

Aproximació a la traducció automàtica parlada amb altaveus intel·ligents

NIU: 1422505

Alumna: Paula Pastur Bustos

Tutor: Eduard Simón

Estudis: Màster en Tradumàtica: Tecnologies de la traducció

Data: 22 de juny de 2020



Universitat Autònoma de Barcelona

Taula de continguts

1. Introducció	1
2. Marc teòric.....	4
2.1. Els sistemes de funcionament per veu	4
2.1.1. La síntesi de veu (text to speech)	5
2.1.2. El reconeixement de veu (speech to text)	15
2.1.3. L'intel·ligència artificial (IA) i el processament de veu (speech to speech)	29
2.1.3.1. Les bases de la intel·ligència artificial	31
2.1.3.2. La intel·ligència artificial a partir dels anys 50	38
2.1.3.3. Situació actual: realitat i desafiaments.....	46
2.1.3.4. El <i>Natural Language Processing</i> (NLP).....	50
2.2. La traducció automàtica	56
3. Marc pràctic	62
3.1. Els altaveus intel·ligents	62
3.1.1. Echo Dot Alexa	63
3. 1. 2. Google Home	66
3.2. Prova de traducció	68
3.2.1. Prova 1: substantius simples de diferents especialitats	69
3.2.1.1. Del castellà a l'anglès	69

3.2.1.2. Del castellà al francès	70
3.2.2. Prova 2: sintagmes de diferents especialitats	71
3.2.2.1. Del castellà a l'anglès	71
3.2.2.2. Del castellà al francès	72
3.2.3. Prova 3: oracions	73
3.2.3.1. Del castellà a l'anglès	73
3.2.3.2. Del castellà al francès	74
3.2.4. Prova 4: preguntes simples.....	75
3.2.4.1. Del castellà a l'anglès	75
3.2.4.2. Del castellà al francès	76
3.2.5. Prova 5: refranys.....	77
3.2.5.1. Del castellà a l'anglès	77
3.2.5.2. Del castellà al francès	78
3.2.6. Prova 6: frases cèlebres	80
3.2.6.1. Del castellà a l'anglès	80
3.2.6.2. Del castellà al francès	81
3.2.7. Resultats	83
4. Conclusions	85
5. Bibliografia.....	87
6. Annexos.....	92

Dades del TFM/ Datos del TFM/ Dissertation data

Títol: Aproximació a la traducció automàtica parlada amb altaveus intel·ligents

Título: Aproximación a la traducción automática hablada con altavoces inteligentes

Title: Spoken machine translation approach using smart speakers

Resum del TFG/ Resumen del TFG/ Abstract

Vivim en un món en què la tecnologia ha arribat a tots el àmbits. Ara com ara els altaveus intel·ligents s'estan provant per fer traduccions, ja que ens permeten fer consultes sense treure les mans del teclat. En aquest treball, provarem de traduir unitats lingüístiques per veu del castellà a l'anglès i al francès, per concloure quin dels altaveus tradueix millor i per tant, ens serviria més.

Vivimos en un mundo en el que la tecnología ha llegado a todos los ámbitos. Hoy en día, los altavoces inteligentes se están probando para hacer traducciones ya que permiten hacer consultas sin despegar las manos del teclado. En este trabajo, traduciremos unidades lingüísticas por voz del español al inglés y al francés para saber cuál de los dos altavoces traduce mejor y, por consiguiente, nos sería más útil.

The world where we live is surrounded by technology in all the areas. Nowadays, smart speakers are being tested for translation because they allow people consulting or searching for things without taking their hands off the keyboard. In this paper, language units will be translated by voice from Spanish into English and into French, in order to find out which speaker (Google Home or Amazon's Alexa) translates better and would, therefore, be more useful.

Paraules clau/ Palabras clave/ Key words

Altaveu intel·ligent, processament de veu, intel·ligència artificial, traducció, síntesi de veu, reconeixement de veu, Google Home, Echo Dot

Altavoz inteligente, procesamiento de voz, inteligencia artificial, traducción, síntesis de voz, reconocimiento de voz, Google Home, Echo Dot

Smart speaker, speech processing, artificial intelligence, translation, speech synthesis, speech recognition, Google Home, Echo Dot

Avís legal/ Aviso legal/ Legal notice

© Paula Pastur Bustos, Barcelona, 2019. Tots els drets reservats.

Cap contingut d'aquest treball pot ser objecte de reproducció, comunicació pública, difusió i/o transformació, de forma parcial o total, sense el permís o l'autorització del seu autor/de la seva autora.

© Paula Pastur Bustos, Barcelona, 2019. Todos los derechos reservados.

Ningún contenido de este trabajo puede ser objeto de reproducción, comunicación pública, difusión y/o transformación, de forma parcial o total, sin el permiso o la autorización de su autor/a.

© *Paula Pastur Bustos, Barcelona, 2019. All rights reserved.*

None of the content of this academic work may be reproduced, distributed, broadcasted and/or transformed, either in whole or in part, without the express permission or authorization of the author.

1. Introducció

Motivació

Quan estudiava el Grau en Traducció i Interpretació a la UAB, vaig adonar-me que m'interessava especialment el món de les tecnologies, ja que sempre havia tingut un especial interès en l'automatització de tasques. Diverses assignatures que vaig tenir durant el grau em van fer aprendre més i tenir encara més ganes de saber-ne. Personalment, considero que, vulguem o no, és un fet que en l'àmbit de la traducció, com en molts més, les tecnologies són ja una realitat. L'automatització de processos ens permeten fer les coses de manera més ràpida i, com a conseqüència, estalviar temps i ser més productius. És per això que vaig decidir-me a estudiar el màster de Tradumàtica: Tecnologies de la Traducció a la UAB. Tot i que al principi tenia la idea mental que aprendria sobre traducció automàtica, no sabia ben bé quins altres processos implicaven les tecnologies en l'àmbit. Per a la meva sorpresa, m'he adonat que hi ha infinites aplicacions de les tecnologies en diferents parts de diversos processos i, fins i tot, he conegut parts que no sabia que existien. Inicialment, vaig pensar en fer un treball purament centrat en la traducció, però a poc a poc vaig adonar-me'n, amb l'ajuda d'alguns professors, que hi ha un camp que no està gaire explorat ara per ara, sobretot la seva aplicació a la traducció: el processament de la veu. És per això que em vaig decidir a fer un treball que inclogués part del meu àmbit d'especialització i també una part d'innovació, d'un àmbit en què tot i no saber-ne gaire, estava segura que aprendria molt i em serviria en un futur. Així doncs, he desenvolupat un treball que m'ha servit per conèixer molts aspectes lingüístics dels altaveus intel·ligents, així com per aprendre el desenvolupament tecnològic que ha portat fins al punt on som: un món en constant desenvolupament, on res és avui com serà demà, i en què, a parer meu, les màquines de processament de veu tenen molt a oferir.

Objectius

L'objectiu principal d'aquest treball és fer un estudi introductori sobre el funcionament dels sistemes de traducció automàtica integrats en els altaveus intel·ligents Google Home i Echo Dot. Analitzem els tipus de traducció automàtica que integren i les seves limitacions a l'hora de traduir. A la secció pràctica interactuem amb els altaveus, fent

que aquests tradueixin diferents tipus de frases, de diferents llargàries, especialitats i formulades de maneres diferents. D'aquesta manera, extraïem algunes conclusions pel que fa a les diferències entre tots dos models . L'exercici es realitza amb els sistemes de TA incorporats de fàbrica i sense instal·lar cap *skill* addicional en cap cas. Pel que fa a l'elecció dels altaveus per aquest exercici, és evident que ara com ara hi ha tres marques d'altaveus intel·ligents al mercat que destaquen sobre les altres quant a desenvolupament i qualitat: el d'Apple (HomePod), el d'Amazon (Echo) i el de Google (Google Home). Per motius logístics, aquest exercici es va haver de limitar a dos d'aquests models. Respecte a les prestacions que ofereix cadascun i quant a desenvolupament de la TA que incorpora cadascun, els tres models estan força igualats. Sí és cert que el d'Apple (amb un preu de venda al públic superior), segons alguns experts amb qui vaig contactar, té una millor qualitat d'àudio, factor que no era excessivament rellevant per aquest exercici. Conseqüentment, la pràctica s'ha realitzat fent servir els altaveus intel·ligents d'Amazon i de Google.

