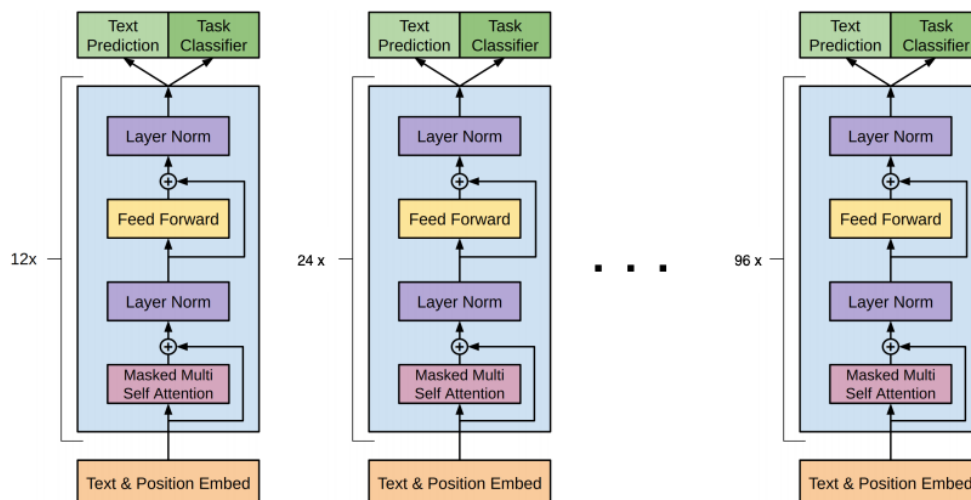
¹⁷⁵ Cf. History of Translation. [Google Translate - Wikipedia](#)

¹⁷⁶ Cf. Joe Sommerlad, GOOGLE TRANSLATE: HOW DOES THE MULTILINGUAL INTERPRETER ACTUALLY WORK?, March 2021, The Independent

2. Modalities used by computers to analyse human languages; cutting-edge GPT

In March 2020, the US company OpenAI¹⁷⁷ launched the language model called Generative Pre-trained Transformer 3, GPT-3. It is a versatile autoregressive language model, which is able to create texts of up to 800 words, very similar to what a person can write and can be set up for other functionalities such as generating quizzes, websites and recommending books according to users' preferences.¹⁷⁸

GPT-3 can create very accurate texts because its architecture is a neural deep learning network with 175 billion self-optimising coefficients, which learn from past events as the software carries out its given tasks. The OpenAI company has used various sets of corpora from different websites including CommonCrawl and Wikipedia Corpus, as training data for the GPT-3.¹⁷⁹



The figure shows the recursive diagram of the GPT-3's recursive diagram.

Source: [GPT-3 An Overview · All things \(dizlab.github.io\)](https://dizlab.github.io/gpt-3-an-overview/)

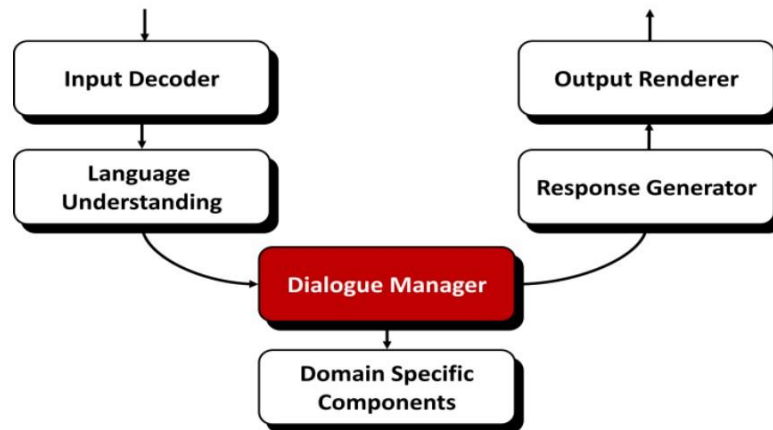
¹⁷⁷ OpenAI is a research and development laboratory for tools based on artificial intelligence, with a mission to offer services and tools for the benefit of the community.

¹⁷⁸ Cf. Aman Anad, Deep Learning Trends: top 20 best uses of GPT-3 by OpenAI, [Deep Learning Trends: top 20 best uses of GPT-3 by OpenAI \(educative.io\)](https://educative.io/deep-learning-trends-top-20-best-uses-of-gpt-3-by-openai/), September 20, 2020

¹⁷⁹ Amol Mavuduru, What is GTP-3 and why is it so powerful? [What is GPT-3 and why is it so powerful? | Towards Data Science](https://towardsdatascience.com/what-is-gpt-3-and-why-is-it-so-powerful/), February 17, 2021

2.1 Spoken Dialogue System

The spoken dialogue system is a computer capable of interacting with users through a natural language, to provide support to or let users access to various system functions. Its architecture is complex.¹⁸⁰



The diagram shows the various components of the Spoken Dialogue System. Source: Punjab Technical University, [1306.4134.pdf \(arxiv.org\)](#)

The first stage is the Input Decoder which receives the voice request and converts it into a written text.¹⁸¹ The second stage is understanding properly what users have said via the natural language understanding analyser.¹⁸² In the next step, the dialogue manager receives the semantic representation of the sentence and understands how the text can be adapted to the context of the input sentence, decides what response the user should receive, resolves linguistic ambiguities and stores the history of previously received sentences.¹⁸³ Occasionally, the sentence can be processed by an additional component performing dedicated tasks.

¹⁸⁰ Cf. Suket Arora, Kamaljeet Batra, Sarabjit Singh, Dialogue System: A Brief Review, Punjab University, p. 1

¹⁸¹ Matthew Zajeckowski, [Automatic Speech Recognition \(ASR\) Software - An Introduction - Usability Geek](#)

¹⁸² IBM Cloud Education, [What is Natural Language Processing? | IBM](#), July 2, 2020

¹⁸³ Ivi. Suket Arora, Kamaljeet Batra, Sarabjit Singh, Dialogue System: A Brief Review, Punjab University, p. 2.

Afterwards, the sentence is sent to the response generator via the dialogue manager, that refines the sentence and decides what information to include and what linguistic register is appropriate. In the last step, the speech generator has to convert the sentence into a voice message.¹⁸⁴

The spoken dialogue system can be implemented in three different ways: finite state-based system, frame-based system and agent-based system.

The finite state-based system guides the person through a pre-set sequence of steps.¹⁸⁵

The frame-based system has a default question path and the answers given by users will be entered into pre-made modules in order to understand which tasks the machine has to perform.¹⁸⁶

The agent-based system allows users to communicate freely as the digital agent is able to reason and argue its answers and remembers the conversation history.

2.2 Rules-based Machine Translation

In 1970, the first machine translator, Systran, was launched on the market and from that moment on, this type of translator became a fundamental resource of today's society.¹⁸⁷ There is a variety of approaches to machine translation.

The first one is rule-based machine translation, which requires both dictionaries and grammar texts of input and output languages. This

¹⁸⁴ Ibid. Dialogue System: A Brief Review, pp. 3-4

¹⁸⁵ Cf. Hagen Böhm, Dialogue management using Finite State Models, Saarland University, 2002, pp. 4-10

¹⁸⁶ Ivi. Cfr. Suket Arora, Kamaljeet Batra, Sarabjit Singh, Dialogue System: A Brief Review, Punjab University, p. 3

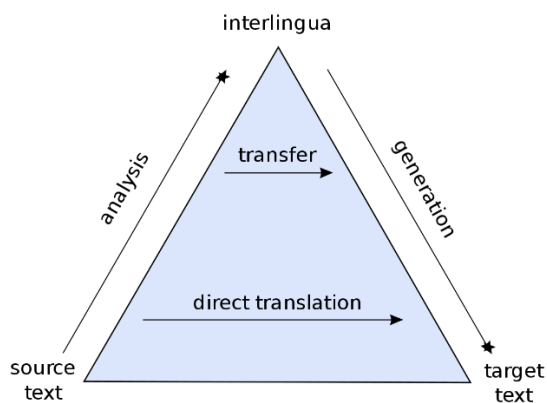
¹⁸⁷ Cf. History Section, [Rule-based machine translation - Wikipedia](#)

translator tries to compose the translated words to have a complete sentence through the morphology analysis.¹⁸⁸ This type of translation can be implemented in three different ways.

- Dictionary-based machine translation: each word of the input sentence will be translated through the use of a bilingual dictionary.¹⁸⁹

- Transfer-Based Machine Translation: firstly, the algorithm grammatically analyses the sentence, then tries to understand its context. After the analytical phase, the sentence is translated, the gender and number concordance are checked, and finally it is returned to the users.¹⁹⁰

- Interlingual machine translation: firstly, the input language is translated into a neutral language, and then translated into the target language. In this way, there is no need for bilingual dictionaries, as any language will be compared with English and then translated into the target language.¹⁹¹



The figure shows the pyramid of Bernard Vauquois, who explains how the three methods of rule-based machine translation work. Source: [1: Bernard Vauquois' pyramid | Download Scientific Diagram \(researchgate.net\)](#)

¹⁸⁸Cf. Kantan, Mt. RBMT vs. SMT, [RBMT vs SMT – KantanMT – Machine Learning – Neural Engines – AI – Language Technology \(kantanmtblog.com\)](#), February 04, 2021

¹⁸⁹ Cf. Agato Translation Company, WHAT IS DICTIONARY-BASED MACHINE TRANSLATION?, [What is Dictionary-Based Machine Translation? \(agatotranslate.ae\)](#)

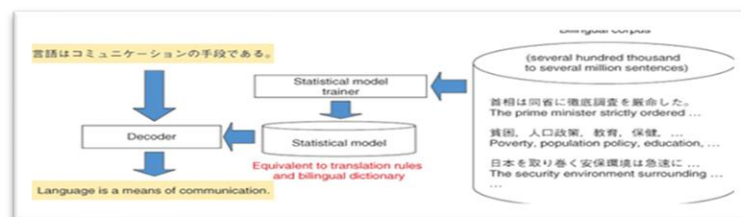
¹⁹⁰ Cf. Transfer-based machine translation, [Transfer-based machine translation - Wikipedia](#)

¹⁹¹Cf. R. Richsen, Interlingual Machine Translation, Silverchair, USA, pp. 145-147, [010144.pdf \(silverchair.com\)](#), August 20, 2021

2.3 Statistical Machine Translation (SMT)

In the mid-1980s, IBM's Thomas Watson research centre re-introduced Statistical Machine Translation (SMT), which became predominant until the beginning of the 21st century.¹⁹²

SMT avails itself of statistical algorithms to predict the best translation output, to process the most appropriate rules for producing the target sentence, and does not need the grammar rules of the languages to translate as training data. In practice, after inputting the text, the SMT will use statistical algorithms to process a series of translations based on training corpora, then the language decoder will choose the translation that has the highest probability to be the most suitable one for the situation in terms of syntax, semantics and context relevance. This approach requires a large quantity of linguistic corpora in both the input and output languages to deliver the most likely translation and to learn how to predict the best translation on future occasions.¹⁹³ The SMT experiences difficulty in translating a text, whose writing style and topic do not match that of the corpora, and translating colloquial expressions.¹⁹⁴



The diagram shows how the statistical machine translation algorithm works.

Source: [Recent Innovations in NTT's Statistical Machine Translation | NTT Technical Review \(ntt-review.jp\)](https://www.ntt-review.jp/column/201701/20170101_001.html)

¹⁹² Ilya Pestov, FreeCodeCamp, Statistical Translation, [A history of machine translation from the Cold War to deep learning \(freecodecamp.org\)](https://www.freecodecamp.org/news/a-history-of-machine-translation-from-the-cold-war-to-deep-learning/)

¹⁹³ Cf. Nabeel T. Alsohybe, Neama Abdulaziz Dahan and Fadl Mutaheer Ba-Alwi, Machine-Translation History and Evolution: Survey for Arabic-English Translations, p. 7, January 19, 2017.

¹⁹⁴ Cf. United Language Corp, THE PROS AND CONS OF STATISTICAL MACHINE TRANSLATION, [The Pros and Cons of Statistical Machine Translation \(unitedlanguagegroup.com\)](https://www.unitedlanguagegroup.com/blog/the-pros-and-cons-of-statistical-machine-translation/)

3. Historical background and overview of Natural Language Processing (NLP)

Natural Language Processing (NLP) is part of artificial intelligence and aims at giving computers the ability to communicate with humans through text and voice messages in the same way as humans would.¹⁹⁵

In the 1980s and 1990s, complex handwritten grammar rules were replaced by the first machine learning algorithms, such as decision trees. At the beginning of the 2000s, with the introduction of recurrent neural networks (RNN), it was possible to create machines that process voice and text which improved voice and messaging artificial assistants. Neural networks are still considered to be the most appropriate choice for creating a digital assistant. Apple's Siri is the first voice assistant developed with natural language processing algorithms and a neural network. Its automated speech recognition module transforms the user's words into digital signals to be interpreted and the voice command system links them to Siri's predefined actions.¹⁹⁶

Natural language processing improved machine translations and voice and messaging artificial assistants. This is achieved thanks to the four elements that constitute NLP: morphology, syntax, semantics and pragmatics, and phonology.¹⁹⁷

Part-of-Speech-Tagging (PoS) is the algorithm that deals with morphology analysis - finding the function of every word. Machines struggle to carry out this task, as well as understanding the ambiguity of

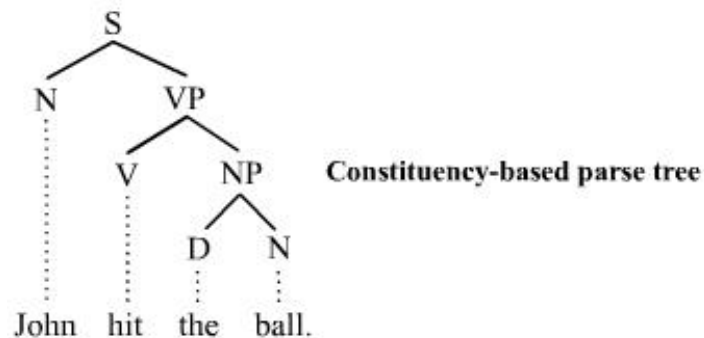
¹⁹⁵ Cf. IBM Cloud Education, Natural Language Processing (NLP), [What is Natural Language Processing? | IBM](#), July 2, 2020.

¹⁹⁶ Cf. Keith D. Foote, A Brief History of Natural Language Processing (NLP), [A Brief History of Natural Language Processing \(NLP\) - DATAVERSITY](#) May 22, 2019.

¹⁹⁷ Cf. Ionos, What is natural language processing?, [How does natural language processing work? - IONOS](#), April 09, 2019.

words, therefore to solve these challenges, PoS uses corpora to teach the NPL machine to distinguish words.¹⁹⁸

The syntax is managed by parse trees which have the task of understanding the meaning of the sentence, by breaking it down into parsing steps, to be analysed individually.¹⁹⁹



The figure shows how a syntactic tree divides the sentence in parsing steps. Sentence (S), Noun (N), Verbal Phrase (VP), Verb (V), Noun Phrase (NP), Determiner (D). Source: [Parse tree - Wikipedia](#)

The next stage is analysing the semantics and pragmatics and it is difficult for a machine to understand what meaning to attribute to a polysemic word. To solve this semantic problem, the machine analyses other words to understand the context with the help of corpora.²⁰⁰ The last step is phonology which is managed by the dialogue manager discussed in chapter 2.1.

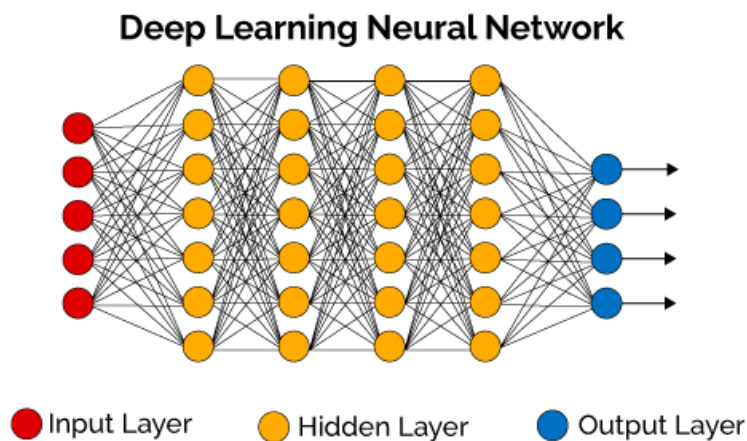
¹⁹⁸ Cf. Aiswarya Ramachandran, NLP Guide: Identifying Part of Speech Tags using Conditional Random Fields, [NLP Guide: Identifying Part of Speech Tags using Conditional Random Fields | by Aiswarya Ramachandran | Analytics Vidhya | Medium](#), October 05, 2018

¹⁹⁹ Ivi. Ionos, What is natural language processing?, [How does natural language processing work? - IONOS](#), April 09, 2019.