Metodologia

En primer lloc, i durant força mesos, em vaig dedicar a cercar informació en llibres, a Internet i enciclopèdies. D'aquesta manera, vaig ser capaç de poder estructurar el meu marc teòric per posar en context tot allò que faria falta saber per entendre com funcionen els altaveus intel·ligents. Així doncs, un cop acabada aquesta part, vaig procedir a fer algunes proves en els altaveus per saber com funcionaven perquè no tenia cap dels dos abans de fer aquest treball. Vaig interactuar amb ells, fer diverses proves per saber quines formulacions faria servir a l'hora de realitzar la pràctica, vaig també fer proves de quant de temps poden escoltar la meua veu, vaig provar també de buscar paraules en el diccionari que tenen integrat (que, per cert, vaig trobar que l'Echo Dot d'Alexa busca les paraules a la RAE i el Google Home al Lexico, la versió espanyola d'Oxford Dictionaries), i, per últim, vaig provar de fer algunes proves amb les algunes paraules que ja havia decidit amb anterioritat. Després d'això, vaig començar a confeccionar la pràctica. En primer lloc, i com ja explicaré quan arribi l'hora més endavant, vaig buscar les paraules i les oracions dels exercicis perquè cada prova estigués destinada obtenir un resultat diferent i, un cop vaig tenir les proves preparades, vaig procedir a dur-les a terme. A continuació, vaig escriure unes breus conclusions que

a la mateixa vegada em van fer reflexionar per després puntuar les traduccions que havien fet l'Alexa i el Google Home. Un cop puntuades totes les unitats lingüístiques, vaig fer uns gràfics il·lustratius i, finalment, vaig descriure els resultats de la pràctica i, per últim, vaig escriure les conclusions del treball.

2. Marc teòric

És un fet que la tecnologia és aquí per quedar-se en molts aspectes de la nostra vida quotidiana i, cada cop més, en la professional.

La traducció és un dels oficis més antics de la història, ja que neix de la necessitat de comunicació entre comunitats per poder intercanviar productes, cultura i fins i tot idees. En un món globalitzat cal aquest intercanvi en l'àmbit econòmic, polític, cultural i, ara com ara, tecnològic, per avançar en tots els aspectes i per a això, la llengua n'és un instrument essencial i com a conseqüència, ho és també la traducció.

Ja han passat milers d'anys des que es va començar a produir aquest canvi i, en el món en què vivim ara com ara, és evident que la tecnologia avança a la velocitat de la llum. Avui en dia, una de les tècniques més "innovadores" és el reconeixement de veu i, tot i que el seu funcionament encara no és perfecte, podria arribar a ser-ho. No obstant això, a continuació no ens centrarem pas a fer un estudi dels precedents de la traducció o en parlar de la traducció. Focalitzarem aquest apartat del marc teòric en l'estudi de la veu, en la seva síntesi, funcionament, desenvolupament i producció, per entendre els aspectes lingüístics i el procés que hi ha al darrere. Així doncs, podrem treure'n el màxim partit possible i els podrem fer servir tot sabent amb què estem treballant i com funcionen, ja que com hem dit abans, en un futur no gaire llunyà, podrien convertir-se en una eina indispensable en el nostre sector.

2.1. Els sistemes de funcionament per veu

La necessitat de disposar de sistemes que ens simplifiquin la nostra interacció home-màquina mitjançant la interfície més natural de la que podem disposar, la nostra pròpia parla. Per explicar aquest tipus de sistemes cal tenir en compte el que anomenem processament de veu, fenomen que inclou, fonamentalment, dues categories: la síntesi del llenguatge i el reconeixement automàtic de la parla. Un sistema de síntesi de veu és, segons Varona (1997), un instrument capaç de produir veu que imita la veu humana, i l'explicarem en més detall en l'apartat 2.1.1. D'altra banda, un sistema de reconeixement de veu és un mecanisme que pot descodificar el senyal acústic que produeix un aparell fonador locutor, en una seqüència d'unitats lingüístiques que

contenen el missatge que el locutor vol transmetre. Totes dues categories estan, avui dia, ja força desenvolupades gràcies a la intel·ligència artificial, tot i que cap de les dues funciona encara a la perfecció. Aquests dos sistemes que componen el processament de la veu, per la seva naturalesa de comunicació persona-màquina, comparteixen conceptes i tècniques que l'engloben el PV en la disciplina del processament de senyals i els converteix, d'aquesta manera, en una eina extrapolable a molts àmbits, com l'ajuda a minusvàlids, la facilitació a l'hora de fer transaccions bancàries i, en el cas concret de l'objecte d'estudi del present treball, el control oral de sistemes de traducció automàtica.

A continuació, en els apartats 2.1.1 i 2.1.2 mostrarem la història cronològicament tant dels sistemes de síntesi de veu com dels de reconeixement. Aquest relat ens ajudarà a entendre el si de la qüestió, el processament de veu, del qual parlarem en última instància.

2.1.1. La síntesi de veu (text to speech)

Els sistemes de síntesi de veu, també anomenats en anglès *speech to text systems*, són aquells que permeten produir veu a partir de text. En aquest apartat explicarem la seva evolució, des de la seva aparició, fins avui dia. Amb això, constituïrem una base per poder entendre, més endavant, com funcionen els que avui anomenem altaveus intel·ligents, que són capaços de 'reconèixer' la veu, com veurem en la secció 2.1.2 sobre el reconeixement de veu, i de produir-la, com veurem a continuació.

El desenvolupament dels sistemes de síntesi de veu és un pas indispensable per tot allò en què va evolucionar (el reconeixement de veu) i tot allò en què es va convertir i en va facilitar la creació (els sistemes de processament de veu).

Tot i que els primers estudis sobre l'estudi de la veu daten del 1668, en què B.J Wilkins va estudiar sobre les posicions de la boca a l'hora de parlar (va elaborar un abecedari de símbols que representaven les posicions de la boca quan es produïen els diferents sorolls), no va ser fins que van aparèixer els primers dispositius que podem parlar dels sistemes que poden sintetitzar la veu humana com a tals o, en altres paraules, el *text-to-speech*, que deu el seu nom a que ens permet produir veu a partir d'un text escrit.

El primer dels intents de crear un sintetitzador de veu el va fer C.G. Kratzenstein l'any 1779 qui, a St. Petesburg, va fer un treball en què explicava les diferències fisiològiques dels cinc sons vocàlics (/a/, /e/, /i/, /o/, i /u/) i va crear un aparell que permetia reproduir-los. L'aparell era un ressonador acústic que tenia cinc tubs de formes diferents per crear les mateixes ressonàncies que les que produïm les persones quan parlem, com podem apreciar en la Figura 1. Per reproduir totes les vocals, excepte la /i/, es bufaven els tubs amb una mena de canya (com els instruments de música) que feien que ressonés. Per produir el so /i/, en canvi, es bufava sense canya ja que era el més fàcil de fer, i feia un soroll semblant al d'una flauta (Flanagan 1972, Flanagan et al. 1973, Schroeder 1993).

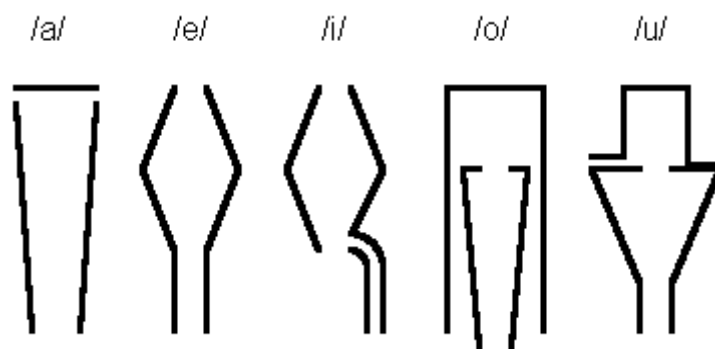


Figura 1. Formes de l'aparell que va presentar Kratzenstein a St. Petesburg (Schroeder 1993)

Alguns anys més tard, W.R. Kempelen, va presentar la seva 'Màquina mecànica-acústica parlant', que era capaç de produir sons vocàlics i consonàntics (Klatt 1987, Schroeder 1993). De fet, Kempelen havia començat abans a treballar en aquest sistema, l'any 1769, però no va ser fins vint anys més tard que va publicar la seva investigació sobre la producció de la parla humana i els experiments que havia fet amb la seva màquina que produïa sons. Les parts de la seva màquina reproduïen el sistema de processament de la parla d'una persona humana. Les parts essencials de la màquina eren una cambra de pressió que simulava els pulmons, una llengüeta vibratòria que actuava com a cordes vocals, i un tub de cuir per al tracte vocal.

Per produir sons vocàlics es modificava la forma del tub de cuir i, depenent de la posició exacta de la mà, en sonava un de diferent. D'altra banda, per produir els consonàntics, hi havia quatre orificis a la llengüeta que es controlaven amb els dits. Si s'obrien, es produïen els sons nasals 'm' i 'n'; si es dividia el corrent d'aire amb el

polze, la 'l'; i si es tapaven els orificis i es destapaven bruscament, sonaven les oclusives 'p', 'b', 't', 'd', 'k' i 'g'. Per als sons 's' i 'sh' es feien servir ressonadors i se subministrava l'aire a diferents nivells, i per últim, per generar 'f', 'v' i 'h' i el so alemany 'ch' es deixava escapar l'aire de la cambra d'aire comprimit de manera suau (més intensament per a la 'f' i més fluix per la 'h'). La màquina, segons l'inventor, podia dur a terme la síntesi de veu en tres setmanes en francès, italià i llatí (l'alemany era més difícil perquè hi ha més sons consonàntics). Abans de les seves demostracions, la laringe es considerava el centre de producció de la parla, però la seva recerca el va portar a la teoria que el tracte vocal, una cavitat entre les cordes vocals i els llavis, era el lloc principal d'articulació acústica. Tot i els avenços que va fer, els científics de l'època no el van prendre gaire seriosament per la màquina que havia presentat anteriorment a tot Europa, la 'Màquina de jugar als escacs'. Es suposava que la màquina podria desafiar l'oponent i sabia exactament què fer en cada moment per guanyar la partida a l'adversari, però anys més tard, i segons ell mateix també va admetre, es va saber que el mecanisme real de la màquina era un exjugador professional d'escacs sense comes que s'amagava sota el tauler (Flanagan et al. 1973, Schroeder 1993). No obstant això, la màquina parlant va resultar d'una gran quantitat d'esforços i de prova-error que el van portar a la conclusió que la millor manera d'imitar les cordes vocals humanes era el brunzit que la llengüeta provocava quan l'aire (que s'emmagatzemava en una cambra d'aire comprimit) hi passava a través.

A mitjan segle XIX, C. Wheatstone va construir una versió de la màquina parlant de Kempelen, com podem veure en la Figura 2, que va presentar a la reunió de l'Associació britànica per al progrés de les ciències a Dublín l'any 1850. Aquesta màquina era una mica més complexa i capaç de produir vocals i la majoria dels sons consonàntics, a més que fins i tot podia generar combinacions de sons i paraules senceres. Per produir les vocals es feia servir la llengüeta vibratòria i es tancaven els orificis, i les ressonàncies es feien deformant el ressonador de cuir, com en la màquina de Kempelen. Les consonants, fins i tot les nasals, es produïen gràcies a un flux turbulent a través d'un passatge adequat amb la llengüeta.

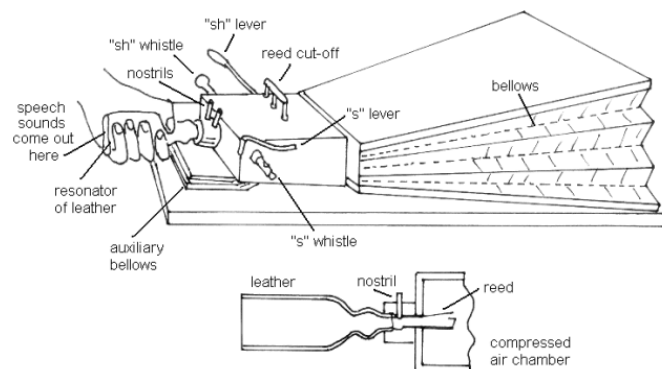


Figura 2. Reinterpretació de Wheatstone de la màquina parlant de von Kempelen (Flanagan 1972).

L'any 1838, Willis va descobrir la connexió que hi havia entre els sons vocàlics i la geometria del tracte vocal, i també va sintetitzar les vocals amb tubs ressonadors, com els tubs d'un òrgan. A més, va descobrir que la qualitat de les vocals no només depenia de la llargada dels tubs sinó també del seu diàmetre (Schroeder 1993).

A finals dels 1800, A.G. Bell, inspirat en la màquina de Wheatstone i juntament amb el seu pare i el seu germà, va crear una màquina parlant del mateix tipus. Bell, a més, va fer alguns experiments controvertits amb el seu gos, al qual es col·locava entre les cames i el feia grunyar i li va modificar el tracte vocal amb les mans perquè produís sons semblants a la parla humana. No obstant això, la màquina que van aconseguir fer, amb materials com goma, cotó i filferro, podia produir sons vocàlics i nasals molt satisfactoris i fins i tot, frases (Flanagan 1972, Shroeder 1993).

El 1846 J. Faber va construir una màquina a la qual va anomenar 'Euphoria', que va suposar un gran avenç respecte a la màquina de Von Kempelen perquè podia variar el to fonamental i, en conseqüència, entonar preguntes, xiuxiejar o produir veu normal. H.LF. Helmutz, l'any 1875 també va crear una màquina que podia variar el to i, a més, feia servir diapasons per produir les vocals artificialment. No obstant això, no va ser fins que es va inventar el gramòfon, a finals del segle XIX, que es va obrir una porta per a la investigació de la veu humana. El primer sistema que es va desenvolupar va ser el fonògraf. L'aparell consistia en un cilindre amb ranures amb forma d'hèlix que estava recobert amb una fulla fina d'estany i que es desplaçava davant d'una bobina acústica que tancava per una banda un diafragma de pergamí, estés sobre una petita caixa

cilíndrica. Un estilet arrodonit que estava pegat sobre el diafragma pressionava la fulla d'estany i produïa protuberàncies i buits per la pressió de les privacions de l'aire que provocava la veu, segons Gendre (1900). Per escoltar la gravació, es col·locava l'estilet al començament del cilindre i la botzina amplificava els sons que s'havien produït. Els cilindres, més tard, es van substituir per discos plans, cosa que va donar lloc a l'aparició del gramòfon.

L'any 1879 W.H. Preece i A. Stroh van examinar les estries del gramòfon per provar de descobrir la naturalesa física dels sons però, després d'adonar-se que no aconseguien cap resultat van provar de fer una aproximació inversa: van crear un sintetitzador mecànic que generava un to complex gràcies a la suma d'un to pur i un número variable d'harmònics. El van construir amb un conjunt de rodes engranades que rotaven a diverses velocitats. La marca que generava el sintetitzador era comparable a la que produïa el gramòfon. La idea de fer servir la síntesi com una ajuda a l'anàlisi del senyal s'ha demostrat que és una gran aproximació al problema (Poulton, 1986).

El 1923 C. Paget va descobrir que hi havia components freqüencials en tots els sons vocàlics i en va fer una taula a través de l'observació, tot i que ara per ara s'accepta que n'hi ha més dels que va observar. Va construir un sistema semblant al dels ressonadors acústics en forma de tub de Kratzenstein, però en aquest cas estaven fets d'argila i, individualment, tots podien produir sons vocàlics i consonàntics. Va ser en aquella època quan es van començar a provar de desenvolupar els primers sintetitzadors de veu elèctrics.

El primer dispositiu de síntesi de veu completament elèctric el va crear J.Q. Stewart l'any 1922 (Klatt, 1987). El sintetitzador comptava amb un interruptor que simulava les cordes vocals i un seguit de circuits que simulaven les ressonàncies de les cordes vocals; i un altre interruptor tallava o permetia el pas del corrent d'aire (com les cordes vocals). L'aparell era capaç de generar sons vocals estàtics simples amb dos formants baixos, però no consonants ni frases. A més a més, Stewart, com Paget, va treballar en la teoria de l'existència de dos circuits ressonants separats que estaven compostos per un nombre variable de bobines i de condensadors. D'aquesta manera, un dispositiu de variables controlava l'amortiment i les intensitats relatives de tots dos ressonadors i el sintetitzador generava sons força semblants als de la veu humana.