²⁰⁰ Ivi. Ionos, What is natural language processing?, [How does natural language processing work? - IONOS](#), April 09, 2019.

3.1 Deep Learning

Deep learning is a research branch of machine learning and enables computers to learn from the results of their past activities to provide better results for future tasks. Applications range from conversation between user and machine to voice and facial recognition.²⁰¹



This figure shows the neural network of a deep learning architecture. Source: [Neural Networks and Deep Learning – John Sisler](#)

In 1949, the first computer based on artificial neural networks was marketed. It was designed by Walter Pitts and Warren McCulloch²⁰² and consisted of a set of algorithms and mathematics called threshold logic to emulate human thought.²⁰³ In later years, deep learning models were considered too complex and the technology of the time could not support them, therefore most projects were shelved. In 1999, when computers began to process data via the graphics processing units (GPU), artificial neural networks were reconsidered again as the speed of computation

²⁰¹ Cf. Wikipedia, Definition, [Deep learning - Wikipedia](#)

²⁰² Walter Pitts and Warren McCulloch are two iconic figures in the field of neural learning because they authored the article "A Logical Calculus of Ideas Immanent in Nervous Activity" (1943), which set out the first mathematical model of an artificial neural network in history.

²⁰³ Cf. Keith D. Foote, A Brief History of Deep Learning, [A Brief History of Deep Learning - DATAVERSITY](#) February 07, 2017

had increased a thousandfold. Additional complex deep learning algorithms were developed with the ability to have long short-term memory - an architecture capable of classifying and processing multiple data simultaneously and providing predictions of the outcome based on previous events.²⁰⁴

In 2012, Google Brain developed a deep learning architecture that learns without supervised learning functions, that means receiving non-tagged data and forcing the computer to distinguish information autonomously.²⁰⁵

Concluding in 2019, Yoshua Bengio, Geoffrey Hinton, and Yann LeCun took deep learning to the next level, bringing improvements in computer vision and speech recognition, which paved the way for automated medical diagnosis and self-driving cars.²⁰⁶

3.1.1 How Deep Learning works

Deep learning aims to provide increasingly accurate results through a machine's self-improvement. Its artificial neural networks are designed to emulate human thinking and learning from past events.²⁰⁷

The implementation of Big Data analytics has improved how neural networks process data, and offers the possibility to perform predictive analysis on future events based on the analysis of past events. Big Data can extrapolate from a huge amount of data important

²⁰⁴ Cf. MLK, Brief History of Deep Learning from 1943-2019 [Timeline], [Brief History of Deep Learning from 1943-2019 \[Timeline\] | MLK - Machine Learning Knowledge](#), November 24, 2019

²⁰⁵ Cf. Wikipedia, Google Brain, [Google Brain - Wikipedia](#)

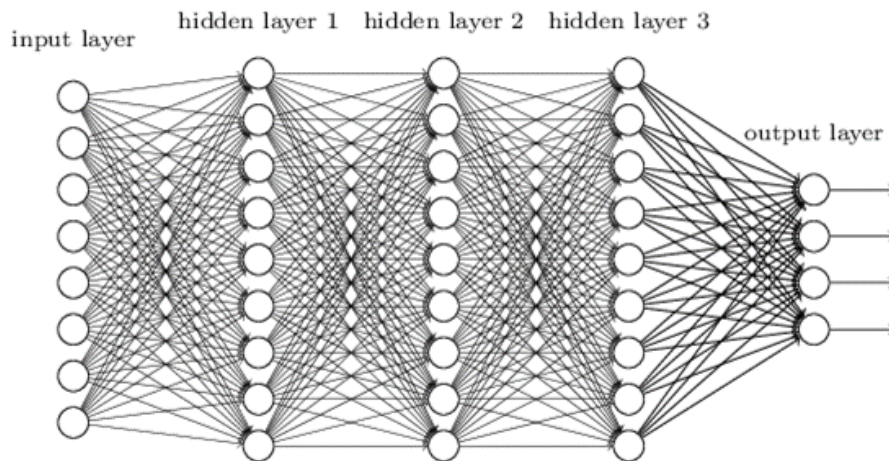
²⁰⁶ Cf. Varun Bansal, [The Evolution of Deep Learning | Towards Data Science](#), April 05, 2021

²⁰⁷ Cf. MathWorks, What is deep learning?, [What Is Deep Learning? | How It Works, Techniques & Applications - MATLAB & Simulink \(mathworks.com\)](#)

information for the company such as market trends and customer behaviour.²⁰⁸

Neural networks consist of three main layers: an input layer, one or more hidden layers and an output layer.²⁰⁹

Deep neural network



The figure portrays the architecture of an artificial deep neural network. There may be more than one hidden layer as required. Source: [Interpretability of Neural Networks | by Stephen Ho | Medium](#)

The artificial neural network is composed of thousands of layers containing millions of neurons that communicate with one another by exchanging thousands of signals. Each node is given a weight according to the data it is analysing, the heaviest node of each layer will influence the analysis of the next layer and the last layer will analyse all the data of the nodes, from previous layers, with the highest weight to provide the final result.²¹⁰ During data processing, neural networks classify information by answering a set of binary true/false questions, which are the result of complex mathematical calculations. In addition, as the

²⁰⁸ Cf. Simplilearn, What is Big Data Analytics and Why Is it Important?, [Big Data Analytics: Types, Tools and Applications \[Updated\] \(simplilearn.com\)](#), July 07, 2021

²⁰⁹ Cf. Ahmed Gad, Beginners Ask “How Many Hidden Layers/Neurons to Use in Artificial Neural Networks?”, [Beginners Ask “How Many Hidden Layers/Neurons to Use in Artificial Neural Networks?” | by Ahmed Gad | Towards Data Science](#), July 27, 2018

²¹⁰ Cf. David Fumo, Gentle introduction to neural network, [A Gentle Introduction To Neural Networks Series — Part 1 | by David Fumo | Towards Data Science](#), August 04, 2021

programme carries out tasks, it also performs autogenous training to increase the accuracy of its answers.

3.2 Neural Machine Translation

The Neuronal Machine Translator has surpassed all other types of translators for the accuracy in its translation and improved memory management.²¹¹

The neuronal machine translator can store data during time steps of the translation and handle sentences of varying lengths. This type of translator uses sequential data which must be stored during the various time steps of the process; therefore, it is equipped with two components built with a Long Short-Term Memory (LSTM), that is a type of Recurrent Neural Network (RNN). On the contrary, a simple artificial neural network is not able to save any sequential data.²¹²

The RNN consists of a connection of nodes performing a task by dividing it into steps over a period of time that allows the network to correct its errors while processing. Thanks to its internal state the RNN can store data strings, up to the length of a 10-line paragraph. To improve the storage size, Long Short-Term Memory (LSTM) was designed to avoid the vanishing gradient problem, thus it carries all data from the initial stages of the analysis to the final stages, by doing so if there are translation errors, they will not be lost but can be corrected.²¹³

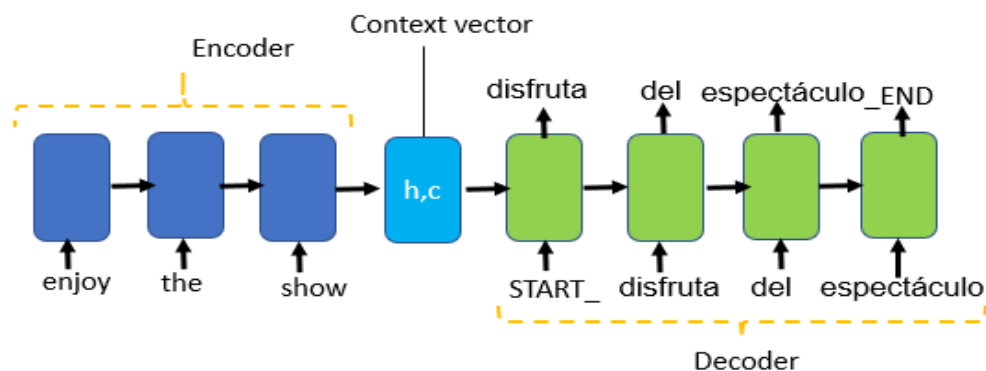
²¹¹ Cf. Sam Yip, TranslateFX, [What is Neural Machine Translation & How does it work?](https://www.translatefx.com) ([translatefx.com](https://www.translatefx.com)) April 18, 2021

²¹² Cf. [Intuitive explanation of Neural Machine Translation | by Renu Khandelwal | Towards Data Science](#)

²¹³ Ivi, [Intuitive explanation of Neural Machine Translation | by Renu Khandelwal | Towards Data Science](#)

The Long Short-Term Memory can achieve this because it has filtering gates, which allow it to decide what information to save and carry to the next series of analysis until the final decisional stage.²¹⁴

The neural machine translator is composed of the Seq2Seq model²¹⁵ which transforms the source sequence into the target sequence by means of two LSMTs called encoder and decoder. The encoder decipheres the words of the input language and inserts them into a value vector called the thought vector. The latter analyses the words and derives their meaning. Afterwards, the encoder returns results that are partly discarded in order to save only those that are deemed correct. The decoder receives the results and will analyse them again. It will assign each string a weight and the one with the highest weight will be the translation to output.²¹⁶



The figure portrays how a neural machine translator translates the input sentence “Enjoy the show”

Source: [Implementing neural machine translation using keras | by Renu Khandelwal | Towards Data Science](#)

²¹⁴ Cf. [Illustrated Guide to LSTM’s and GRU’s: A step by step explanation | by Michael Phi | Towards Data Science](#)

²¹⁵ Seq2Seq is a special type of Recurrent Neural Network, specialised in solving complex language problems such as translations and answering questions.

²¹⁶ Ibid. [Intuitive explanation of Neural Machine Translation | by Renu Khandelwal | Towards Data Science](#)

4. Examples of how NLP works in the area of voice communication between humans and machines

In the modern world, humans and computers can easily communicate via voice messages, yet the latter finds it difficult because communicating requires the combination of reasoning and physical domains.

Communication can be divided into 6 stages:²¹⁷

1. Thoughts are converted into words
2. The vocal cords generate sounds belonging to a natural language
3. The air transmits these sounds and vibrations to the listener's ears
4. The auditory nerves transmit the sound to the brain
5. The brain converts the sound into a natural language
6. The brain processes the sound to obtain the meaning of words.

The interdisciplinary field Automatic Speech Recognition (ASR) can enable machines to convey voice messages by splitting the communication in three steps: speech synthesis, speech recognition and speech understanding.

4.1 Automatic Speech Recognition (ASR)

In 1952, Bell Labs created the first voice recognition system, that could recognise numbers from 0 to 10. In 1962, IBM launched the voice-activated machine "Shoebbox", which was capable of recognising 16 words and was used in switchboards to ask users which option they wanted to choose via the telephone keypad. In 1980, IBM created

²¹⁷ Cf. [Speech recognition is hard — Part 1 | by SAmin | Towards Data Science](#) January 14, 2019

Tangora: a voice-activated typewriter, with a vocabulary of 20,000 words, that helped people with visual impairment.²¹⁸

In 1984, ACT Ltd releases the first laptop on the market with a built-in voice recognition system, with a vocabulary of 4096 words. Users can activate every computer functionality with their voice.

Towards the beginning of the 2000s, IBM created Home Page Reader, which can read entire Internet pages to help people with visual impairments. In 2011, Apple released the first versatile voice assistant "Siri", which can activate iPhone functionalities via voice commands. In 2017, Google updated its translator, making it the first software able to record the users' voice, in more than 100 languages, and translate it.²¹⁹

4.2 Problems and functions of automatic speech recognition (ASR)

In verbal communication machines face manifold challenges.²²⁰

- Humans have two ears that pick-up sounds, whereas a computer should have a microphone for incoming sound waves and a device to convert them into digital signals.

- When speaking in a public place, humans need to separate words from surrounding sounds. The computer has to do the same.

- Sometimes people speak very quickly and this leads to a difficult understanding of punctuation. A machine should understand when each word and sentence begin and end.

²¹⁸ Cf. IBM Cloud Education, [What is Speech Recognition? | IBM](#) September 2, 2020

²¹⁹ Cf. [Speech recognition - Wikipedia](#)

²²⁰ Cf. S. Amin, Speech recognition is hard, Towards Data Science, [Speech recognition is hard — Part 1 | by SAmin | Towards Data Science](#) June 12, 2021

- Every person speaks differently, depending on their social status, gender, age and origin. The computer should manage the variability of speech, and recognise the various accents.

- Homophones exist and the machine should be able to understand what they mean and what function they have, according to the context.

- Finally, the machine must have complete knowledge of syntax, the context, the semantics of both input and output language.

To respond to these above-mentioned challenges, the ASR is designed with a complex architecture.²²¹

- The sound receiver is responsible for transforming sound waves into digital signals.

- The voice selector scans the digital signal and isolates the part containing the speech and transforms it into an acoustic parameter.

- Training data consists of written documents and recorded voices.

- The acoustic apparatus breaks down the acoustic wave of speech and divides it into pieces and calculates the probability of which phonemes are most likely to be used.²²²

- The pronunciation module associates the various words of speech with their respective sounds.²²³

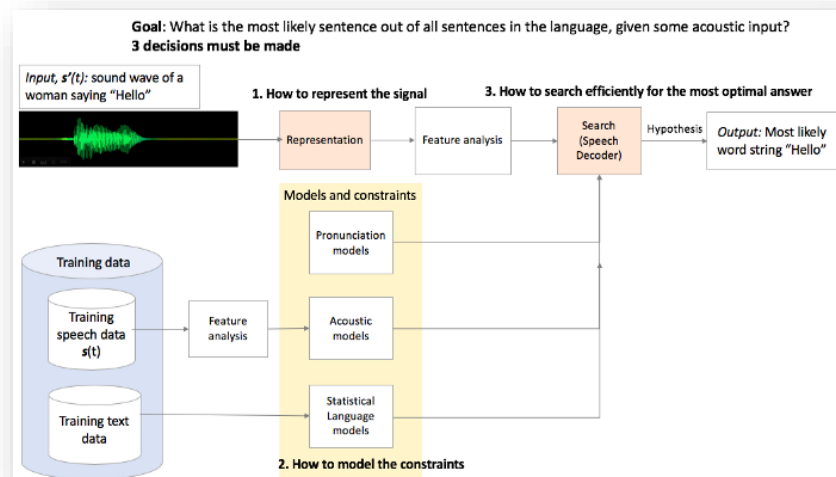
- The linguistic module predicts the most probable sequence that words must form to create a meaningful sentence.

- The decoder combines the phonetic and linguistic results to produce a most likely text string based on the input sentence.²²⁴

²²¹ Cf. Ketan Doshi, Audio Deep Learning Made Simple: Automatic Speech Recognition (ASR), How it Works, [Audio Deep Learning Made Simple: Automatic Speech Recognition \(ASR\), How it Works | by Ketan Doshi | Towards Data Science](#), March 25, 2021

²²² Cf. It Chronicles, Speech Recognition Algorithm, [Speech Recognition Algorithm - Brought to you by ITChronicles](#)

²²³ Ibid It Chronicles, Speech Recognition Algorithm, [Speech Recognition Algorithm - Brought to you by ITChronicles](#)



This diagram shows how the ASR works. Source: [Speech recognition is hard — Part 1 | by SAmin | Towards Data Science](#)

In order to have an accurate output, the most used algorithm is the Hidden Markov Model - a statistical model capable of calculating the probability of a set of future variables by analysing the current ones. According to Markov, knowing all the historical variables of the process is not necessary because once the probability of the future data has been established, the model will determine the best possible set, which is the one that comes closest to the sequence of current variables.²²⁵

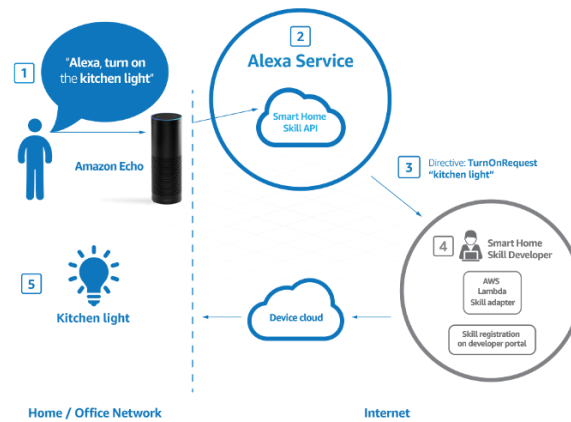
4.3 Digital assistant: combining artificial intelligence and linguistics

In 2013, Amazon launched Alexa, an artificial intelligence-based virtual assistant built into the Amazon speaker called Echo. Alexa quickly climbed the sales charts of home digital assistants worldwide. In 2020, 53.6 million Alexas were sold and in 2020, Alexa received

²²⁴ Ivi. S. Amin, Speech Recognition is hard, Towards Data, Science [Speech recognition is hard — Part 1 | by SAmin | Towards Data Science](#)

²²⁵ Cf. Sanjay Dorairaj, Hidden Markov Models Simplified, [Hidden Markov Models Simplified. Sanjay Dorairaj | by Sanjay Dorairaj | Medium](#)

250,000 marriage proposals. In 2021, 70% of digital assistants in US homes are branded Amazon.²²⁶



The figure shows how the communication between Alexa and users takes place.