Wagner va fer també un sintetitzador del mateix tipus (Flanagan, 1972) que consistia en quatre ressonadors elèctrics connectats en paral·lel als quals exercitava un brunzit. Les sortides dels quatre ressonadors es combinaven i produïen, també, sons vocàlics. L'any 1932 els investigadors japonesos Obata i Teshima van descobrir un tercer formant de les vocals (Schroeder, 1993). Els tres primers formants es consideren suficients per produir una veu sintètica intel·ligible.

El primer dispositiu que es va considerar un sintetitzador de veu va ser el VODER (Voice Operating Demonstrator) que va presentar H. Dudley a la Fira Mundial de Nova York del 1939. Per fer servir el dispositiu (Figura 3), calia un entrenament d'aproximadament un any, atès que l'operador manipulava 14 claus amb els dits per controlar l'estructura de tracte vocal que ressonava i un pedal de peu dret que permetia que els tons variessin. Un cop se sabia com fer-lo servir, es podia produir un discurs intel·ligible continu. El VODER tenia dues fonts de so: l'oscil·lador que generava un brunzit periòdic com el de Stewart per als sons sonors, i un soroll aleatori per als sons sords. La secció de ressonadors era més complexa que la de Stewart i contenia 10 filtres que comprenien les freqüències de senyal. La qualitat de la parla del VOCODER no era la millor, però oferia molt de potencial a l'hora de produir-ne.

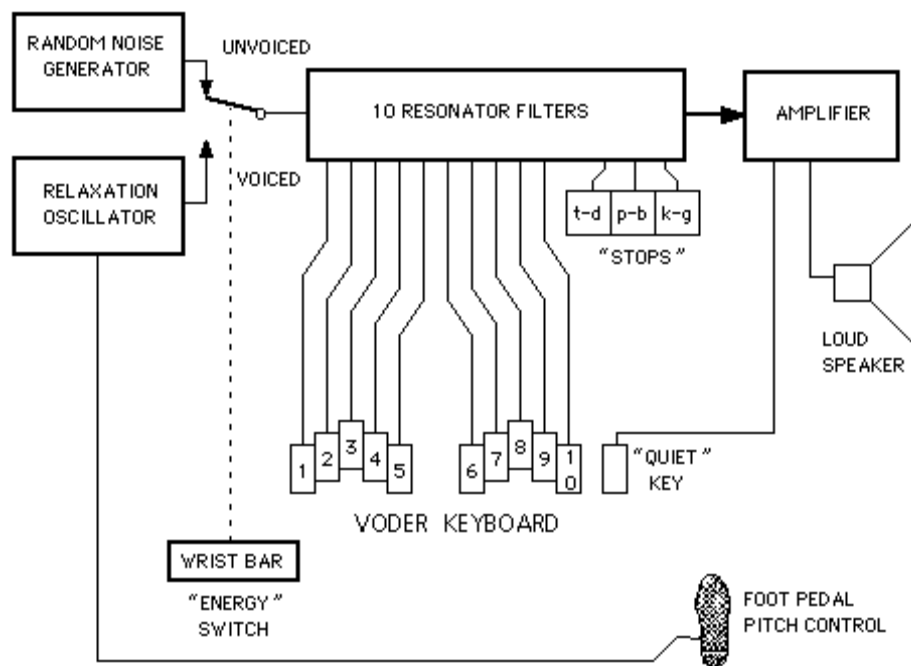


Figura 3. Diagrama esquemàtic del sintetitzador VODER

Paral·lelament, va sorgir el VOCODER 1939 (Voice Coder), un compressor de banda ampla de telefonia. El senyal de veu (en aquest cas, real) es codificava de manera que es podia transmetre a l'altra banda de la línia telefònica i a través de filtres, separava les diferents bandes de freqüència del senyal. Per caracteritzar el senyal acústic original se seguien uns paràmetres que detectaven si el so era sord o sonor i també es mesurava el to fonamental de vibració de les cordes vocals, l'amplitud de senyal de cada banda, etc. que es transmetien per la línia telefònica. L'aparell receptor feia l'operació inversa amb un demultiplexor i es reconstruïa el senyal. No obstant això, l'equipament que calia per dur a terme aquesta operació era molt car en aquella època, motiu pel qual el VOCODER només es va fer servir per a casos molt especialitzats.

En general, després de l'aparició del VODER el món científic es va interessar, cada cop més, en la síntesi de la veu i, finalment, es va demostrar que es podia produir la parla intel·ligible artificialment. De fet, l'estructura bàsica i la idea de VODER són molt semblants a les dels sistemes actuals.

L'any 1822, el matemàtic S.B.J. Fourier va donar a conèixer el seu teorema que totes les ones complexes es descomponen en elements sinusoidals simples i que es podia analitzar qualsevol mena de so. A principis del segle XX, abans de la Primera Guerra Mundial, Siemens ja va ser capaç de descompondre els elements. Aquell espectròmetre separava els harmònics d'ones complexes i, més tard, durant la Segona Guerra Mundial, van ser els Estats Units els qui van perfeccionar l'espectrògraf (anomenat Sona-graph i desenvolupat als laboratoris Bell als anys quaranta). En la mateixa època apareixen les primeres descripcions de l'aparell i les primeres mostres dels espectrogrames, un llibre que, com el seu mateix nom indica, '*Visible Speech*', tenia la intenció de convertir la parla en visible.

L'espectrògraf dibuixa la distribució de l'energia en el domini del temps i de la freqüència que s'anomena espectrograma, però aquest dibuix requereix unes dimensions que el paper no té (només en té dos). Així doncs, en el paper, el temps el representa l'eix horitzontal i la freqüència, el vertical. Quan hi ha períodes de silenci no s'hi dibuixa res i quan hi ha sons de mitjana intensitat hi apareixen tonalitats grises (Varona, 1997).

Al llarg dels anys cinquanta, els fonòlegs van començar a estudiar la producció de l'aparell fonador humà i a sistematitzar els resultats a partir dels quals es van extreure les teories fonològiques. Paral·lelament, H.K. Dunn va aconseguir millorar la síntesi de veu del VODER. El sistema que va desenvolupar es basava en una font d'energia elèctrica que simulava les cordes vocals i un model de línies de transmissió que representaven el tracte vocal (condensadors i bobines). Amb filtres passabaix es retardava l'ona sonora a través de la cavitat bucal. Dunn, gràcies a imatges de rajos X va aconseguir calcular freqüències de ressonància aproximant la forma del tracte vocal a cilindres. Els resultats coincidien amb les mides experimentals de l'espectrograma (en els tres primers formants). El to del so d'aquest sintetitzador només es podia canviar manualment.

No obstant això, el 1958 G. Rosen, va crear el primer sintetitzador que es podia controlar dinàmicament. Es basava en el mateix sistema que Dunn (feia servir condensadors i bobines) i es controlava el dispositiu amb un retard que seleccionava una seqüència de configuracions de tracte vocal. Les transicions entre una configuració o l'altra se suavitzaven electrònicament. Com l'espectrograma d'un senyal acústic és adequat i contempla les característiques importants de la veu, és evident que s'hauria de poder reconstruir la veu. Els sintetitzadors que es basaven en l'estudi de formants també es podien controlar dinàmicament. L'any 1962, J. Anthony i W. Lawrence van representar el primer sintetitzador que contenia tres filtres de formants variables i una de fixa, connectades. Aquest sintetitzador estava control·lat per un dispositiu electromecànic que llegia l'esquema de formants dibuixat amb tinta conductora.

De la mateixa manera, els laboratoris Haskins de Nova York van desenvolupar una síntesi del llenguatge de manera completa a partir d'espectrograma en l'anomenat reproductor de patrons (Pattern Playback) de Haskins (Klatt 1987, Flanagan et al. 1973). El Playback consisteix a dibuixar sobre una banda de cel·luloide transparent un espectrograma inspirat en un real o un d'inventat. Com el coneixement de l'anàlisi espectrogràfica permet conèixer els elements i les parts principals de cada so, quan es dibuixaven es podia aconseguir que l'aparell les pogués llegir i les pronunciés. La banda de cel·luloide passa a una velocitat adequada per un sistema de cèl·lules fotoelèctriques i de vibradors que en cada moment barreja els components amb les amplituds

proporcionals al dibuix dels espectres que apareixen sobre la banda. D'aquesta manera, reconstrueix les fluctuacions dels objectes esquematitzats al llarg del temps, els amplifica i els pronuncia a través d'un altaveu o un magnetòfon, Martínez [1986].

Un altre sintetitzador de veu que es pot controlar de manera dinàmica i que també es basa en espectrograma però que no necessita dibuixar-lo és el PAT (Parametric Artificial Talker) que es mostra a la Figura 4. És un anàleg acústic de l'aparell fonador humà en què uns circuits elèctrics ressonen quan els mou un estímul acústic semblant a la vibració de la laringe, semblant al que passa a les cavitats de ressonància de l'aparell fonador humà.

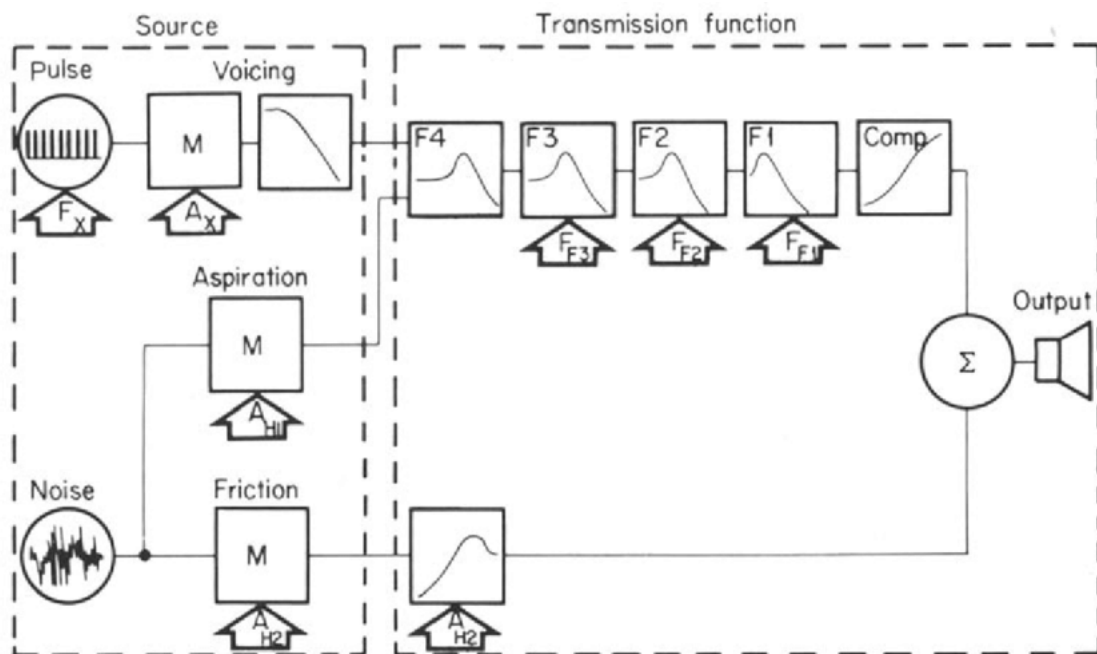


Figura 4. Esquema del PAT

Per imitar l'acústica d'un so calen dues fonts d'energia: una de so periòdic i l'altra de so aperiòdic. Totes dues fonts es controlaven a partir de vuit paràmetres variables que produïen els diferents sons consonàntics: amplitud laríngia, freqüència laríngia (f_0), primer formant dels sons amb característiques vocàliques, segon formant, tercer formant, freqüència de soroll, amplitud de soroll i soroll dels formants. La informació necessària per a cadascun dels vuit paràmetres s'aconseguia a través de l'espectrograma de dues maneres diferents. La primera, restituint amb tinta conductora els paràmetres,

sobre un paper calibrat sobre el mateix espectrograma. La segona, fent servir els comandaments i proporcionant les dades que calgui comprovar.

Un aparell semblant n'és l'OVE II, que va crear G. Fant, entre d'altres, l'any 1956 als laboratoris de transmissió de parla d'Estocolm. El posterior OVE III, el controlava un computador digital CDC 1700. El computador s'encarregava de variar els paràmetres a través de convertidors digitals/analògics. Es va guardar en la memòria del computador una llibreria que contenia tots els paràmetres per a vocals i consonants sueques. Gràcies a l'efectivitat que van aconseguir els computadores digitals, es va fer palès que s'havien de fer processos de síntesi complets i no només tasques de control.

En la mateixa època hi va haver grans avenços en programació com l'algoritme de la transformada ràpida de Fourier (FFT) que va permetre fer l'espectrograma de manera molt més ràpida. Per sintetitzar una paraula cal una gran quantitat de càlculs i si es pretén que el sistema funcioni a temps real, és a dir, al ritme normal de la llengua parlada, hi ha un temps limitat per fer operacions. Hi ha dues solucions: o bé fer que el computador sigui més ràpid o bé fer programes més eficients (Brigham, 1974).

El 1975, es va llançar MUSA, un dels primers sistemes de síntesi de veu. Consistia en un *hardware* d'ordinador independent i un *software* especialitzat que li permetia llegir en italià. L'any 1978 es va llançar una segona versió que també era capaç de llegir l'italià. Els sistemes dominants entre 1980 i 1990 van ser els DECTalk, basats en el treball de Dennis Klatt al MIT, i el sistema Bell Labs. Aquest últim sistema va ser un dels primers sistemes multilingües independents de llengua, que va fer un ampli ús dels mètodes de processament del llenguatge natural. L'electrònica de mà amb síntesi de veu va començar a sorgir l'any 1970. Un dels primers va ser el Telesensory Systems Inc. (TSI) Speech+, una calculadora portàtil per a cecs, el 1976. Hi va haver també altres dispositius amb finalitats educatives, como la joguina Speak & Spell que va produir Texas Instruments el 1978. Fidelity va presentar una versió parlant del seu ordinador d'escacs electrònic al 1979. El primer videojoc amb síntesi de veu va ser el joc d'*arcade* del 1980, Stratovox (al Japón, Speak & Rescue), de Sun Electronics. El primer joc d'ordinador personal amb síntesi de veu va ser Manbiki Shoujo (Shoplifting Girl), que es va presentar al 1980 per al PET 2001. El desenvolupador del joc Hiroshi Suzuki, va desenvolupar una tècnica de programació 'creuada per zero' per produir una

forma d'ona de veu sintetitzada. Un altre exemple va ser la versió d'*arcade* de Berzerk, també del 1980. L'empresa Milton Bradley va produir el primer joc electrònic multijugador amb síntesi de veu el mateix any: Milton.

Els primers sintetitzadors electrònics de veu sonaven robòtics i sovint eren intel·ligibles. La qualitat de la parla sintetitzada ha millorat, avui dia, considerablement, per a partir del 2016 la producció dels sistemes de síntesi de veu continua sent clarament distingible de la veu humana real, com és el cas dels assistents de veu dels quals parlarem més endavant.

2.1.2. El reconeixement de veu (speech to text)

A continuació, com en l'anterior apartat, explicarem de manera cronològica l'evolució de la tecnologia del reconeixement de veu o, en altres paraules, de com, al llarg del temps, s'han inventat diferents màquines capaces de plasmar textualment les paraules orals, és a dir, la parla. Cal tenir en compte, però, que aquest és un mètode més complex, atès que presenta diversos reptes. Quan tenim una conversa o parlem en general, normalment, no fem gaires pauses (que representarien, de manera escrita, les comes, punts, espais, etc.) i sovint, repetim paraules o ens autocorregim. També cal tenir en compte que existeixen aspectes com les varietats dialectals, el registre, la pronunciació o el to de veu, que fan que els problemes per dur a terme el reconeixement de la veu s'hagin vist agreujats en el transcurs de la història.

Al contrari que en el desenvolupament dels sistemes de síntesi en què primer es van desenvolupar els sistemes mecànics i elèctrics, en el reconeixement de veu els principals avenços es van aconseguir amb l'aparició dels ordinadors en els anys setanta. No obstant això, hi va haver alguns sistemes elèctrics a principis del segle XX.

El que ens permet fer un sistema de reconeixement de veu és comunicar-nos amb una màquina de manera oral. De fet, podem considerar el sistema de síntesi de veu com una boca artificial, i el sistema de reconeixement de veu com una oïda artificial. Si pensem en com funcionen aquests sistemes, la resposta pot semblar fàcil: reben un senyal acústic i actuen en conseqüència. No obstant això, el seu funcionament va més enllà. Sovint no es limita a transcriure un missatge oral en format textual, sinó que pot,

fins i tot, entendre i obeir una ordre pronunciada de manera natural per un humà i executar determinades accions.