Source: [How Amazon Alexa works? Your guide to Natural Language Processing \(AI\) | by Alexandre Gonfalonieri | Towards Data Science](#)

Amazon uses natural language processing combined with automatic speech recognition to communicate with users.

Once the voice message is received, the Alexa Voice Service (AVS) server breaks down the voice command into small sounds that will be compared with those in the database. When comparing sounds, AVS tries to find the keyword on which it will focus its search. After finding the information, AVS will send the result to Alexa that will provide it to the user. If the command is ambiguous, Alexa will ask the user to confirm it by repeating a message that closely sounds to the voice command received, this method also serves to train Alexa to learn new information and recognise different words and accents.²²⁷ Furthermore, Alexa can learn new functions, as Amazon allows third parties to create apps to train it, so users can always enjoy new functions.

²²⁶ Cf. Smiljanic Stash, Amazon Alexa Statistics, Facts, and Trends, [Amazon Alexa Statistics, Facts, and Trends | PolicyAdvice](#), February 05, 2021

²²⁷ Cf. Alexandre Gonfalonieri, How Amazon Alexa Works?, Towards Data Science, [How Amazon Alexa works? Your guide to Natural Language Processing \(AI\) | by Alexandre Gonfalonieri | Towards Data Science](#), May 24, 2021

5. Advantages and disadvantages of Natural Language Processing

Thanks to the development of natural language processing (NLP), it has been possible to find a solution to the difficulty of managing the huge amount of training data. Machines can now analyse and organise documents, understand linguistic nuances, offer accurate translations that help people in everyday life. Companies can use NLP to screen CVs against specific criteria, without bias and reduce selection time. In the hospitality industry, NLP can help to facilitate customers to find information through a digital assistant and be directed to the most appropriate operator, in this way customers are more satisfied. Moreover, it simplifies the work for employees by reducing repetitive actions, that can slow them down and depress them, to ensure that productivity levels and morale remain high at all times.²²⁸

However, NLP presents considerable challenges.

Setting up an NLP architecture is not easy because the training data must be chosen to give both general and deep knowledge to machines on many different topics. Too little or too much heterogeneous data can reduce the effectiveness of NLP. Although NLP checks the context to understand the meaning of sentences, it still encounters problems in understanding the subtleties of human thought, for example irony and sarcasm used to express feelings and ideas.²²⁹

²²⁸ Cf. Capacity, What are the advantages of Natural Language Processing (NLP), [What are the advantages of Natural Language Processing \(NLP\)? | Capacity](#)

²²⁹ Cf. Sciforce, Medium, Biggest Open Problems in Natural Language Processing, [Biggest Open Problems in Natural Language Processing | by Sciforce | Sciforce | Medium](#) July 04, 2021

5.1 Machine learning and deep learning in comparison

Machine learning is one of the first algorithms providing effective translations, better analysing the users' voice in order to formulate a response, processing data and returning results in the shortest possible time.²³⁰

Its advantages are manifold.²³¹

- It can examine a large amount of data and easily process trends and habits of people.
- It does not need human supervision because it is able to learn from past events and is capable of making predictions.
- Every time it performs a task it improves in efficiency and accuracy because it is able to learn from past events.

Although machine learning can process natural languages better than other algorithms, there are significant drawbacks.²³²

- A huge amount of non-homogeneous, labelled data is required to train the algorithm.
- It requires a huge use of financial resources.
- The algorithm works autonomously but is sensitive to errors and if it cannot correct them, it will repeat them endlessly.

Managing issues of machine learning have led to the development of a branch called deep learning. It is composed of a hierarchy of layers - usually 150 to 200²³³- which are made up of nodes that break down

²³⁰ Cf. Wikipedia, Machine Learning, [Machine learning - Wikipedia](#)

²³¹ Cfr. Data Flair, Advantages and Disadvantages of Machine Learning Language, [Advantages and Disadvantages of Machine Learning Language - DataFlair \(data-flair.training\)](#)

²³² Sciforce, Medium, Biggest Open Problems in Natural Language Processing, [Biggest Open Problems in Natural Language Processing | by Sciforce | Sciforce | Medium](#)

²³³ Cf. RfWireless-World, Advantages of Deep Learning, [Advantages of Deep Learning | disadvantages of Deep Learning \(rfwireless-world.com\)](#) , June 20, 2021

inputs into simpler concepts. The latter are assigned to each node, of the various layers, which after having analysed the concept, assigns a weight to it, the higher the weight, the higher are the chances that the data goes to the next layer until it becomes part of the result.

Deep learning has a number of advantages and improvements:

- The same neural network can perform different tasks and analyse different data - images, text, sound, video - without changing the internal structure.²³⁴

- Most companies have data in various formats - images, videos, documents - and the deep learning algorithm is designed to cross-reference and analyse data of different nature, whereas machine learning cannot.²³⁵

- The algorithm has a flexible structure in which new layers and nodes can be added to create new analysis and learning models.²³⁶

Although deep learning has many improvements compared to other artificial neural networks, it needs further refinements.

- It is difficult to understand how deep learning reasons and comes to a particular solution, as it is almost impossible to study each step of the processing of each node and layer. In fact, the reasoning of the neural network is the result of thousands of neurons arranged in interconnected layers and it is difficult to understand what reasoning they carry out.²³⁷

²³⁴ Cf. Valeryia Shchutskaya, Deep Learning: Strengths and Challenges, [Deep Learning: Strengths and Challenges – InData Labs Blog](#)

²³⁵ Cf. Alice Kinth, 32 Advantages and Disadvantages of Deep Learning [32 Advantages and Disadvantages of Deep Learning | by Alice Kinth | Medium](#)

²³⁶ Ivi. Valeryia Shchutskaya, Deep Learning: Strengths and Challenges, [Deep Learning: Strengths and Challenges – InData Labs Blog](#)

²³⁷ Cf. Aysegul Takimoglu, What is Deep Learning?, [What is Deep Learning? Explore Use Cases / Examples, Benefits \(aimultiple.com\)](#)

- The algorithm is capable of learning on its own how to perform various tasks, so it is difficult to understand how its reasoning evolves over time.²³⁸

- Although this algorithm is faster than the others, it requires a very high computing power of the graphic card during the training phase, which is the most expensive phase in terms of time and financial resources.²³⁹

²³⁸ Ivi Alice Kinth, 32 Advantages and Disadvantages of Deep Learning [32 Advantages and Disadvantages of Deep Learning | by Alice Kinth | Medium](#)

²³⁹ Cf. Stanford University, Drawbacks of Deep Learning, [Drawbacks of Deep Learning | MS&E 238 Blog \(stanford.edu\)](#)

6. The future of interaction between artificial intelligence and natural languages

There are two ways in which interaction between artificial intelligence and human language can take place in the future. On the one hand, life will be easier thanks to digital assistants performing various functions and help disadvantaged people. On the other hand, there are some concerns about how the way people communicate will change and how the labour market will be affected. By 2026, almost 80% of smartphone users will only use voice commands to communicate with their digital assistant, by 2023, 8 billion digital voice assistants will be in use and 25% of human-machine interactions in the workplace will be via voice commands.²⁴⁰

Two companies will soon launch their cutting-edge digital assistants on the market, which are capable of carrying out tasks ranging from managing IT devices to talking on the phone on the users' behalf.

Amazon developed Amazon Alexa For Business, that can organise a meeting calendar and manage audio visual equipment connected to a meeting room's WI-FI network via users' voice commands.²⁴¹

Google has developed two innovative products that will forever revolutionise our interaction with mobile phones and how we communicate with people.²⁴² The first product is Google Pixel 6 digital assistant that can read any text in the original language or in any another language. It creates subtitles for videos and during a call it transcribes in real time on a document what the interlocutor says to help people with

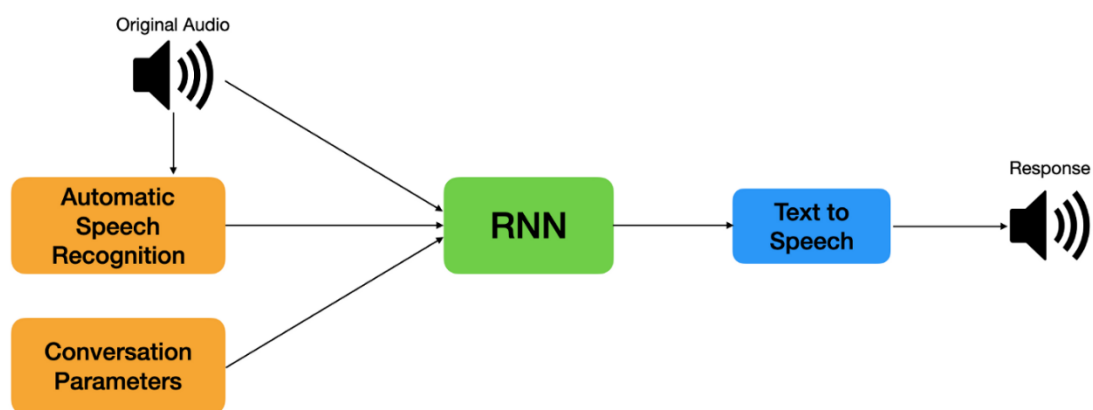
²⁴⁰ Cf. Adobe Press, 15 mind-blowing stats about voice assistants, [15 mind-blowing stats about voice assistants \(adobe.com\)](#)

²⁴¹ Cf. Eric Bieller, [What's The Future Of Human-Computer Interaction \(HCI\)? \(careerfoundry.com\)](#)

²⁴² Cf. TIMOTHY B. LEE, [Why Google believes machine learning is its future | Ars Technica](#)

hearing problems to talk over the phone. In addition, the users with their voice can access any application, browse the image gallery and write messages.²⁴³

The second product is Google Duplex. When users want to make a call, organise an event or make a reservation at a restaurant, they simply enter the information and the Google Assistant will start Duplex which will call and talk to the interlocutor. At the end of the call users will receive a summary of the call.²⁴⁴ Duplex uses natural language processing and recurrent neural network (RNN) to translate and write what the speaker says and formulate response sentences. Automatic Speech Recognition (ASR) is connected to the RNN and recognises the voice and contextualises the speaker's sentences. The text-to-speech (TTS) receives the result from the RNN and converts it into spoken messages.²⁴⁵



The figure shows how Google Duplex functions

Source: [NLP in the Real World: Google Duplex | by Jerry Wei | Towards Data Science](#)

²⁴³ In this video, we can see Google Digital Assistant features. [Next Generation Google Assistant: Demo 2 at Google I/O 2019 - YouTube](#)

²⁴⁴ Cf. The video shows Duplex features, [Google's AI Assistant Can Now Make Real Phone Calls - YouTube](#)

²⁴⁵ Cf. Jerry Wei, NPL in the Real World, Towards Data Science, [NLP in the Real World: Google Duplex | by Jerry Wei | Towards Data Science](#)

6.1 Perplexities and concerns about an artificial intelligence-based future

Although society can benefit from artificial intelligence because it makes life easier, especially for people from disadvantaged groups, citizens and national bodies are raising concerns.

Many US citizens asked the government whether it is morally acceptable that a digital assistant talks like a human without the interlocutor's knowledge. In addition, many wonder whether ill-intentioned individuals use the Google Assistant to call people in order to steal their bank account details and resell them.²⁴⁶

The US Senator Chris Coons raised awareness about Amazon's privacy policy related to the domestic use of Alexa. Amazon explained that they check voice messages and the transcripts of users' voice recordings which are stored on Amazon's servers indefinitely.²⁴⁷ Many economists and scientists, such as Paul Romer (Nobel Prize winner) and Nick Bloom (professor at Stanford University), are concerned about the phenomenon of technological unemployment. There is an increasing cutback in employment in various sectors such as manufacturing and agriculture. Moreover, office employees have been reduced by 30% in the US because computers are now able to analyse resumés and screen candidates for jobs faster and without bias.²⁴⁸

Stephen Hawking, Nobel Prize in Physics, described artificial intelligence as the evil of society, if national governments in

²⁴⁶ Cf. Chris Butler, Is Google Duplex ethical and moral?, Towards Data Science [Is Google Duplex ethical and moral? | by Chris Butler | Towards Data Science](#)

²⁴⁷ Cf. Chris Coons, [Amazon responds to Sen. Coons' concerns about consumer privacy practices for Alexa devices \(senate.gov\)](#)

²⁴⁸ Eric Rosenbau, IBM artificial intelligence can predict with 95% accuracy which workers are about to quit their jobs, <https://www.cnbc.com/2019/04/03/ibm-ai-can-predict-with-95-percent-accuracy-which-employees-will-quit.html>

collaboration with the European Union did not pass laws to direct artificial intelligence towards humanitarian purposes and the improvement of living standards.²⁴⁹

Concluding, artificial intelligence and natural language processing improve societies, for example, people with hearing impairments can speak freely on the phone and documents can be translated into various languages in seconds. However, there are still concerns about the best way to manage user data without violating privacy protocols and how to protect people from those who want to use artificial intelligence to harm others.

²⁴⁹ Cf. Arjun Kharpal, Stephen Hawking says A.I. could be ‘worst event in the history of our civilization’, <https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html>

7.Conclusion

The aim of this thesis is to show the intrinsic relationship between artificial intelligence and natural languages. The analytical study of the latter has been fundamental in the development of algorithms, whose analysis is close to human thought. NLP allows increasingly complex functions to be performed in a short time in order to optimise time and resources. In addition, this thesis highlights the complex hardware architecture, which makes it possible to perform operations such as translating, giving commands to a voice assistant to search for information, comparing data and assisting with work. These tasks, which may seem easy, require a thorough analysis of human language and communication, and the support of a well-trained neural network.

Continuous digitisation and improvements in natural language processing (NLP) have created jobs and made our life easier. People can communicate in different languages thanks to neural translators, documents can be analysed in seconds, and for those with hearing problems, digital assistants help them speak on their mobile phones.

However, I agree with the concerns about the misuse of artificial intelligence and the protection of sensitive user data.

In addition, there is the problem of unemployment caused by artificial intelligence, which has affected the secondary and tertiary sectors of our society. This phenomenon increases social inequality, as middle-class workers are replaced by computers, for instance companies employ fewer workers and prefer to utilise neural translators, digital assistants

organising administrators' agendas, and programmes capable of creating barrier tests for recruitment processes.²⁵⁰

Another aspect is the need to have laws to manage the use of artificial intelligence to avoid abuses and practices aimed at harming citizens. Moreover, we must also look at the bigger picture and point out that the reduction of jobs in some fields does not only lead to unemployment but paves the way to new jobs.

Concluding, the development of artificial intelligence and natural language processing should be seen as a way to progress and improve our lives. It is important to emphasise that each of us, international organisations and national governments utilise technology for the good of society. I hope that we can become more aware of the benefits that natural language processing offers and make good use of it.

²⁵⁰ Cf. United Nations Press, [Will robots and AI cause mass unemployment? Not necessarily, but they do bring other threats | United Nations](#)

Sección española

Introducción

La inteligencia artificial estudia como diseñar softwares capaces de emular los procesos intelectuales humanos, como el razonamiento y el aprendizaje a partir de la experiencia pasada. Estas capacidades les permiten de realizar tareas como traducir textos, proporcionar soporte al cliente y buscar información de todo tipo para uso empresarial y privado.



La imagen muestra el resultado de combinar la mente humana y las máquinas. Fuente: [Beginning Your Journey to Implementing Artificial Intelligence | SoftwareONE Blog](#)

La tesis examina en detalle el desarrollo cronológico y técnico de la relación entre la inteligencia artificial y las lenguas naturales, que son sistemas de sonidos y estructuras gramaticales que permiten a las personas interactuar entre sí.

En 1949, el Padre Busa²⁵¹ comenzó a digitalizar la obra de Tomás de Aquino, lo que dio lugar a la creación de la lingüística computacional. En los años 50 los estudios del lingüista Chomsky se convirtieron en la base de los algoritmos de codificación capaces de comprender las lenguas naturales.

Estudios posteriores permitieron a los investigadores construir redes neuronales artificiales capaces, por ejemplo, de asistir a los usuarios en

²⁵¹ Cfr. IBM, Gli albori della linguistica computazionale e la più grande opera a stampa mai prodotta al mondo, https://www.ibm.com/ibm/history/ibm100/it/it/stories/linguistica_computazionale.html

diversas tareas, traducir casi como lo harían los expertos traductores, analizar datos y estudiar el comportamiento de los usuarios.