L'any 1916, Flowers va ser la primera persona a dissenyar una màquina capaç de transcriure veu gràcies al seu coneixement sobre la transmissió de missatges per cable entre submarins, que es transmetien mitjançant un codi alfabètic especial que registrava una línia ondulada sobre paper que després es transcrivia. Saber això va portar Flowers a desenvolupar una màquina que podia convertir els sons en ones com les que es transmetien entre submarins, i les va anomenar alfabet fonogràfic. Crear aquest dispositiu (Figura 5) també va fer que s'adonés que existien tons a diverses freqüències del senyal acústic. És especialment important mencionar que aquesta primera màquina de Flowers no feia servir cap mena de component elèctric. Consistia de dos electroimants i dos condensadors que cobrien tota l'audiofreqüència (Varona, 1997).

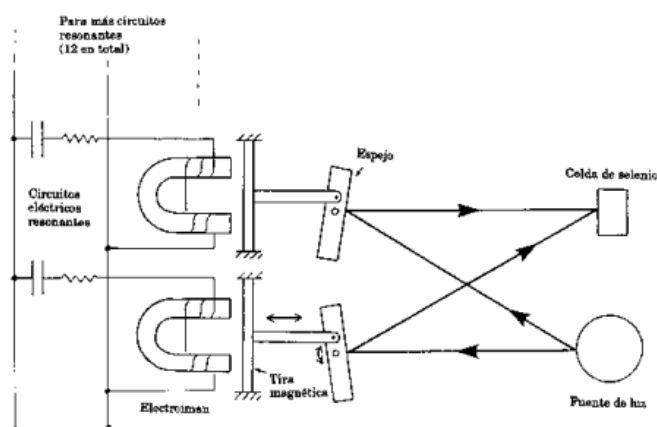


Figura 5. Disseny de la primera màquina que permetia transcriure veu l'any 1916.

Segons Poulton (1983), l'any 1930, les autoritats patents alemanyes van desestimar atorgar una patent al T. Nemes per al seu transcriptor fonètic, ja que consideraven que era impossible fer un transcriptor fonètic. Nemes proposava un dispositiu òptic i, tot i que va demostrar que la seva teoria tenia sentit, no va poder desenvolupar-la a causa de la guerra.

Els primers avenços en el reconeixement de la parla comencen als anys quaranta, quan es va desenvolupar un dispositiu capaç de visualitzar el senyal acústic sobre paper: l'espectrògraf. Aquest dispositiu permetia registrar l'energia de les bandes de freqüència

d'una paraula o frase en funció dels temps (espectrograma). A partir de llavors, és quan es va començar a creure realment que es podien fer sistemes de reconeixement de la parla.

L'any 1952, Davis, Biddulph, i Balashek, dels Laboratoris Bell, van construir un sistema elèctric que reconeixia els deu dígit anglesos pronunciats de manera aïllada per un únic locutor. Dos filtres dins de dues bandes de freqüència separaven el senyal de veu segons la seva freqüència (més de 900 Hz i menys). La seqüència principal a cada banda es localitzava a partir de contar el nombre d'encreuaments per zero, cosa que determinava la freqüència amb què el senyal canviava de polaritat. Això significa que el senyal complet es trencava en dos senyals que representaven la freqüència dels dos primers formants i variaven de manera suau quan es pronunciaven. Aquests senyals s'aplicaven als canals X i Y en un oscil·loscopi (Davis, 1952).

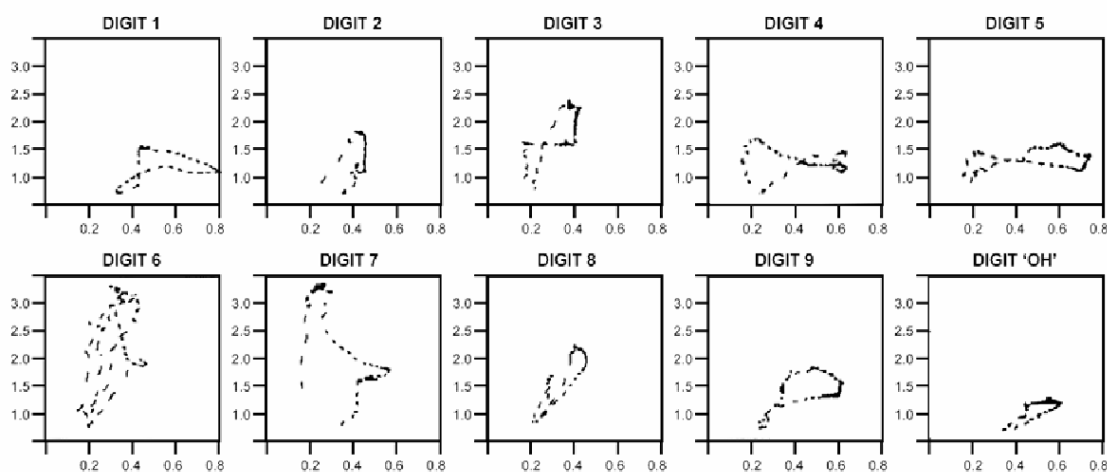


Figura 6. Imatges de la representació dels 10 dígitos en formants.

Quan es pronunciava un dígit a través del micròfon, es dibuixava una corba de dues dimensions en la pantalla de l'oscil·loscopi, que eren força diferents per cadascun dels dígitos anglesos. Per comparar aquestes corbes era necessari quantificar-les d'alguna forma, de manera que es va dividir el dibuix en dues dimensions dins de quadres i es mesurava la part de la corba que ocupava cada quadre. Hi havia 28 quadres per representar les diferents combinacions dels formants 1 i 2. Aquests quadres estaven definits per un circuit de vàlvules que activava un dels 28 relés segons la forma del dibuix. Cada relé corresponia a un dels quadres en què s'havia dividit la pantalla de l'oscil·loscopi, i en funció dels relés activats s'encenia un dels indicadors corresponents

a un número reconegut. Les mesures dels patrons es van fer amb 100 repeticions de cadascun dels 10 dígit i s'obtenien taxes de reconeixement entre un 97% i un 99%.

Segons Rabiner (1993), paral·lelament i de manera independent, es va desenvolupar un altre sistema als Laboratoris RCA. Olson i Bellar van desenvolupar una màquina d'escriure fonètica. El senyal que s'obtenia a través del micròfon passava a través d'un amplificador i un compressor i s'aplicava a un banc de 8 filtres. D'aquesta manera, es pretenia ajustar el valor mitjà del senyal, perquè fos semblant entre els locutors que parlaven en un to alt i un to baix. Els filtres s'ajustaven per calcular l'envolupant. Aquesta informació es transmetia a un banc de relés mitjançant un interruptor que actuava cada 40 ms. des del moment en què començava la pronunciació. S'obtenia, d'aquesta manera, un dibuix característic per cadascun dels diferents sons a causa de les diferents configuracions dels relés.

Olson i Bellar es van preguntar des d'un primer moment quines unitats lèxiques seria millor reconèixer: fonemes, síl·labes o paraules. En un primer moment van provar amb tots tres conjunts: amb les transcripcions de fonemes s'obtenien resultats poc ajustats a la realitat; amb les síl·labes milloraven; i amb les paraules completes la transcripció era perfecta (si es reconeixia correctament la paraula). No obstant això, per a vocabularis raonables hi havia moltes paraules amb molts fonemes i síl·labes, de manera que si es feien servir paraules, la quantitat de memòria per desar les característiques havia de ser força gran. Així doncs, es van decidir per desenvolupar un sistema de reconeixement basat en síl·labes. Les dades que proporcionaven els relés (memòria espectral) es descodificaven en síl·labes i després en lletres individuals per actuar sobre les tecles de la màquina d'escriure. El sistema es va provar amb frases constituïdes per 10 síl·labes en diverses permutacions diferents. Si l'entonació de la frase era cuidada, resultava un percentatge d'encert d'un 98%. Van continuar els seus experiments i van arribar a analitzar paraules aïllades traduint-les a diferents llengües. El sistema reconeixia paraules angleses i franceses i les podia traduir a l'anglès, al francès, a l'alemany i al castellà.

Es van desenvolupar altres treballs en l'època basats en dispositius analògics, com el de Wiren i Stubbs l'any 1956, que obtenia contingut espectral dels senyals i feien servir com a criteri de classificació la freqüència de ressonància de les vocals.

Utilitzaven un arbre binari de selecció, de manera que en cada fulla es feia una separació. Per exemple, si en una fulla de l'arbre se separaven els fonemes sords i sonors, en la següent inferior en què totes les unitats eren sordes, se separaven en fricatives o no, etc. Els circuits necessaris per fer aquestes separacions incloïen una gran quantitat de vàlvules perquè el transistor era un element massa nou per a l'època.

A més dels Laboratoris RCA i els Laboratoris Bell, alguns laboratoris japonesos també van demostrar la seva capacitat de construir equips per dur a terme el reconeixement de veu. Per a l'any 1970 ja s'havien fet una gran quantitat de progressos en els nivells bàsics del reconeixement automàtic de veu. Per exemple, Reddy, l'any 1967 va descriure un sistema que consistia en dos ordinadors connectats entre ells (un IBM 7090 i un PDP1). Aquest sistema identificava els fonemes vocàlics i consonàntics, però el principal problema era com aplicar els coneixements de nivell lingüístic superior (fonètics, sintàctics, semàntics i pragmàtics) per ajudar a la comprensió del missatge pronunciat. Fant, als setanta, va suggerir un model en què la tasca del reconeixement es dividia en una seqüència de 5 passos: extracció de paràmetres, detecció de segments, transcripció fonètica, identificació de paraules i interpretació semàntica. Al contrari, el sistema HEARSY (de les paraules angleses HEAR=escoltar i SAY=dir) que va proposar Reddy estava basat en un model de reconeixement en paral·lel que incloïa tres reconeixedors per separat: acústic, sintàctic i semàntic. El HEARSY I era una continuació de l'anterior i també contenia els tres reconeixedors independents. La frase es processava per extraure les seves característiques i el resultat es passava als tres mòduls. Cada reconeixedor feia les seves pròpies hipòtesis tot basant-se en els diferents nivells de coneixement i, al final, un sistema mediador que tenia tota la informació controlava el sistema i donava els resultats. També a mitjans dels anys setanta, Itakura, Rabiner i Levinson van proposar aplicar la tecnologia de reconeixement dels patrons fonamentals al reconeixement de la veu. En aquest període Martin va fundar la primera companyia comercial de reconeixement de veu anomenada Threshold Technology, Inc. i va desenvolupar el primer producte real de reconeixement automàtic de veu: el sistema VIP-100. Aquest sistema només es va fer servir en aplicacions simples, com en empreses de fabricació de plaques frontals de televisió (per al control de qualitat) i per FedEx (per classificar paquets en una cinta transportadora), però va ser important perquè va influir en el desenvolupament del programa ARPA-SUR, que veurem a

continuació. Després d'això hi va haver un optimisme generalitzat entre els investigadors i els organismes inversors per arribar en poc temps a desenvolupar sistemes capaços de reconèixer de forma precisa gairebé qualsevol frase, pronunciada per un locutor qualsevol de manera contínua. Amb aquest objectiu en ment, es van impulsar dos grans projectes d'investigació en els quals es pretenia arribar a les menors restriccions gramaticals possibles de les frases a reconèixer, així com del lèxic utilitzat. Es van començar a desenvolupar aquests projectes a diferents països, però és als Estats Units on es fa el projecte més gran fins llavors del reconeixement automàtic de la veu. L'anomenat Advanced Research Projects Agency - Departament de Defensa - Speech Understanding Research (ARPA-SUR) amb un pressupost de quinze milions de dòlars i una duració de cinc anys.

Un grup d'experts dirigits pel professor Newell proposa una llarga llista d'especificacions. El sistema havia d'acceptar un discurs continu, permetre una adaptació fàcil a nous locutors i el diccionari havia de ser de 10000 paraules amb una sintaxi artificial adequada a una tasca restringida. El sistema havia de gestionar 100 milions d'instruccions per segon i estar llest per al 1973. Abans que es fessin totes les seves deliberacions, es van fer algunes modificacions: es va passar a exigir un vocabulari de 1000 paraules en comptes de 10000 i es va ampliar el termini de presentació dels resultats fins al 1976.

Tot i que els objectius no es van obtenir per complet, les aportacions del projecte ARPA-SUR van contribuir de manera molt notable a conèixer millor les propietats de la parla i les limitacions dels sistemes de reconeixement automàtic de la veu, així com a conscienciar-se de la necessitat d'investigar més en el camp per resoldre les limitacions. Principalment hi havia dues aproximacions al problema: els models estructurals estocàstics i els models ocults de Markov. Amb aquesta aproximació la correlació entre les característiques espectrals pròpies del senyal acústic i les corresponents unitats sublèxiques s'aprenia de manera automàtica a partir d'un ampli conjunt de mostres d'aprenentatge, i els sistemes basats en el coneixement en els quals tot el coneixement que s'aporta al sistema es basa en regles sintàctiques, gramaticals, etc., donades o obtingudes a través d'un estudi del problema concret.

El sistema que més es va aproximar als objectius proposats va ser el HARPY de la Universitat Carnegie-Mellon i desenvolupat per Klatt. El sistema HARPY modelava totes les fonts de coneixement (fonològica, lèxica, sintàctica i semàntica) en una única xarxa d'estats finits, que s'obtenia prèviament a partir de subxarxes que modelaven les diferents paraules. Tot i que el seu antecessor, el sistema DRAGON, feia servir un model de Markov i els mètodes associats, el sistema HARPY no es basava explícitament en cadenes de Markov, sinó que assumia una estructura en xarxa particularitzada al problema específic, i introduïa una important modificació a l'algorisme de Viterbi de reconeixement: la cerca en feixos (Beam Search). Aquest sistema va aconseguir una taxa d'errors semàntics d'aproximadament el 5%, amb un vocabulari de 1011 paraules, una sintaxi artificial de baixa perplexitat, amb adaptació al locutor, i amb un temps de resposta d'aproximadament 80 vegades el temps real.

La Universitat Carnegie-Mellon va desenvolupar, també en el marc de l'ARPA-SUR, un altre sistema totalment diferent; l'EARSAY-II, una continuació de l'EARSAY-I. Aquest sistema va introduir l'arquitectura pissarra (*blackboard*) i el concepte de font de coneixement, com a processos paral·lels, independents, cooperatius i asíncrons. Aquests processos els controlava un planificador, i es comuniquen entre si a través d'una estructura de dades global (la pissarra) on s'annotaven totes les hipòtesis emeses. Les fonts de coneixement s'autoactivaven per la satisfacció de certes condicions d'estat de la pissarra, especificades en estímuls (*stimulus frame*), que permetien fer les accions especificades gràcies a les seves respostes corresponents (*response frame*). Les respostes es dedicaven a introduir o actualitzar informacions de la pissarra, sobre les que tenien competència. Aquest sistema va aconseguir taxes d'error del 9% al nivell purament semàntic, i del 26% al nivell sintàctic i semàntic, amb un consum de recursos de 3 a 4 vegades superior que el HARPY. No obstant això, els mèrits del HEARSY-II no s'han de valorar en base a aquests resultats, sinó que s'han de valorar les seves aportacions a l'arquitectura de sistemes complexos intel·ligents, anomenats sistemes basats en el coneixement. També dins del projecte ARPA-SUR, es va desenvolupar el sistema HWIM (Hear What I Mean) de Bolt Beranek and Newman Inc. L'HWIM estava basat en una filosofia (no arquitectura) similar al HEARSY-II, però posant èmfasi en aconseguir unes estratègies d'hipòtesi i uns resultats òptims. Aquest sistema constava de quatre mòduls bàsics: el processador acústic, el mòdul d'accés lèxic, el component

lingüístic i el mòdul de control. El processador acústic estava en contacte directe amb el mòdul d'accés lèxic, que s'encarregava de subministrar, a petició del mòdul de control, hipòtesis sobre quines eren les millors paraules en un segment del senyal, amb una mesura de la qualitat de la comparació. El component lingüístic s'encarregava de determinar, també a petició del mòdul de control, si una seqüència de paraules es podia interpretar com una subseqüència d'alguna sentència sintàctica, semàntica o pragmàtica correcta. L'objectiu principal del mòdul de control era descobrir quina seqüència òptima de paraules cobria la totalitat del senyal vocal, i es podia acceptar també pel component lingüístic. Els resultats que es van aconseguir amb aquest sistema quan es va acabar el projecte ARPA-SUR, van ser força pitjors dels que van obtenir els sistemes anteriors (56% d'error semàntic amb un consum de recursos aproximadament 200 cops més gran que el HARPY). Paral·lelament als esforços que va iniciar ARPA, als anys setanta van començar a prendre forma dos grans direccions en la investigació sobre el reconeixement de veu en què IBM i AT&T Bell Laboratories representaven les dues escoles de pensament que tenien objectius comercials diferents.