1. Fundamentos de la inteligencia artificial y los intentos de crear máquinas capaces de entender las lenguas naturales

Las empresas informáticas siempre han intentado construir ordenadores capaces de interactuar con las personas, utilizando lenguas naturales para realizar funciones como traducir un documento, organizar una agenda de forma automatizada y asistir a un médico durante una operación. Estos procesos han sido posibles gracias a la lingüística computacional, que se centra en el desarrollo de sistemas capaces de analizar, interactuar y comprender una o varias lenguas naturales²⁵².

En 1946, el jesuita Roberto Busa, con la ayuda de los ordenadores de IBM, fue el pionero en utilizar un ordenador para que analice la obra *Omnia* de Tomás de Aquino. En 30 años de trabajo, creó el *Index Thomisticus* que consta de 70.000 páginas con un total de 20 millones de líneas²⁵³. En 1950, el matemático Alan Turing publicó en la revista académica *Mind* el artículo "Computing Machinery and Intelligence" en el que intentaba encontrar una respuesta a la pregunta "¿Puede pensar una máquina?"²⁵⁴ Para buscar la respuesta creó el famoso "test de Turing", donde dos jugadores, una persona y un ordenador, tienen que responder a las preguntas del interrogador, que debe identificar cuál de los dos jugadores es el ordenador en función de las respuestas recibidas. Ningún ordenador ha sido capaz de superar la prueba²⁵⁵.

En 1957, el lingüista Noam Chomsky, mientras que estaba estudiando el desarrollo del lenguaje natural en los seres humanos y la

²⁵² Cfr. Salvatore Sorace, *Introduzione alla linguistica computazionale*, [Microsoft PowerPoint - SEI1213_01_Linguistica_Computazionale.ppt \(unipa.it\)](#), p 5.

²⁵³ Cfr. Martina Dell'Utri, [International Busa Machines: come un gesuita ha insegnato la lettura ai computer | Controverso](#), 11/12/2020.

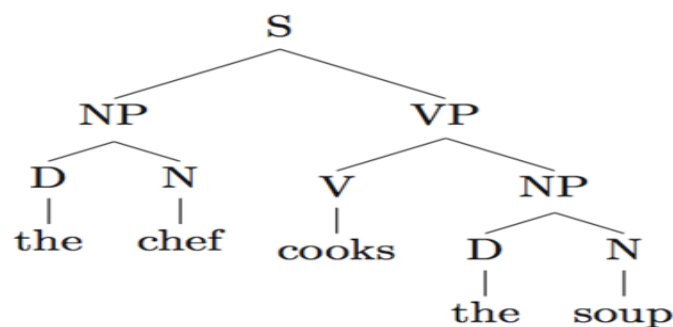
²⁵⁴ Cfr. Alan Turing, *Mind*, Volume LIX, Issue 236, October 1950, Oxford, Pages 434. [I.— COMPUTING MACHINERY AND INTELLIGENCE | Mind | Oxford Academic \(oup.com\)](#)

²⁵⁵ Ivi P. 436 - 439.

capacidad de los niños para comunicar sin educación formal, contribuyó a codificar algoritmos capaces de analizar y crear frases²⁵⁶. Chomsky desarrolló la gramática generativa, un conjunto de reglas que describen cómo formular oraciones en lenguas naturales, inspirándose en la gramática formal, centrándose en cómo el cerebro formula las oraciones y destacando la capacidad innata de los niños para formar y comprender oraciones que nunca han oído antes. Esta capacidad permite entender si una frase tiene sentido gramatical²⁵⁷.

Por ejemplo, los niños aprenden inconscientemente que se dice “puede hablar con ella” en lugar de, “puede hablar a ella”.

Chomsky realizó una investigación sobre un grupo de niños de hasta 10 años y observó que ellos aprenden que las palabras tienen un orden lógico y que cada parte tiene una tarea específica²⁵⁸. Esta competencia innata es necesaria para aprender cualquier lengua y Chomsky se basó en ella para crear la teoría de la gramática universal, que proporciona un sistema para el aprendizaje de lenguas.



Esquema mental inconsciente que utilizan los niños cuando aprenden a hablar. Chomsky lo denominó diagrama de árbol de la gramática generativa. Oración (S); frase sustantiva (NP); frase verbal (VP); determinante (D), sustantivo (N); verbo (V). Fuente: [Parse tree - Wikipedia](#)

²⁵⁶ Cfr. Francesca Fiore, Il linguaggio universale secondo Noam Chomsky – Introduzione alla Psicologia, [Il linguaggio universale secondo Noam Chomsky - Psicologia \(stateofmind.it\)](#), 03/12/15.

²⁵⁷ Cfr. Generative Grammar, [Generative grammar | Britannica](#)

²⁵⁸ Cfr. [The Chomsky Experiments | by Thomas Roeper | The New York Review of Books \(nybooks.com\)](#) 16/03/1981

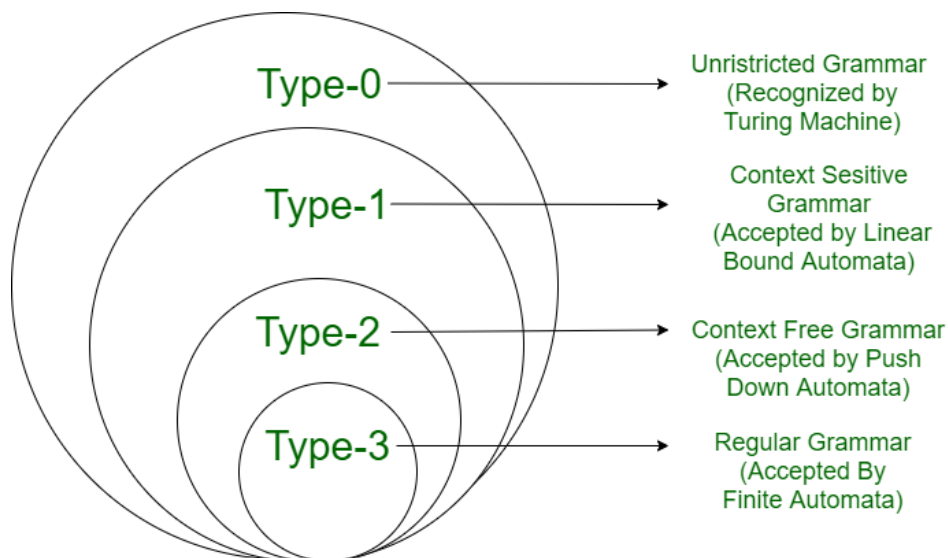
Chomsky organizó la gramática universal y las distintas gramáticas generativas en su jerarquía homónima formada por cuatro grupos²⁵⁹:

La gramática de tipo 0 engloba todas las gramáticas formales.

La gramática de tipo-1 es un tipo de gramática sensible al contexto.

La gramática de tipo-2 es libre de contexto y ofrece la posibilidad de desarrollar algoritmos para el análisis léxico de las frases.

La gramática de tipo 3 permite diseñar algoritmos para la búsqueda de palabras y frases en documentos, navegadores y diversos programas de escritura y traducción.



El gráfico muestra la jerarquía de Chomsky.

Fuente: [Chomsky Hierarchy in Theory of Computation - GeeksforGeeks](#)

Entre 1959 y 1963, John Backus y Peter Naur, científicos de IBM, desarrollaron la "Notación de Backus-Naur", que establece las reglas formales que deben utilizarse para ilustrar claramente la sintaxis de un lenguaje informático y los protocolos de red²⁶⁰.

²⁵⁹ Cfr. James Rogers, Formal language theory: refining the Chomsky hierarchy, [Formal language theory: refining the Chomsky hierarchy | Philosophical Transactions of the Royal Society B: Biological Sciences \(royalsocietypublishing.org\)](#), pp 2-20, 19/07/2012

²⁶⁰ Cfr. Sapere, [Backus-Naur Form | Sapere.it](#)

En 1968, en el MIT (Instituto Tecnológico de Massachusetts), el profesor Marvin Minsky²⁶¹ estaba diseñando un ordenador que pudiera comunicar con las personas y mover objetos geométricos en un entorno virtual. Para desarrollar la parte de la comunicación, Minsky preguntó al estudiante de doctorado Terry Winograd²⁶² si quería, para su tesis, codificar un algoritmo para que el ordenador pudiera entender las instrucciones en lengua natural. Durante los dos años siguientes, con la combinación de estudios de sintaxis, semántica y método deductivo, Terry Winograd desarrolló SHRDLU, un algoritmo capaz de comunicar en inglés²⁶³.

En 1949, el ingeniero Warren Weaver²⁶⁴ escribió el vademécum titulado "Traducción", donde se explica cómo se puede codificar programas capaces de traducir sin supervisión humana. El proceso consistía en sustituir cada palabra del texto por la correspondiente en el otro idioma y utilizar la estadística aplicada para calcular la frecuencia con la que se utilizaba cada palabra en un contexto específico²⁶⁵.

En 1952, Yehoshua Bar-Hillel²⁶⁶, profesor del MIT, propuso la innovadora idea de dotar a los ordenadores de una enciclopedia universal²⁶⁷. IBM, el MIT, la Universidad de Brown y la Universidad de Oxford empezaron a crear corpus lingüísticos para formar a los traductores²⁶⁸. A partir de 1975, con la introducción del traductor

²⁶¹ Cfr. Michael Aaron Dennis, [Marvin Minsky | American scientist | Britannica](#), 5/08/2021

²⁶² Información sobre Terry Winograd. [Terry Winograd - Wikipedia](#)

²⁶³ Cfr. Terry Winograd, Procedures as a representation of data in a computer program for understanding natural language, MIT, January 1971, p 18 - 24. [AITR-235.pdf \(stanford.edu\)](#)

²⁶⁴ Warren Weaver fue un científico y matemático estadounidense, considerado el padre de la traducción automática.

²⁶⁵ Cfr. Warren Weaver, Translation, p 15 - 24, [weaver001.pdf \(stanford.edu\)](#)

²⁶⁶ Cfr. Kasher, Essay in memory of Yehoshua Bar-Hillel, Kasher, Boston, p 1-2.

²⁶⁷ Ivi. P 5-6.

²⁶⁸ Cfr. Brown University, Brown Corpus Manual For Use With Digital Computer, USA, 1964, pp 3-8. [Brown Corpus Manual \(uib.no\)](#)

Systran, la precisión de la traducción mejoró, aunque sólo en la aeronáutica²⁶⁹.

En la década de 2000, Google presentó su propio traductor que era similar al de IBM, pero con más corpus disponibles²⁷⁰ y en 2016, la compañía decidió utilizar la traducción automática neural porque es más rápida y precisa²⁷¹.

²⁶⁹ Cfr. The Lingua File, Systran: a brief history, [The Lingua File from TLF Translation: SYSTRAN: A Brief History of Machine Translation](#)

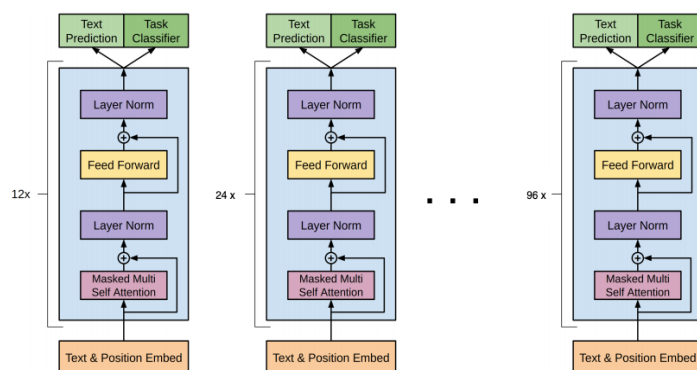
²⁷⁰ Cfr. History of Translation. [Google Translate - Wikipedia](#)

²⁷¹ Cfr. Joe Sommerlad, GOOGLE TRANSLATE: HOW DOES THE MULTILINGUAL INTERPRETER ACTUALLY WORK?, Marzo 2021. [Google Translate: How does the multilingual interpreter actually work? | The Independent | The Independent](#)

2. Modos utilizados por los ordenadores para analizar las lenguas humanas y la tecnología de vanguardia del GPT-3

En marzo de 2020, la empresa estadounidense OpenAI²⁷² ofreció al mercado informático el Generative Pre-trained Transformer 3, abreviado GPT-3.

GPT-3 es un modelo de lenguaje autorregresivo, que utiliza el aprendizaje neuronal profundo para crear textos de hasta 800 palabras, muy similares a un texto escrito por una persona, y que puede configurarse para otras funciones como generar cuestionarios, páginas web y recomendar libros según las preferencias del usuario²⁷³. Puede crear textos precisos porque está compuesto por 175.000 millones de coeficientes que se auto optimizan y que aprenden de los acontecimientos pasados mientras el software realiza el trabajo. La empresa OpenAI ha utilizado un enorme conjunto de corpus de varias fuentes – Wikipedia, Brown y Oxford University - y una gran colección de volúmenes como base de conocimiento para GPT-3²⁷⁴.



La imagen muestra el diagrama recursivo del software GPT-3. Fuente: [GPT-3 An Overview · All things \(dzlab.github.io\)](https://github.com/dzlab/gpt-3-overview)

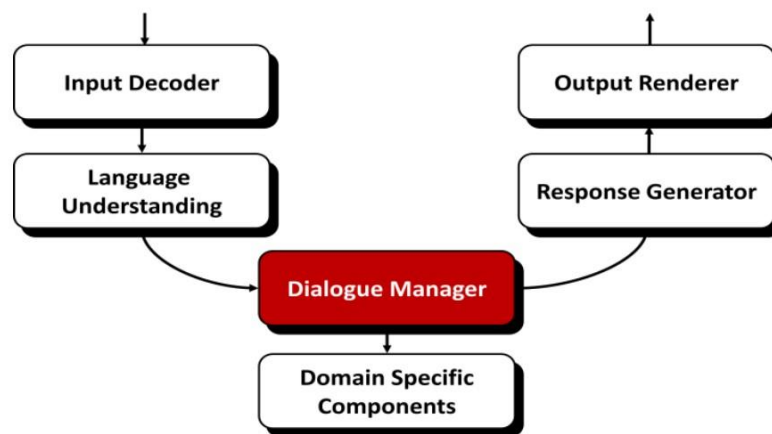
²⁷² OpenAI es una empresa de investigación y desarrollo de herramientas basadas en la inteligencia artificial, con la misión de proporcionar servicios y herramientas en beneficio de todo.

²⁷³ Cfr. Aman Anad, Deep Learning Trends: top 20 best uses of GPT-3 by OpenAI, [Deep Learning Trends: top 20 best uses of GPT-3 by OpenAI \(educative.io\)](https://www.educative.io/blog/top-20-best-uses-of-gpt-3-by-openai), 20/09/2020

²⁷⁴ Cfr. Amol Mavuduru, What is GTP-3 and why is it so powerful? [What is GPT-3 and why is it so powerful? | Towards Data Science](https://towardsdatascience.com/what-is-gpt-3-and-why-is-it-so-powerful/), 17/02/2021

2.1 Spoken Dialogue System

El spoken dialogue system es un sistema informático capaz de interactuar con los usuarios mediante una lengua natural, para dar soporte o permitir a los usuarios acceder a diversas funciones del sistema²⁷⁵.



El diagrama muestra los distintos componentes del Spoken Dialogue System. Fuente: Punjab Technical University, [1306.4134.pdf \(arxiv.org\)](https://arxiv.org/pdf/1306.4134.pdf)

La primera etapa es el *Input Decoder*, que recibe el comando de voz y lo convierte en un texto escrito²⁷⁶. La segunda etapa consiste en comprender correctamente lo que los usuarios han dicho a través del analizador de comprensión de la lengua natural²⁷⁷. En la siguiente etapa, *dialogue manager* recibe la representación semántica de la frase y entiende cómo se puede adaptar el texto al contexto general de la frase de entrada, decide qué respuesta debe recibir el usuario y resuelve las ambigüedades lingüísticas²⁷⁸. Tras ser analizada por el dialogue manager, la frase puede ser procesada por un componente adicional que

²⁷⁵ Cfr. Suket Arora, Kamaljeet Batra, Sarabjit Singh, Dialogue System: A Brief Review, Punjab University, p 1

²⁷⁶ Cfr. Matthew Zajechowski, [Automatic Speech Recognition \(ASR\) Software - An Introduction - Usability Geek](https://www.usabilitygeek.com/automatic-speech-recognition-asr-software-an-introduction/)

²⁷⁷ Cfr. IBM Cloud Education, [What is Natural Language Processing? | IBM](https://www.ibm.com/cloud/learn/natural-language-processing/), 2/07/2020

²⁷⁸ Ivi. Suket Arora, Kamaljeet Batra, Sarabjit Singh, Dialogue System: A Brief Review, Punjab University, p 2.

realiza tareas específicas como la traducción y la comparación. Después, la frase se envía al generador de respuestas a través del *dialogue manager*, que refina la frase y decide qué información incluir y qué registro lingüístico usar. En el último paso, el generador de voz tiene que convertir la frase en un mensaje de voz²⁷⁹.

El *spoken dialogue system* se puede implementar de tres maneras diferentes: *finite state-based system*, *frame-based system* and *agent-based system*.

El *finite state-based system* guía a la persona a través de una secuencia preestablecida de pasos²⁸⁰.

El *frame-based system* tiene una ruta de preguntas predeterminada y las respuestas dadas por los usuarios se introducirán en varios módulos para entender qué tareas la máquina tiene que realizar²⁸¹.

El *agent-based system* permite a los usuarios comunicarse libremente, ya que el agente digital es capaz de razonar y argumentar sus respuestas.

2.2 Rules-based Machine Translation

En 1970 se lanzó al mercado el primer traductor automático, Systran, y a partir de ese momento este tipo de traductor se convirtió en un recurso fundamental de la sociedad²⁸². Hay diferentes enfoques para la traducción automática.

La primera es la *rule-based machine translation*, que requiere diccionarios y textos gramaticales de las lenguas de entrada y salida.

²⁷⁹ Ibid. Dialogue System: A Brief Review, p 3-4

²⁸⁰Cfr. Hagen Böhm, Dialogue management using Finite State Models, Saarland University, 2002, pp. 4-10

²⁸¹ Ivi. Cfr. Suket Arora, Kamaljeet Batra, Sarabjit Singh, Dialogue System: A Brief Review, Punjab University, p 3

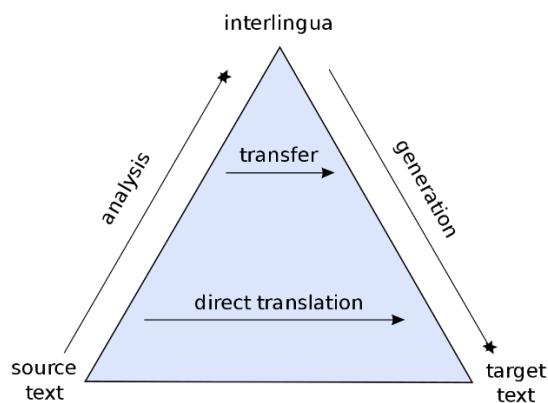
²⁸² Cfr. History Section, [Rule-based machine translation - Wikipedia](#)

Este traductor intenta componer las palabras traducidas para tener una frase completa mediante el análisis morfológico²⁸³. Este tipo de traducción puede implementarse de tres maneras diferentes.

- *Dictionary-based machine translation*: cada palabra de la frase de entrada se traducirá mediante el uso de un diccionario bilingüe²⁸⁴.

- *Transfer-Based Machine Translation*: en primer lugar, el algoritmo analiza gramaticalmente la frase y luego intenta comprender su contexto. Tras la fase de análisis, se traduce la frase, se comprueba la concordancia de género y número y, finalmente, los usuarios reciben la oración²⁸⁵.

- *Interlingual machine translation*: en primer lugar, la lengua de entrada se traduce a una lengua neutra, que suele ser el inglés, y luego se traduce a la lengua de destino. De este modo, no se necesitan diccionarios bilingües, ya que cualquier lengua se comparará con el inglés y luego se traducirá a la lengua de salida²⁸⁶.



La imagen muestra la pirámide de Bernard Vauquois, que explica cómo funcionan los tres métodos de rule-based machine translation. Fuente: [1: Bernard Vauquois' pyramid | Download Scientific Diagram \(researchgate.net\)](https://www.researchgate.net/publication/350144441)

²⁸³ Cfr. Kantan, Mt. RBMT vs SMT, [RBMT vs SMT – KantanMT – Machine Learning – Neural Engines – AI – Language Technology \(kantanmtblog.com\)](https://kantanmtblog.com/), 02/2021

²⁸⁴ Cfr. Agato Translation Company, WHAT IS DICTIONARY-BASED MACHINE TRANSLATION?, [What is Dictionary-Based Machine Translation? \(agatotranslate.ae\)](https://agatotranslate.ae/)

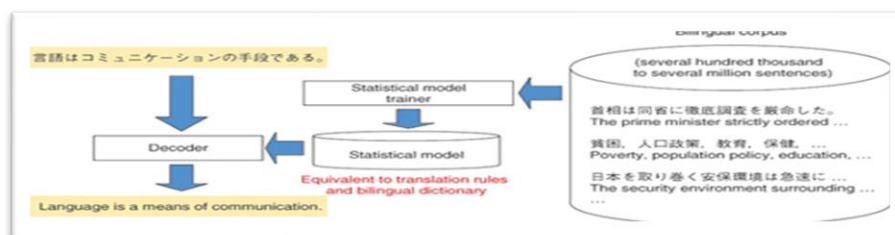
²⁸⁵ Cfr. Transfer-based machine translation, [Transfer-based machine translation - Wikipedia](https://en.wikipedia.org/wiki/Transfer-based_machine_translation)

²⁸⁶ Cfr. R. Richsen, Interlingual Machine Translation, Silverchair, USA, pp 145-147, [010144.pdf \(silverchair.com\)](https://www.silverchair.com/010144.pdf) 20/08/2021

2.3 Traducción automática estadística

A mediados de los años 80, IBM reintrodujo la traducción automática estadística (SMT), que fue predominante hasta principios del siglo 21²⁸⁷.

La SMT se sirve de algoritmos estadísticos para predecir el mejor resultado de la traducción y no necesita las reglas gramaticales de las lenguas a traducir como datos de entrenamiento. Tras introducir el texto, la SMT utilizará algoritmos estadísticos para procesar una serie de traducciones basadas en corpus de entrenamiento, y luego el decodificador lingüístico elegirá la traducción que tenga la mayor probabilidad de ser la más adecuada para la situación. Este enfoque requiere una gran cantidad de corpus lingüísticos en la lengua de entrada y en la de salida para obtener la traducción más probable y aprender a predecir la mejor traducción en futuras ocasiones. Además, para garantizar un resultado de calidad, los corpus lingüísticos deben estar especializados en un tema²⁸⁸ y el traductor tiene dificultades al traducir un texto cuyo estilo de escritura y tema no coinciden con los de los corpus, y al traducir expresiones coloquiales²⁸⁹.



La imagen muestra cómo funciona el algoritmo de traducción automática estadística. Fuente: [Recent Innovations in NTT's Statistical Machine Translation | NTT Technical Review \(ntt-review.jp\)](#)

²⁸⁷ Cfr. Ilya Pestov, FreeCodeCamp, Statistical Translation, [A history of machine translation from the Cold War to deep learning \(freecodecamp.org\)](#)

²⁸⁸ Cfr. Nabeel T. Alsohybe, Neama Abdulaziz Dahan and Fadi Mutaheer Ba-Alwi, Machine-Translation History and Evolution: Survey for Arabic-English Translations, pag 7, 19/01/2017.

²⁸⁹ Cfr. United Language Corp, THE PROS AND CONS OF STATISTICAL MACHINE TRANSLATION, [The Pros and Cons of Statistical Machine Translation \(unitedlanguagegroup.com\)](#)

3. Breve historia y descripción general del algoritmo de Procesamiento de Lenguaje Natural (PLN)

El Procesamiento del Lenguaje Natural (PLN) tiene como objetivo dotar a los ordenadores de la capacidad de comunicarse con los seres humanos a través de mensajes de texto y de voz de una manera similar a la humana²⁹⁰.

En los años 80 y 90, las reglas gramaticales escritas a mano fueron sustituidas por los primeros algoritmos de aprendizaje automático. A principios de la década de 2000, con la introducción de las redes neuronales recurrentes (RNN), fue posible crear máquinas que procesan voz y texto y esto mejoró los asistentes artificiales. Las redes neuronales siguen considerándose la opción más adecuada para crear un asistente digital. Siri, de Apple, es el primer asistente de voz desarrollado con un algoritmo de procesamiento del lenguaje natural y una red neuronal. Su módulo de reconocimiento de voz automatizado transforma las palabras del usuario en señales digitales que se interpretan y el sistema de comandos de voz las vincula a las acciones predefinidas de Siri²⁹¹.

Para llevar a cabo estas tareas, el proceso de PNL se compone de cuatro pasos: morfología; sintaxis; semántica y pragmática, y fonología²⁹².

El *Part-of-Speech-Tagging* (PoS) es el algoritmo que se ocupa del análisis morfológico: encontrar la función de cada palabra. Las máquinas tienen dificultades para realizar esta tarea, así como para entender la ambigüedad de las palabras, por ello para resolver estos retos PoS utiliza

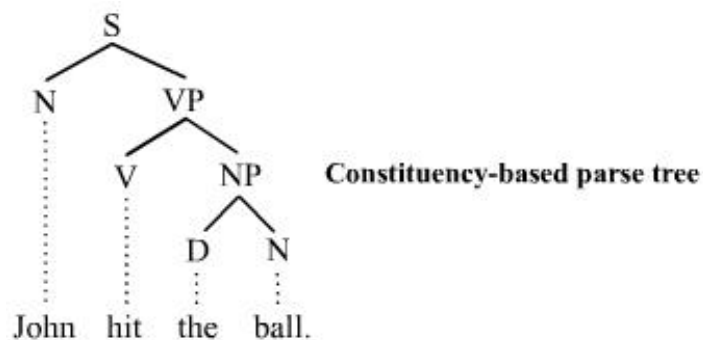
²⁹⁰ Cfr. Ibm Cloud Education, Natural Language Processing (NLP), [What is Natural Language Processing? | IBM](#), 2/07/2020.

²⁹¹ Cfr. Keith D. Foote, A Brief History of Natural Language Processing (NLP), [A Brief History of Natural Language Processing \(NLP\) - DATAVERSITY](#) 22/05/2019.

²⁹² Cfr. Ionos, What is natural language processing?, [How does natural language processing work? - IONOS](#), 09/04/2019.

corpus para enseñar a la máquina NPL a distinguir las palabras. Por ejemplo, los corpus enseñan a las máquinas que un verbo precedido de un artículo se convierte en un sustantivo²⁹³.

La sintaxis se gestiona mediante los *parse trees* que tienen la tarea de comprender el significado de la frase, descomponiéndola en pasos de análisis sintáctico, que deben ser analizados singularmente²⁹⁴.



La imagen muestra cómo un árbol sintáctico divide la frase en pasos de análisis sintáctico. Oración (S), sustantivo (N), frase verbal (VP), verbo (V), frase nominal (NP), determinante (D). Fuente: [Parse tree - Wikipedia](#)

La siguiente etapa es analizar la semántica y el pragmatismo, es difícil para una máquina entender qué significado atribuir a una palabra polisémica y para resolver este problema semántico, la máquina analiza otras palabras de la frase para entender el contexto²⁹⁵.

El último paso es la fonología, que es gestionada por dialogue manager tratado en el capítulo 2.1.

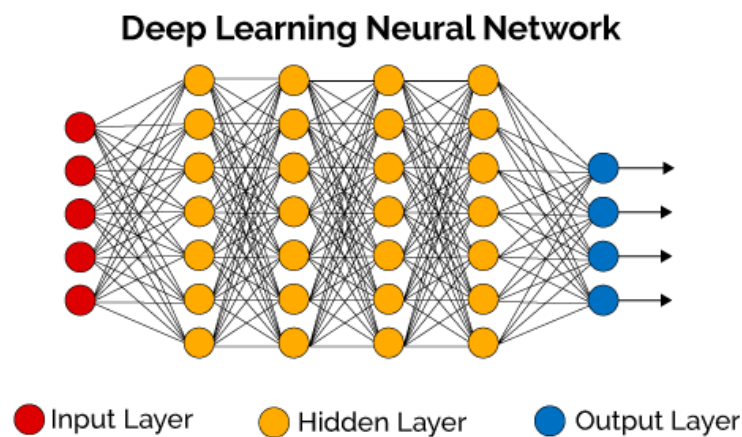
²⁹³ Cfr. Aiswarya Ramachandran, NLP Guide: Identifying Part of Speech Tags using Conditional Random Fields, [NLP Guide: Identifying Part of Speech Tags using Conditional Random Fields | by Aiswarya Ramachandran | Analytics Vidhya | Medium](#), 05/10/2018

²⁹⁴ Ivi. Ionos, What is natural language processing?, [How does natural language processing work? - IONOS](#), 09/04/2019.

²⁹⁵ Ibidem. Ionos, What is natural language processing?, [How does natural language processing work? - IONOS](#), 09/04/2019.

3.1 Aprendizaje profundo

El aprendizaje profundo (*deep learning*) es una rama de investigación del aprendizaje automático y permite a los ordenadores aprender de los resultados de sus actividades pasadas para ofrecer mejores resultados en tareas futuras. Las aplicaciones van desde la conversación entre el usuario y la máquina hasta el reconocimiento facial y de voz²⁹⁶.



Esta imagen muestra la red neuronal de una arquitectura de aprendizaje profundo. Fuente: [Neural Networks and Deep Learning – John Sisler](#)

En 1949 se comercializó el primer prototipo, un ordenador basado en redes neuronales humanas, diseñado por Walter Pitts y Warren McCulloch²⁹⁷ y que consistía en un conjunto de algoritmos y matemáticas llamado lógica umbral para emular el pensamiento humano²⁹⁸. En años posteriores, los modelos de aprendizaje profundo fueron archivados por ser demasiado complejos y la tecnología de la época no podía soportarlos.

²⁹⁶ Cfr. Wikipedia, Definición, [Deep learning - Wikipedia](#)

²⁹⁷ Walter Pitts y Warren McCulloch son dos figuras icónicas en el campo del aprendizaje neuronal porque escribieron el artículo "A Logical Calculus of Ideas Immanent in Nervous Activity" (1943), que describió el primer modelo matemático de una red neuronal artificial de la historia.

²⁹⁸ Cfr. Keith D. Foote, A Brief History of Deep Learning, [A Brief History of Deep Learning - DATAVERSITY](#) 07/02/2017

En 1999, cuando los ordenadores empezaron a procesar datos a través de las unidades de procesamiento gráfico (GPU), se replantearon los sistemas de redes neuronales. Se desarrollaron algoritmos de aprendizaje profundo más complejos con capacidad de *long short-term memory*, una arquitectura capaz de clasificar y procesar múltiples datos simultáneamente y proporcionar predicciones del resultado basadas en eventos anteriores²⁹⁹.

En 2012, Google Brain desarrolló una arquitectura de aprendizaje profundo que aprende sin recibir datos no etiquetados y obligando al ordenador a distinguir la información de forma autónoma³⁰⁰.

Concluyendo en 2019, Yoshua Bengio, Geoffrey Hinton y Yann LeCun mejoraron la visión por ordenador y el reconocimiento del habla³⁰¹.

3.1.1 Cómo funciona el aprendizaje profundo

El aprendizaje profundo proporciona resultados cada vez más precisos gracias a sus complejas redes neuronales, que están diseñadas para emular el pensamiento humano y aprender de eventos pasados³⁰². La implementación de la analítica de Big Data es crucial porque mejora el procesamiento de datos y extrae de una enorme cantidad de datos información clave para el negocio, como las tendencias del mercado, el

²⁹⁹ Cfr. MLK, Brief History of Deep Learning from 1943-2019 [Timeline], [Brief History of Deep Learning from 1943-2019 \[Timeline\] | MLK - Machine Learning Knowledge](#), 24/11/2019

³⁰⁰ Cfr. Wikipedia, Google Brain, [Google Brain - Wikipedia](#)

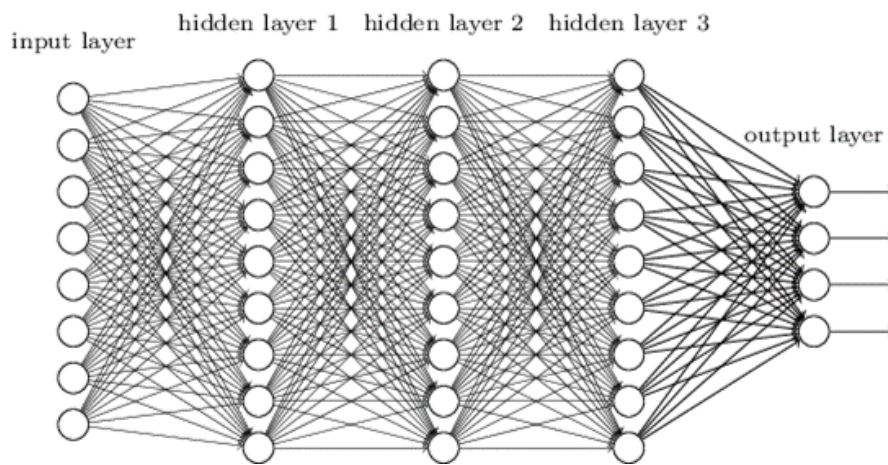
³⁰¹ Cfr. Varun Bansal, [The Evolution of Deep Learning | Towards Data Science](#), 05/04/2021

³⁰² Cfr. MathWorks, What is deep learning?, [What Is Deep Learning? | How It Works, Techniques & Applications - MATLAB & Simulink \(mathworks.com\)](#)

comportamiento de los usuarios y la evaluación de la calidad de los productos³⁰³.

Las redes neuronales constan de tres capas principales: capa de entrada, una o varias capas ocultas y capa de salida. Cada capa realiza tareas muy específicas -transcripción de textos, análisis de sonidos e imágenes- para administrar el aprendizaje profundo³⁰⁴.

Deep neural network



La imagen muestra la arquitectura de una red neuronal profunda. Fuente: [Interpretability of Neural Networks | by Stephen Ho | Medium](#)

La red neuronal artificial está compuesta por miles de capas que contienen millones de neuronas que se comunican entre sí intercambiando miles de señales. A cada nodo se le asigna un peso en función de los datos que analiza, el nodo más pesado de cada capa influirá en el análisis de la capa siguiente y la última capa analizará todos los datos de los nodos, procedentes de las capas anteriores, con el mayor peso para proporcionar el resultado final. Esta compleja estructura de

³⁰³ Simplilearn, What is Big Data Analytics and Why It is Important?, [Big Data Analytics: Types, Tools and Applications \[Updated\] \(simplilearn.com\)](#), 7/07/2021.

³⁰⁴ Cfr. Ahmed Gad, Beginners Ask “How Many Hidden Layers/Neurons to Use in Artificial Neural Networks?”, [Beginners Ask “How Many Hidden Layers/Neurons to Use in Artificial Neural Networks?” | by Ahmed Gad | Towards Data Science](#), 27/07/2018

capas y nodos necesita una potente arquitectura de hardware, ya que los cálculos de aprendizaje profundo pueden llevar días o semanas³⁰⁵.

La cantidad de datos necesarios para que un sistema de aprendizaje profundo proporcione un resultado preciso es enorme. Un traductor necesita millones de palabras etiquetadas individualmente y un número igual de frases etiquetadas para devolver una oración traducida que tenga sentido³⁰⁶.

3.2 Traducción automática neuronal

El Traductor Automático Neuronal utiliza una red neuronal artificial para traducir una frase de longitud variable. Este tipo de traductor ha eclipsado a todos los tipos de traductores por su precisión de traducción y su mejor gestión de la memoria³⁰⁷.

Las características de un traductor automático neuronal son: mantener los datos en las fases temporales de procesamiento y manejar frases de longitud variable. Para ello, requiere la ayuda de dos componentes con una arquitectura de *Recurrent Neural Network* (RNN) del tipo *Long Short Term Memory* (LSTM), que puede almacenar datos durante el proceso de traducción³⁰⁸.

Recurrent Neural Network (RNN) es una red de nodos que realizan la tarea en un marco temporal. Gracias a su memoria interna, puede procesar cadenas de datos de hasta 10 líneas. Sin embargo, para poder

³⁰⁵ Cfr. David Fumo, Gentle introduction to neural network, [A Gentle Introduction To Neural Networks Series — Part 1 | by David Fumo | Towards Data Science](#), 04/08/2021

³⁰⁶ Cfr. Sambit Mahapatra, Why Deep Learning over Traditional Machine Learning?, [Why Deep Learning over Traditional Machine Learning? | by Sambit Mahapatra | Towards Data Science](#), 07/07/2021

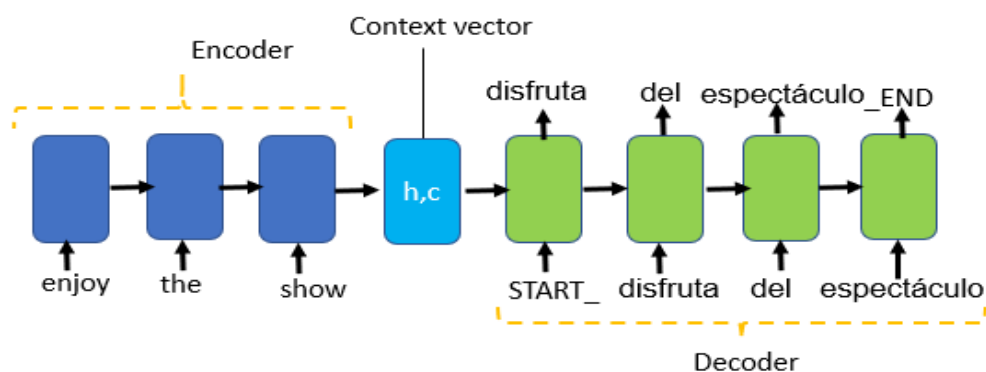
³⁰⁷ Cfr. Sam Yip, TranslateFX, [What is Neural Machine Translation & How does it work? \(translatefx.com\)](#) 18/04/2021

³⁰⁸ Cfr. Renu Khandelwal, Intuitive explanation of Neural Machine Translation, Towards Data Science, [Intuitive explanation of Neural Machine Translation | by Renu Khandelwal | Towards Data Science](#)

manejar textos más largos, se creó la *Long Short Term Memory* (LSTM), que lleva todos los datos desde las etapas iniciales del análisis hasta las etapas finales y, de esta manera, si hay errores de traducción, se pueden resolver³⁰⁹.

La *Long Short Term Memory* es capaz de hacer esto a través de puertos, que le permiten regular el flujo de datos y decidir qué información debe proceder en la siguiente serie de análisis. Esta característica es esencial para un traductor automático neural que maneja datos secuenciales³¹⁰.

En lo específico, el traductor automático neural utiliza el modelo Seq2Seq que transforma la secuencia de origen en la secuencia de destino utilizando dos LSMTs llamados codificador y decodificador. El codificador descifra las palabras de entrada de la lengua a traducir y las inserta en un vector que analiza las palabras, extrae su significado y lo inserta junto a las palabras analizadas. Después, el codificador devolverá resultados que se introducirán en el decodificador. Este último asignará a cada resultado un peso y lo que tenga mayor peso será la traducción³¹¹.



La figura muestra cómo un traductor automático neural traduce la frase entrante "Enjoy the show".

Fuente: [Implementing neural machine translation using keras | by Renu Khandelwal | Towards Data Science](#)

³⁰⁹ Cfr. [Recurrent neural network - Wikipedia](#)

³¹⁰ Cfr. Michael Phi, [Illustrated Guide to LSTM's and GRU's: A step by step explanation | by Michael Phi | Towards Data Science](#)

³¹¹ Ivi, Renu Khandelwal, [Intuitive explanation of Neural Machine Translation | by Renu Khandelwal | Towards Data Science](#)

4. Ejemplos de cómo funciona el procesamiento del lenguaje natural en la comunicación oral entre humanos y máquinas

En el mundo moderno, los humanos y los ordenadores pueden comunicarse fácilmente a través de mensajes de voz. Pero las computadoras les resulta difícil porque la comunicación requiere la combinación de los ámbitos del razonamiento y de la mecánica.

La comunicación puede dividirse en 6 etapas³¹²:

1. Los pensamientos se convierten en palabras;
2. Las cuerdas vocales generan sonidos pertenecientes a un lenguaje natural;
3. El aire transmite estos sonidos y vibraciones a los oídos del oyente;
4. Los nervios auditivos transmiten el sonido al cerebro;
5. El cerebro convierte el sonido en un lenguaje natural;
6. El cerebro procesa el sonido para obtener el significado de las palabras.

El campo interdisciplinar del Reconocimiento Automático del Habla (*Automatic Speech Recognition*, ASR) es capaz de permitir a las máquinas transmitir mensajes de voz dividiendo la comunicación en tres pasos: síntesis del habla, reconocimiento del habla y comprensión del habla.

³¹² Cfr. S.Amin, [Speech recognition is hard — Part 1 | by SAmin | Towards Data Science](#) 14/01/2019

4.1 Sistema de reconocimiento automático del habla (*Automatic Speech Recognition, ASR*)

En 1952, el Bell Labs creó Audrey, el primer sistema de reconocimiento de voz, y podía reconocer los números del 0 al 10. En 1962, IBM lanzó la máquina "Shoebbox" activada por voz, que podía reconocer 16 palabras y se utilizaba en las centralitas. En 1980, IBM creó Tangora: una máquina de escribir activada por voz con un vocabulario de 20.000 palabras y ayudaba a leer a las personas con discapacidad visual³¹³.

En 1984, ACT Ltd comercializó el primer ordenador portátil con un sistema de reconocimiento de voz integrado y con un vocabulario de 4.096 palabras. El usuario podía activar todas las funciones del ordenador con su voz.

A principios de la década de 2000, IBM creó el *Home Page Reader*, que podía leer páginas enteras de Internet de forma comprensible para las personas con deficiencias visuales. En 2011, Apple sorprende con el asistente de voz Siri; el usuario puede acceder a muchas funciones del teléfono móvil Iphone 4 mediante comandos de voz. Por ejemplo, puede llamar a una persona de la agenda, escribir y leer mensajes. En 2017, Google realiza una actualización para su Traductor que lo hace capaz de grabar la voz del usuario, en más de 100 idiomas, y traducirla³¹⁴.

³¹³ Cfr. IBM Cloud Education, [What is Speech Recognition? | IBM 2/09/2020](#)

³¹⁴ Cfr. History section, [Speech recognition - Wikipedia](#)

4.2 Problemas y funcionamiento del sistema de reconocimiento automático del habla

La comunicación verbal entre dos personas es muy compleja y los ordenadores se enfrentan a muchos retos³¹⁵:

- Los humanos tienen dos orejas que captan los sonidos, mientras que un ordenador debe tener un micrófono para las ondas sonoras entrantes y un dispositivo para convertirlas en señales digitales.

- Cuando se habla en un lugar público, hay que separar las palabras de los sonidos circundantes. El ordenador tiene que hacer lo mismo.

- A veces hablamos muy deprisa y esto complica la comprensión de la puntuación y cuando empieza y termina cada palabra. Una máquina debe entender cuándo empieza y termina cada palabra y cada frase.

- Cada persona habla de forma diferente, según su estatus social, su edad y su procedencia. El ordenador debe gestionar la variabilidad del habla, tener un amplio vocabulario y una base de datos de sonidos grabados para reconocer los distintos acentos e inflexiones.

- Los homófonos existen en las lenguas y la máquina debe ser capaz de entender qué significan y qué función tienen.

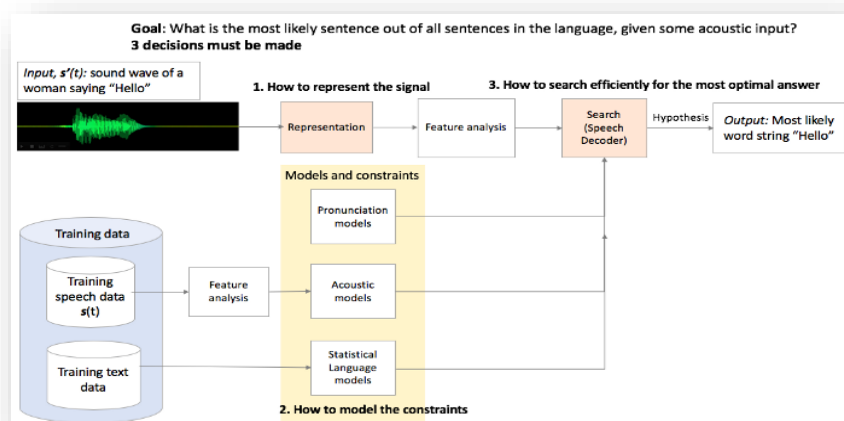
- Por último, el ordenador debe conocer la semántica de la lengua a traducir y la de destino, así como el contexto.

Para hacer frente a todos estos problemas, el sistema de reconocimiento automático del habla consta de una arquitectura compleja³¹⁶:

³¹⁵ Cfr. S. Amin, Speech recognition is hard, Towards Data Science, [Speech recognition is hard — Part 1 | by SAmin | Towards Data Science](#) 12/06/2021

³¹⁶ Cfr. Ketan Doshi, Audio Deep Learning Made Simple: Automatic Speech Recognition (ASR), How it Works, [Audio Deep Learning Made Simple: Automatic Speech Recognition \(ASR\). How it Works | by Ketan Doshi | Towards Data Science](#) , 25/03/2021

- El receptor de sonido se encarga de transformar las ondas sonoras en señales digitales.
- El selector de voz analiza la señal digital y aísla la parte que contiene el habla para transformarla en un parámetro acústico.
- La base de datos consiste en corpus de documentos escritos y voces grabadas para que la computadora pueda aprender a manejar un flujo de palabras uno tras otro.
- El aparato acústico descompone la onda acústica del habla y calcula la probabilidad de los fonemas más probables³¹⁷.
- El módulo de pronunciación asocia las distintas palabras del discurso con sus respectivos sonidos³¹⁸.
- El módulo lingüístico predice la secuencia más probable que deben formar las palabras para crear una frase con sentido.
- El descodificador combina los resultados fonéticos y lingüísticos para producir una cadena de texto más probable basada en la frase de entrada.



Este diagrama muestra cómo funciona el sistema de reconocimiento automático de voz.

Fuente: [Speech recognition is hard — Part 1 | by SAmin | Towards Data Science](#)

³¹⁷ Cfr. It Chronicles, Speech Recognition Algorithm, [Speech Recognition Algorithm - Brought to you by ITChronicles](#)

³¹⁸ Cfr. Samudravijaya K, ASR, <http://www.iitg.ac.in/samudravijaya/tutorials/asrTutorial.pdf>, p 9-10

El modelo oculto de Márkov es el algoritmo más utilizado para transcribir el sonido entrante y procesar la mejor respuesta. Se trata de un modelo estadístico capaz de calcular la probabilidad de un conjunto de variables futuras mediante el análisis de las presentes, sin tener en cuenta las variables analizadas en procesos anteriores³¹⁹.

El sistema de reconocimiento automático del habla se utiliza en diversos campos, como:

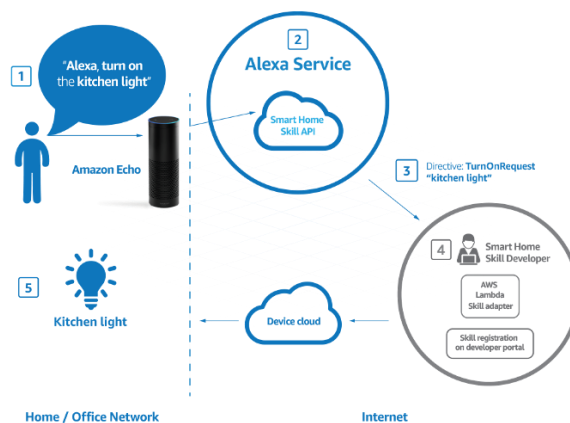
- En el sector del automóvil: con la voz el usuario puede encender todas las funciones del coche;
- En la sanidad: el médico puede dictar un informe con la ayuda de un asistente digital en lugar de escribirlo a mano;
- En telefonía: los asistentes de voz pueden ayudar a los usuarios a realizar búsquedas, marcar números de teléfono y realizar pagos.

4.3 Asistente virtual: combinación de inteligencia artificial y lingüística

En 2013, Amazon comercializó Alexa, un asistente virtual basado en inteligencia artificial, integrado en el altavoz Amazon Echo. Desde el principio, Alexa ha escalado las listas de ventas de asistentes digitales para el hogar en todo el mundo. En 2020 se vendieron 53,6 millones de Alexas y el asistente virtual recibió 250.000 propuestas de matrimonio. En 2021, el 70% de los asistentes digitales en los hogares estadounidenses son de la marca Amazon³²⁰.

³¹⁹Cfr. Sanjay Dorairaj, Hidden Markov Models Simplified, [Hidden Markov Models Simplified. Sanjay Dorairaj | by Sanjay Dorairaj | Medium](#)

³²⁰ Cfr. Smiljanic Stash, Amazon Alexa Statistics, Facts, and Trends, [Amazon Alexa Statistics, Facts, and Trends | PolicyAdvice](#), 05/02/2021



La imagen muestra cómo se produce la comunicación entre Alexa y los usuarios.

Fonte: [How Amazon Alexa works? Your guide to Natural Language Processing \(AI\) | by Alexandre Gonfalonieri | Towards Data Science](#)

Amazon utiliza el procesamiento del lenguaje natural combinado con el reconocimiento automático del habla para comunicarse con los usuarios.

Una vez recibido el mensaje de voz, el servidor de *Alexa Voice Service* (AVS) descompone el comando de voz en pequeños sonidos que serán comparados con los de la base de datos. Al comparar los sonidos, AVS intenta encontrar la palabra clave en la que centrará su búsqueda. Tras encontrarla, AVS enviará el resultado a Alexa, que se lo proporcionará al usuario. Si el comando es ambiguo, Alexa pedirá al usuario que lo confirme repitiendo un mensaje que se parezca mucho al comando de voz recibido, este método también sirve para entrenar a Alexa a aprender nueva información y reconocer diferentes palabras y acentos³²¹. Además, Alexa puede aprender nuevas funciones, ya que Amazon permite a terceros crear aplicaciones para entrenarla, así que los usuarios siempre pueden disfrutar con nuevas funciones.

³²¹ Cfr. Alexandre Gonfalonieri, How Amazon Alexa Works?, Towards Data Science, [How Amazon Alexa works? Your guide to Natural Language Processing \(AI\) | by Alexandre Gonfalonieri | Towards Data Science](#), 24/05/2021

5. Ventajas y problemas del procesamiento del lenguaje natural

Gracias al desarrollo del procesamiento del lenguaje natural (PLN), se ha podido encontrar una solución a la dificultad de gestionar la enorme cantidad de datos de entrenamiento. Las máquinas pueden ahora analizar y organizar los documentos, comprender los matices lingüísticos y ofrecer traducciones precisas que ayuden a las personas en la vida cotidiana. Las empresas pueden utilizar el PNL para filtrar los currículos en función de criterios específicos, sin prejuicios y reduciendo el tiempo de selección. En el sector de la hostelería, el PNL puede ayudar a facilitar a los clientes la búsqueda de información a través de un asistente digital y ser dirigidos al operador más adecuado, de esta forma los clientes quedan más satisfechos. Además, simplifica el trabajo de los empleados al reducir las acciones repetitivas, que pueden ralentizarlos y deprimirlos, para garantizar que los niveles de productividad y la moral se mantengan altos en todo momento³²².

Sin embargo, el PNL presenta retos considerables.

Crear una arquitectura de PNL no es fácil, ya que los datos de entrenamiento deben elegirse de forma que proporcionen a las máquinas conocimientos tanto generales como profundos sobre distintos temas. Unos datos demasiado poco o demasiado heterogéneos pueden reducir la eficacia del PNL. Aunque el PNL comprueba el contexto para entender el significado de las frases, sigue encontrando problemas para entender las sutilezas del pensamiento humano, por ejemplo, la ironía y el sarcasmo utilizados para expresar sentimientos e ideas³²³.

³²² Cfr. Capacity, What are the advantages of Natural Language Processing (NLP), [What are the advantages of Natural Language Processing \(NLP\)? | Capacity](#)

³²³ Cfr. Sciforce, Medium, Biggest Open Problems in Natural Language Processing, [Biggest Open Problems in Natural Language Processing | by Sciforce | Sciforce | Medium](#) 04/07/2021

5.1 Comparación entre el aprendizaje automático y el aprendizaje profundo

El aprendizaje automático permitió hacer más eficientes las traducciones, analizar mejor la voz del usuario para formular una respuesta, procesar los datos y devolver un resultado en el menor tiempo posible. Sus puntos fuertes son múltiples³²⁴:

- Puede examinar una gran cantidad de datos y procesar fácilmente las tendencias y los hábitos de las personas.
- No necesita supervisión humana porque es capaz de aprender de los acontecimientos pasados y es capaz de hacer predicciones.
- Cada vez que realiza una tarea mejora en eficiencia y precisión porque es capaz de aprender de los acontecimientos pasados.

Aunque el aprendizaje automático hace que las traducciones sean más precisas y mejora los asistentes digitales, tiene algunas desventajas no despreciables³²⁵:

- Para entrenar el algoritmo se necesita una gran cantidad de datos etiquetados no homogéneos.
- El algoritmo de aprendizaje automático requiere un enorme uso de recursos financieros y electricidad para funcionar.
- El algoritmo funciona de forma autónoma, pero es sensible a los errores y si no los corrige, los repetirá indefinidamente y sin una intervención temprana será difícil detectar el error.

Los problemas mencionados anteriormente llevaron al desarrollo del aprendizaje profundo.

³²⁴ Cfr. Data Flair, [Advantages and Disadvantages of Machine Learning Language - DataFlair \(dataflair.training\)](https://dataflair.training)

³²⁵ Cfr. Rf Wireless-World, [Advantages of Deep Learning | disadvantages of Deep Learning \(rfwireless-world.com\)](https://rfwireless-world.com)

Este algoritmo se compone de una jerarquía de capas -entre 150 y 200³²⁶ - formadas por nodos, que descomponen los datos entrantes en conceptos más simples y les asignan un peso. Cuanto mayor sea la ponderación, más pasarán los datos a la siguiente capa hasta formar parte del resultado. Por ejemplo, un traductor con este algoritmo descompone una frase en palabras y, tras la traducción de cada palabra, las ensambla para formar una frase completa³²⁷.

Hay muchas razones por las que empresas como IBM y Google han decidido implementar este algoritmo en sus productos³²⁸.

- La misma red neuronal puede realizar diferentes tareas y analizar distintos datos -imágenes, texto, sonido, vídeo- sin cambiar la estructura interna, como ocurre en una red neuronal con un algoritmo de aprendizaje automático.

- Este algoritmo no necesita ser diseñado cada vez para aprender a realizar una tarea específica, porque puede aprender por sí mismo.

- El algoritmo tiene una estructura flexible en la que se pueden añadir nuevas capas y nodos para crear nuevos modelos de análisis y aprendizaje.

Aunque el aprendizaje profundo tiene muchas mejoras con respecto a otras redes neuronales, hay desventajas³²⁹.

- Es difícil entender cómo este algoritmo con su red neuronal llega a una solución, ya que es casi imposible estudiar cada paso del procesamiento de cada nodo y capa. De hecho, el pensamiento de la red neuronal es el resultado de miles de neuronas dispuestas en capas

³²⁶ Cfr. Deep Learning, [Aprendizaje profundo - Wikipedia, la enciclopedia libre](#)

³²⁷ Cfr. InData Labs Blog, [Deep Learning: Strengths and Challenges – InData Labs Blog](#)

³²⁸ Cfr. Alice Kinth, 32 Advantages and Disadvantages of Deep Learning [32 Advantages and Disadvantages of Deep Learning | by Alice Kinth | Medium](#)

³²⁹ ³²⁹ Cfr. Valeryia Shchutskaya, Deep Learning: Strengths and Challenges [Deep Learning: Strengths and Challenges – InData Labs Blog](#)

interconectadas y es difícil entender qué razonamiento han hecho. Esta falta de transparencia en el razonamiento del algoritmo no da la posibilidad de predecir cuándo pueden producirse errores.

- El algoritmo es capaz de aprender por sí mismo cómo realizar diversas tareas, por lo que es difícil entender cómo evoluciona su razonamiento a lo largo del tiempo.

- Aunque este algoritmo es más rápido que los demás, requiere una potencia de cálculo muy elevada de la tarjeta gráfica durante la fase de entrenamiento, que es la más cara en términos de tiempo y recursos financieros.

6. El futuro de la interacción entre la inteligencia artificial y las lenguas naturales

El futuro de la interacción entre la inteligencia artificial y el lenguaje humano se divide en dos vías. Por un lado, la vida será más fácil gracias a los asistentes digitales que realizarán diversas funciones y ayudarán a las personas desfavorecidas. Por otro lado, preocupa cómo cambiará la forma de comunicarse de las personas y cómo se verá afectado el mercado laboral. Para 2026, casi el 80% de los usuarios de teléfonos inteligentes utilizarán únicamente comandos de voz para comunicarse con su asistente digital, para 2023 se utilizarán 8.000 millones de asistentes de voz digitales y el 25% de las interacciones entre humanos y máquinas en el lugar de trabajo se realizarán mediante comandos de voz³³⁰.

Dos empresas comercializarán sus vanguardistas asistentes digitales, capaces de realizar tareas que van desde la gestión de dispositivos informáticos hasta hablar por teléfono en nombre de los usuarios.

Amazon ha desarrollado *Amazon Alexa For Business*, que puede organizar un calendario de reuniones y gestionar los equipos audiovisuales conectados a la red WI-FI de una sala de reuniones a través de los comandos de voz de los usuarios³³¹.

Google ha desarrollado dos productos innovadores que revolucionan nuestra interacción con los teléfonos móviles y la forma de comunicarnos con la gente³³². El primer producto es el asistente digital Google Pixel 6, que puede leer cualquier texto en el idioma original o en

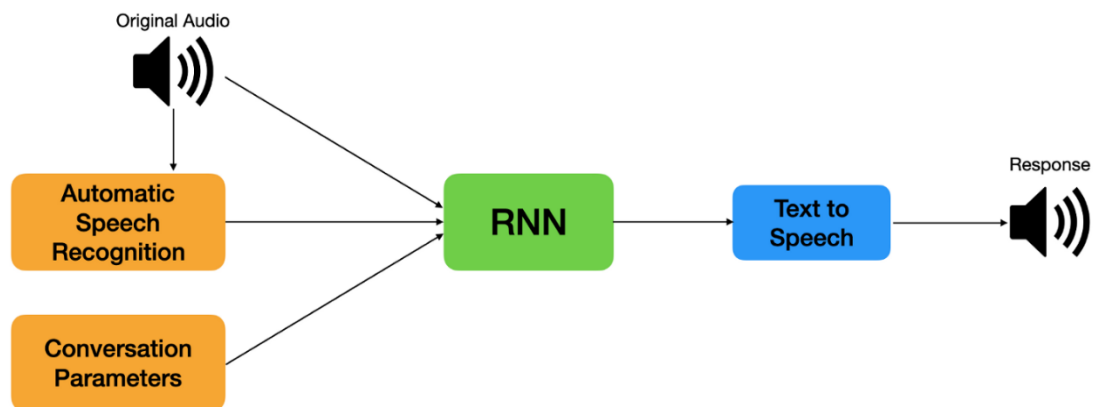
³³⁰ Cfr. Adobe Press, 15 mind-blowing stats about voice assistants, [15 mind-blowing stats about voice assistants \(adobe.com\)](#)

³³¹ Cfr. Eric Bieller, [What's The Future Of Human-Computer Interaction \(HCI\)? \(careerfoundry.com\)](#)

³³² Cfr. TIMOTHY B. LEE, [Why Google believes machine learning is its future | Ars Technica](#)

otro idioma, crea subtítulos para vídeos y durante una llamada transcribe en tiempo real en un documento lo que dice el interlocutor para ayudar a las personas con problemas de audición a hablar por teléfono. Además, los usuarios con su voz pueden acceder a cualquier aplicación, navegar por la galería de imágenes y escribir mensajes³³³.

El segundo producto es Google Duplex. Cuando los usuarios quieren llamar, organizar un evento o hacer una reserva en un restaurante, simplemente introducen la información y el sistema operativo activa Duplex que llamará y hablará con el interlocutor. Al final de la llamada los usuarios recibirán un resumen de la misma. Duplex utiliza el procesamiento del lenguaje natural y la red neuronal recurrente (*recurrent neural network*, RNN) para traducir y escribir lo que dice el interlocutor y formular frases de respuesta. El reconocimiento automático del habla (ASR) está conectado a la RNN y reconoce la voz y contextualiza las frases del interlocutor. El sistema *text-to-speech* (TTS) recibe el resultado de la RNN y lo convierte en mensajes hablados³³⁴.



La imagen muestra el funcionamiento de Google Duplex

Fonte: [NLP in the Real World: Google Duplex | by Jerry Wei | Towards Data Science](#)

³³³ In this video, we can see Google Digital Assistant features. . [Next Generation Google Assistant: Demo 2 at Google I/O 2019 - YouTube](#)

³³⁴ Cfr. Jerry Wei, NPL in the Real World, Towards Data Science, [NLP in the Real World: Google Duplex | by Jerry Wei | Towards Data Science](#)

6.1 Perplejidades y preocupaciones sobre un futuro basado en la inteligencia artificial

Aunque la sociedad puede beneficiarse de la inteligencia artificial porque facilita la vida, especialmente a las personas de grupos desfavorecidos, los ciudadanos y los organismos nacionales plantean su preocupación.

Muchos ciudadanos de EE.UU. se preguntan si es moralmente aceptable que un asistente digital hable como un humano sin que el interlocutor lo sepa. Además, muchos se preguntan si personas malintencionadas utilizan el Asistente de Google para llamar a personas con el fin de robar sus datos bancarios y revenderlos³³⁵.

El senador estadounidense Coons llamó la atención a la política de privacidad de Amazon en relación con el uso doméstico de Alexa. Amazon explicó que controlan los mensajes de voz y las transcripciones de las grabaciones de voz de los usuarios se almacenan en los servidores de Amazon de forma indefinida³³⁶. Muchos economistas y científicos, como Paul Romer (premio Nobel) y Nick Bloom (profesor de la Universidad de Stanford), están preocupados por el fenómeno del desempleo tecnológico. La reducción de personal en diversos sectores es más frecuente, como en la industria manufacturera y la agricultura. Además, los empleados de oficina se redujeron en un 30% en EE.UU. porque ahora los ordenadores son capaces de analizar los currículos y seleccionar a los candidatos de forma más rápida y sin prejuicios³³⁷.

³³⁵ Cfr. Chris Butler, Is Google Duplex ethical and moral?, Towards Data Science [Is Google Duplex ethical and moral? | by Chris Butler | Towards Data Science](#)

³³⁶ Cfr. Chris Coons, [Amazon responds to Sen. Coons' concerns about consumer privacy practices for Alexa devices \(senate.gov\)](#)

³³⁷ Eric Rosenbau, IBM artificial intelligence can predict with 95% accuracy which workers are about to quit their jobs, <https://www.cnbc.com/2019/04/03/ibm-ai-can-predict-with-95-percent-accuracy-which-employees-will-quit.html>

Stephen Hawking, Premio Nobel de Física, describió la inteligencia artificial como el mal de la sociedad, si los gobiernos nacionales en colaboración con la Unión Europea no aprueban leyes para orientar la inteligencia artificial hacia lo humanitario y para mejorar la calidad de vida³³⁸.

En conclusión, la inteligencia artificial y el procesamiento del lenguaje natural mejoran las sociedades; por ejemplo, las personas con deficiencias auditivas pueden hablar libremente por teléfono y los documentos pueden traducirse a varios idiomas en segundos. Sin embargo, sigue preocupando sobre cómo gestionar los datos de los usuarios sin violar los protocolos de privacidad y cómo proteger a las personas de quienes quieren utilizar la inteligencia artificial para perjudicar a otros.

³³⁸ Cfr. Arjun Kharpal, Stephen Hawking says A.I. could be ‘worst event in the history of our civilization’, <https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html>

7. Conclusiones

El objetivo de esta tesis es mostrar la relación intrínseca entre la inteligencia artificial y las lenguas naturales. El estudio analítico de estas últimas ha sido fundamental en el desarrollo de algoritmos, cuyo análisis se acerca al pensamiento humano. El PNL permite realizar funciones cada vez más complejas en poco tiempo para optimizar el tiempo y los recursos. Además, la tesis destaca la compleja arquitectura del hardware, que permite realizar operaciones como traducir, dar órdenes a un asistente de voz, comparar datos y ayudar en el trabajo. Estas tareas, que pueden parecer fáciles, requieren un análisis exhaustivo del lenguaje y la comunicación humanos, y el apoyo de una red neuronal bien entrenada.

La continua digitalización y las mejoras en el procesamiento del lenguaje natural (PLN) han creado puestos de trabajo y facilitado nuestra vida. Las personas pueden comunicarse en diferentes idiomas gracias a los traductores neuronales, los documentos pueden analizarse en segundos y, para las personas con problemas de audición, los asistentes digitales les ayudan a hablar por el móvil.

Sin embargo, estoy de acuerdo con la preocupación por el mal uso de la inteligencia artificial y la protección de los datos sensibles de los usuarios.

Además, existe el problema del desempleo provocado por la inteligencia artificial, que ha afectado a los sectores secundario y terciario de nuestra sociedad. Este fenómeno aumenta la desigualdad social. De hecho, los trabajadores de clase media se ven sustituidos por ordenadores, por ejemplo, las empresas emplean pocos trabajadores y prefieren utilizar traductores neuronales, asistentes digitales que

organizan las agendas de los administradores y programas capaces de seleccionar a los candidatos para las empresas³³⁹.

Otro aspecto es la necesidad de contar con leyes que gestionen el uso de la inteligencia artificial para evitar abusos y prácticas que perjudiquen a los ciudadanos. Por otra parte, también hay que mirar el panorama general y señalar que la reducción de puestos de trabajo en algunos campos no sólo conduce al desempleo, sino abre el camino a nuevos puestos de trabajo.

En conclusión, el desarrollo de la inteligencia artificial y el procesamiento del lenguaje natural debe considerarse una forma de progresar y mejorar nuestras vidas. Es importante destacar que cada uno de nosotros, las organizaciones internacionales y los gobiernos nacionales utilizan la tecnología para el bien de la sociedad. Espero que seamos más conscientes de las ventajas que ofrece el procesamiento del lenguaje natural y hagamos un buen uso de él.

³³⁹ Cfr. United Nations Press, [Will robots and AI cause mass unemployment? Not necessarily, but they do bring other threats | United Nations](#)

Ringraziamenti

Sono altamente fiero per il lavoro svolto, ringrazio tutti coloro che mi sono stati accanto in un periodo non facile della mia vita e che hanno contribuito alla buona riuscita della tesi.

I miei ringraziamenti, in particolar modo, vanno alla relatrice e dirigente universitaria, Professoressa Adriana Bisirri, per avermi consigliato nella scelta dell'argomento trattato, alle mie correlatrici di lingua inglese, Professoressa Maria Nocito, e spagnolo, Professoressa Luciana Banegas, per avermi guidato al fine di scrivere un elaborato eccellente dal punto di vista linguistico e alla mia correlatrice, Professoressa d'Informatica Applicata, Claudia Piemonte, per avermi guidato alla stesura dell'elaborato dal punto di vista stilistico e informatico.

Infine, dedico la tesi a me stesso, in quanto fra molte vicissitudini, con la costanza, la tenacia e la determinazione sono riuscito a scrivere questa tesi unica nel suo genere, che contiene un volume di informazioni considerevole e che verte su un argomento tanto attuale quanto complesso, le cui informazioni sono presenti principalmente in inglese e sono quasi inesistenti in italiano.

BIBLIOGRAFIA

Shelly Fan, *Will AI Repleca Us?*, 1st edition, 2019, United Kingdom, Thames and Hudson.

Tom Taulli, *Artificial Intelligence Basics*, 1st edition, 2019, United States of America, Apress.

Sitografia

- A. Gonfalonieri, How Amazon Alexa works? Your guide to Natural Language Processing (AI), <https://towardsdatascience.com/how-amazon-alexa-works-your-guide-to-natural-language-processing-ai-7506004709d3>, consultato il 20/04/2021
- A. Kinth, 32 Advantages and Disadvantages of Deep Learning, <https://medium.com/@deeplearning1000/32-advantages-and-disadvantages-of-deep-learning-564d81bc0287>, consultato il 04/05/2021
- Agato Translation, What is dictionary-based machine translation?, <https://agatotranslate.ae/what-is-dictionary-based-machine-translation/>, consultato il 03/02/2021
- Ai Multiple, What is Deep Learning? Explore Use Cases / Examples, Benefits, <https://research.aimultiple.com/deep-learning/>, consultato il 03/03/2021
- Ai Multiple, What is Deep Learning? Explore Use Cases / Examples, Benefits, <https://research.aimultiple.com/deep-learning/>, consultato il 2/05/2021
- Amand Anan, Deep Learning Trends: top 20 best uses of GPT-3 by OpenAI <https://www.educative.io/blog/top-uses-gpt-3-deep-learning>, consultato il 26/02/2021
- Amanda Porter, What are the advantages of Natural Language Processing (NLP)?, <https://capacity.com/enterprise-ai/faqs/what-are-the-advantages-of-natural-language-processing-nlp/>, consultato il 14/03/2021
- Amol Movuduru, What is GPT-3 and why is it so powerful?, <https://towardsdatascience.com/what-is-gpt-3-and-why-is-it-so-powerful-21ea1ba59811>, consultato il 26/02/2021
- Andrew Kuo, Noam Chomsky on the future of deep learning, <https://towardsdatascience.com/noam-chomsky-on-the-future-of-deep-learning-2beb37815a3e>, consultato il 18/01/2021
- Arjun Kharpal, Stephen Hawking says A.I. could be ‘worst event in the history of our civilization’, <https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html>, consultato il 12/06/2021
- B. Lee, Why Google believes machine learning is its future, <https://arstechnica.com/gadgets/2019/05/googles-machine-learning-strategy-hardware-software-and-lots-of-data/>, consultato il 08/06/2021
- B. Maar, Are Alexa And Siri Considered AI?, <https://bernardmarr.com/are-alexa-and-siri-considered-ai/>, consultato il 28/04/2021
- C. Butler, Is Google Duplex ethical and moral?, <https://towardsdatascience.com/is-google-duplex-ethical-and-moral-f66a23637640>, consultato il 10/06/2021
- C. Coons, Amazon responds to Sen. Coons’ concerns about consumer privacy practices for Alexa devices, <https://www.coons.senate.gov/news/press-releases/amazon-responds-to-sen-coons-concerns-about-consumer-privacy-practices-for-alexa-devices>, consultato il 10/06/2021

Cristina Bosco, Linguistica Computazionale, <http://www.di.unito.it/~bosco/lingue2017/NLP-2017-1-.pdf>, consultato 30/01/2021

D. Linskey, Alexa, are you invading my privacy?' – the dark side of our voice assistants, <https://www.theguardian.com/technology/2019/oct/09/alexa-are-you-invading-my-privacy-the-dark-side-of-our-voice-assistants>, consultato il 12/06/2021

Data Flair, Advantages and Disadvantages of Machine Learning Language, <https://data-flair.training/blogs/advantages-and-disadvantages-of-machine-learning/>, consultato il 14/03/2021

DeepAI, Machine Translation, <https://deepai.org/machine-learning-glossary-and-terms/neural-machine-translation>, consultato il 18/02/2021

Dimitris Glezos, Everything you need to know about machine translation, <https://www.transifex.com/blog/2015/machine-translation-101/>, consultato il 31/01/2021

E. Bieller, What's The Future Of Human-Computer Interaction (HCI)?, <https://careerfoundry.com/en/blog/ui-design/the-future-of-human-computer-interaction-whats-next/>, consultato il 06/05/2021

E. Rosenbaum, IBM artificial intelligence can predict with 95% accuracy which workers are about to quit their jobs, <https://www.cnbc.com/2019/04/03/ibm-ai-can-predict-with-95-percent-accuracy-which-employees-will-quit.html>, consultato il 11/06/2021

for Languages with Limited Resources, https://www.cs.cmu.edu/~jgc/publication/A_Trainable_Transfer_Based_Machine_EAMT_2004.pdf, consultato il 12/02/2021

Francesca Fiore, Linguaggio Universale Secondo Noam Chomsky, <https://www.stateofmind.it/2015/12/linguaggio-universale-psicologia/>, consultato il 06/01/2021

Francisco Alonso, Neural Network and Deep Learning, <https://www.futurespace.es/en/redes-neuronales-y-deep-learning-capitulo-1-preludio/>, consultato il 01/03/2021

Free Code Camp, A history of machine translation from the Cold War to deep learning, <https://www.freecodecamp.org/news/a-history-of-machine-translation-from-the-cold-war-to-deep-learning-f1d335ce8b5/>, consultato il 16/02/2021

G. Kesten, 15 mind-blowing stats about voice assistants, <https://blog.adobe.com/en/publish/2020/09/21/mind-blowing-stats-voice-assistants.html#gs.c7pded>, consultato il 06/05/2021

Google, Google's AI Assistant Can Now Make Real Phone Calls, https://www.youtube.com/watch?v=JvbHu_bVa_g&t=56s, consultato il 09/06/2021

Google, Next Generation Google, <https://www.youtube.com/watch?v=QsX7uEEKCDE>, consultato il 23/05/2021

Google, Project Euphonia, <https://www.youtube.com/watch?v=OAdegPmkK-o&t=41s>, consultato il 08/06/2021

Hagen Böhm, Dialogue Management Using Finite State Models, <https://www.coli.uni-saarland.de/~korbay/Courses/DM-SS02/DM-slides/hagen-fsm-slides.pdf>, consultato il 01/02/2021

IBM 100, Gli albori della linguistica computazionale, [IBM100 - IBM100 - Gli albori della linguistica computazionale - Italia](#), consultato il 09/01/2021

IBM Cloud Education, Artificial Intelligence, <https://www.ibm.com/cloud/learn/what-is-artificial-intelligence>, consultato il 10/01/2021

IBM Cloud Education, Natural Language Processing, <https://www.ibm.com/cloud/learn/natural-language-processing>, consultato il 22/02/2021

IBM Cloud, Speech Recognition, <https://www.ibm.com/cloud/learn/speech-recognition>, consultato il 16/04/2021

Ionos, How does Natural Language Processing work?, <https://www.ionos.com/digitalguide/online-marketing/online-sales/how-does-natural-language-processing-work/>, consultato il 07/03/2021

J. Wei, NLP in the Real World: Google Duplex, <https://towardsdatascience.com/nlp-in-the-real-world-google-duplex-d96160d3770b>, consultato il 09/06/2021

James Rogers, Formal Language Theory : Refining The Chomsky Hierarchy, <https://royalsocietypublishing.org/doi/10.1098/rstb.2012.0077>, consultato il 06/01/2021

Jason Brownlee, A Gentle Introduction to Neural Machine Translation, <https://machinelearningmastery.com/introduction-neural-machine-translation/>, consultato il 16/02/2021

Jason Brownlee, How to Develop a Neural Machine Translation System from Scratch, <https://machinelearningmastery.com/develop-neural-machine-translation-system-keras/>, consultato il 22/02/2021

Joe Sommerlad, GOOGLE TRANSLATE: HOW DOES THE MULTILINGUAL INTERPRETER ACTUALLY WORK?, <https://www.independent.co.uk/life-style/gadgets-and-tech/news/how-does-google-translate-work-b1821775.html>, consultato il 30/01/2021

K. Foote, A Brief History of Artificial Intelligence, <http://www.dataversity.net/brief-history-artificial-intelligence/>, consultato il 27/02/2021

K. Probst, A Trainable Transfer-based Machine Translation Approach

K. Reyes, What is Deep Learning and How Does It Works, https://www.simplilearn.com/tutorials/deep-learning-tutorial/what-is-deep-learning?source=sl_frs_nav_playlist_video_clicked, consultato il 27/02/2021

Keith D. Foote, A Brief History of Natural Language Processing, <https://www.dataversity.net/a-brief-history-of-natural-language-processing-nlp/>, consultato il 06/03/2021

M. Phi, Illustrated Guide to LSTM's and GRU's: A step by step explanation, <https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21>, consultato il 10/04/2021

M.T.Kantan, RBMT vs SMT, <https://kantanmtblog.com/2014/02/13/rbmt-vs-smt/>, consultato il 01/02/2021

Machine Learning Knowledge, Brief History of Deep Learning from 1943-2019 [Timeline], https://machinelearningknowledge.ai/brief-history-of-deep-learning/#Deep_Learning_History_Timeline, consultato il 28/02/2021

Magoosh, How Does Siri Work? The Science Behind Siri, <https://magoosh.com/data-science/siri-work-science-behind-siri/>, consultato il 10/03/2021

MIT Media Lab, Marvin Minsky, “father of artificial intelligence,” dies at 88, <https://news.mit.edu/2016/marvin-minsky-obituary-0125>, consultato il 26/01/2021

Nigel Ward, Shrdlu, <http://www.cs.utep.edu/nigel/papers/shrdlu.pdf>, consultato il 23/01/2021

NTT, Machine Translation, <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201312fa2.html>, consultato 05/02/2021

Oxford Academic, Mind, <https://academic.oup.com/mind/article/LIX/236/433/986238>, consultato il 11/01/2021

Oxford Academic, Mind, <https://academic.oup.com/mind/pages/About>, consultato il 11/01/2021

P. Gupta, Decision Trees in Machine Learning, <https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052>, consultato il 10/03/2021

Q. Lanners, Neural Machine Translation, <https://towardsdatascience.com/neural-machine-translation-15ecf6b0b>, consultato il 14/04/2021

Q. Lanners, Neural machine translation, <https://towardsdatascience.com/neural-machine-translation-15ecf6b0b>, consultato il 18/02/2021

R. Khandelwal, Implementing neural machine translation using keras, <https://towardsdatascience.com/implementing-neural-machine-translation-using-keras-8312e4844eb8>, consultato il 19/02/2021

R. Khandelwal, Intuitive explanation of Neural Machine Translation, <https://towardsdatascience.com/intuitive-explanation-of-neural-machine-translation-129789e3c59f>, consultato il 20/02/2021

RfWireless, Advantages of Deep Learning | disadvantages of Deep Learning, <https://www.rfwireless-world.com/Terminology/Advantages-and-Disadvantages-of-Deep-Learning.html>, consultato il 02/03/2021

S. Amin, Speech recognition is hard — Part 1, <https://towardsdatascience.com/speech-recognition-is-hard-part-1-258e813b6eb7>, consultato il 17/04/2021

S. Arora, K. Batra, S. Singh, Dialogue System, <https://arxiv.org/ftp/arxiv/papers/1306/1306.4134.pdf>, consultato il 02/02/2021

S. Mahapatra, Why Deep Learning over Traditional Machine Learning?, <https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063>, consultato il 01/03/2021

S. Mahapatra, Why Deep Learning over Traditional Machine Learning?, <https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063>, consultato il 02/05/2021

S. Perez, Amazon responds senator's inquiry confirms alexa voice records are kept indefinitely, <https://techcrunch.com/2019/07/03/amazon-responds-to-a-u-s-senators-inquiry-confirms-alexa-voice-records-are-kept-indefinitely/>, consultato il 11/06/2021

S. Perez, Nearly 70% of Us smart speaker owners use amazon echo devices, <https://techcrunch.com/2020/02/10/nearly-70-of-u-s-smart-speaker-owners-use-amazon-echo-devices/>, consultato il 25/04/2021

S. Yip, What is Neural Machine Translation & How does it work?, <https://www.translatefx.com/blog/what-is-neural-machine-translation-engine-how-does-it-work?lang=en>, consultato il 21/02/2021

Safe at last, Incredible Amazon Alexa Statistics You Need to Know in 2021, <https://safeatlast.co/blog/amazon-alexa-statistics/>, consultato il 23/04/2021

Salvatore Sorce, Introduzione alla Linguistica Computazionale, http://www1.unipa.it/sorce/didattica/sei1213/SEI1213_01_Linguistica_Computazionale_intro.pdf, consultato 27/01/2021

Sapere, Backus-Naur Form, <https://www.sapere.it/enciclopedia/Backus-Naur+Form.html>, consultato il 23/01/2021

Sciforce, Biggest Open Problems in Natural Language Processing, <https://medium.com/sciforce/biggest-open-problems-in-natural-language-processing-7eb10ccfc9>, consultato il 11/03/2021

Simpli Learn, What is Big Data Analytics and Why It is Important?, <https://www.simplilearn.com/what-is-big-data-analytics-article>, consultato il 29/02/2021

Smartling, What Is Machine Translation?, <https://www.smartling.com/resources/101/state-of-machine-translation/>, consultato il 23/02/2021

Smartling, What is Machine Translation?, <https://www.smartling.com/resources/101/state-of-machine-translation/>, consultato il 25/03/2021

Terry Winograd, Procedure As Representation For Data In A Computer For Understanding Natural Language, <http://hci.stanford.edu/winograd/shrdlu/AITR-235.pdf>, consultato 26/01/2021

The Guardian, A robot wrote this entire article. Are you scared yet, human?, <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>, consultato il 25/02/2021

Unife, Breve riassunto della teoria chomskyana del linguaggio, http://www.unife.it/lettere/ filosofia/comunicazione/insegnamenti/fondamenti_comunicazione_musicale/materiale_didattico/comunicazione-musicale-2013-2014-materiali-per-la-prima-parte/breve-riassunto-della-teoria-chomskyana-del-linguaggio, consultato il 23/01/2021

United Language Group, The Pros And Cons Of Statistical Machine Translation, <https://www.unitedlanguagegroup.com/blog/pros-and-cons-statistical-machine-translation>, consultato il 03/02/2021

V. Kurama, The past, present and future of deep learning, <https://towardsdatascience.com/the-past-present-and-future-of-deep-learning-adb4d60eaf24>, consultato il 12/06/2021

V. Shchutskaya, Deep Learning: Strengths and Challenges, <https://indatalabs.com/blog/deep-learning-strengths-challenges>, consultato il 02/03/2021

W.N.Francis, Brown Corupus, <http://icame.uib.no/brown/bcm.html>, consultato il 30/01/2021

Wade Roush, Marvin Minsky on common sense and computers that emote, <https://www.technologyreview.com/2006/07/13/228604/marvin-minsky-on-common-sense-and-computers-that-emote/>, consultato il 22/01/2021

Warren Weaver, Translation, <https://web.stanford.edu/class/linguist289/weaver001.pdf>, consultato 28/01/2021

Wikipedia, Regular Language, https://en.wikipedia.org/wiki/Regular_language, consultato il 10/01/2021