L'esforç d'IBM, dirigit per Fred Jelinek, tenia com a objectiu crear una màquina d'escriure activada per veu, de manera que es pretenia convertir una frase parlada en una seqüència de lletres i paraules que es poguessin mostrar o escriure en paper. El sistema de reconeixement, anomenat Tangora, era un sistema dependent de la persona que parlava, és a dir, cada usuari individual havia d'entrenar la seva pròpia màquina d'escriure. L'atenció tècnica se centrava en el volum de vocabulari que reconeixia (com més gran millor) i en l'estructura del model de llengua (la gramàtica), que es representava amb regles sintàctiques estadístiques que descrivien quina era la probabilitat que els símbols lingüístics (per exemple fonemes o paraules) apareguessin en el senyal de veu. Aquest fenomen és el que coneixem com transcripció. El conjunt de regles estadístiques gramaticals o sintàctiques es denominava model de llengua, del qual el model d'*n-grames*, que definia la probabilitat que aparegués una seqüència ordenada de paraules, era la variant més utilitzada. Tot i que tant el model de llenguatge d'*n-grames* com una gramàtica tradicional són manifestacions de les regles de la llengua, les seves funcions eren diferents. El model d'*n-grames*, que caracteritzava la relació de les paraules en un lapse d'*n* paraules, era una representació estadística de la gramàtica. La seva eficàcia es va fer palesa gràcies a un joc de paraules de Claude Shannon: una

competició entre un humà i un ordinador. En aquesta competició es demanava a l'humà i a la màquina endevinar la següent paraula en una frase arbitrària. L'humà endevinava segons la seva experiència nativa de la llengua i l'ordinador, gràcies a les estadístiques de paraules acumulades, de manera que triava segons la probabilitat de freqüència de les paraules i es va demostrar que la màquina podia guanyar l'humà. Des que es van introduir als vuitanta, els models de llenguatge d'*n-grames* i les seves variants es van convertir en indispensables per als sistemes de reconeixement de veu amb grans vocabularis.

Als Laboratoris Bell de AT&T, l'objectiu del programa d'investigació era proporcionar serveis de telecomunicació automatitzats al públic, com el marcatge per veu i el comandament i el control de l'encaminament de les trucades telefòniques. S'esperava que aquests sistemes automatitzats funcionessin bé per la majoria de la població sense haver d'entrenar-los individualment. Així doncs, el focus dels Laboratoris Bell estava en el disseny d'un sistema independent d'altaveus que fossin capaços de tractar amb la variabilitat acústica en els senyals de la parla de molts parlants diferents, sovint amb diferents accents diferents. Això va causar que es creessin algoritmes d'agrupació de la veu per crear patrons de referència de paraules i sons (en un principi plantilles però finalment models estadístics) que es podien fer servir en una ampla gamma de parlants i accents. A més, les investigacions per comprendre i controlar la variabilitat acústica de diverses representacions de la veu van portar al fet que es fes un estudi d'un seguit de mesures de distància espectral i tècniques de modelatge estadístic que produïen unes representacions suficientment bones de les expressions de la població. Com les seves aplicacions (marcatge de veu i encaminament de trucades) només implicaven pronunciacions curtes de vocabulari limitat, es va posar especial atenció en la investigació del model acústic (la representació espectral dels sons o les paraules) i no en el model de la llengua (la sintaxi i la gramàtica). En aquests laboratoris també va ser de gran importància el concepte de la detecció de paraules clau com una manera primitiva de comprensió del discurs. Aquesta tècnica de localització de paraules clau tenia com a objectiu detectar una paraula clau o una frase clau amb algun significat particular. La necessitat d'aquesta detecció de paraules clau era incloure els parlants que preferien parlar amb frases naturals en lloc de fer servir seqüències rígides a l'hora de sol·licitar serveis (és a dir, parlar com si ho estiguessin fent amb un operador

humà). En una aplicació de domini limitat, la presència d'una paraula o frase clau era generalment suficient per indicar la intenció de la persona. La paraula o frase clau detectada desencadenaria una acció prescrita (o una seqüència d'accions) com a resposta a l'expressió oral de l'interlocutor. La tècnica de localització de paraules clau requeria l'ampliació del paradigma habitual de reconeixement de patrons a un que donés suport a la comprovació de la hipòtesi.

Les aproximacions d'IBM i AT&T Bell Laboratories al reconeixement de veu, van tenir una gran influència en l'evolució de la tecnologia de comunicació persona-màquina. Tot i les seves diferències, el desenvolupament accelerat dels mètodes estadístics dels anys vuitanta van fer que tots dos sistemes coincidissin.

La investigació sobre el reconeixement de veu en els anys vuitanta es va caracteritzar per un canvi en la metodologia: de la intuïtiva i basada en plantilles (un paradigma directe de reconeixement de patrons) a una modelació estadística més rigorosa. Tot i que la idea bàsica del model de Markov (HMM) es va entendre i adoptar per la majoria de laboratoris (per exemple IBM), la metodologia no es va completar fins a mitjans dels anys vuitanta i fins després de la publicació generalitzada de la teoria, el model de Markov no es va convertir en el mètode preferit per al reconeixement de la veu. La popularitat i l'ús de l'HMM com la base principal per al reconeixement automàtic de veu i els sistemes de comprensió van continuar constants en les dues dècades següents a causa del constant flux de millores i refinaments de la tecnologia.

El model ocult de Markov, que és un procés doblement estocràtic, modela la variabilitat intrínseca del senyal de veu (i les funcions espectrals que en resulten) i també l'estructura de la llengua parlada en un marc estadístic. Com és sabut, un senyal de veu realista és inherentment molt variable (pronunciació, accent, factors ambientals, soroll, etc.). Quan la gent pronuncia una mateixa paraula, els senyals acústics no són idèntics tot i que l'estructura lingüística subjacent (la sintaxi i la gramàtica) pot ser la mateixa. El formalisme de l'HMM és una probabilitat mesurada que fa servir una cadena de Markov per representar l'estructura lingüística i un conjunt de distribucions de la probabilitat per considerar la variabilitat acústica dels sons de l'enunciat. Amb un conjunt d'enunciats coneguts (text etiquetat) que representen una col·lecció suficient de les variacions de les paraules d'interès (conjunt d'entrenament), es pot fer servir una

estimació eficient anomenada algoritme Baum-Welch, per obtenir el millor conjunt de paràmetres que defineixen el o els models corresponents. L'estimació dels paràmetres que defineixen el model és equivalent a la formació i l'aprenentatge. El model resultant es fa servir llavors per indicar la probabilitat que un enunciat desconegut sigui efectivament la paraula o paraules que es representen en el model. La mesura de la probabilitat representada pel model ocult de Markov és un component essencial d'un sistema de reconeixement de veu que segueix un patró estadístic basat en la teoria de la decisió de Bayes. La metodologia HMM va representar un gran pas des del simple reconeixement de patrons i els mètodes acústics i fonètics que es feien servir abans en els sistemes de reconeixement automàtic de la veu.

La idea del model ocult de Markov sembla haver sorgit per primera vegada a finals dels anys seixanta a l'Institut d'Anàlisi de Defensa (IDA). Baum es va referir a l'HMM com un conjunt de les funcions probabilístiques d'una cadena de Markov que, per definició, implica dues distribucions que van de la mà: una que pertany a la cadena de Markov i l'altra a un conjunt de distribucions de probabilitat, cadascuna associada a un estat de la cadena de Markov, respectivament. El model HMM mira d'abordar les característiques d'una seqüència probabilística d'observacions que poden no ser una funció fixa sinó que canvia segons una cadena de Markov. Aquest procés doblement estocàstic va resultar útil en diverses aplicacions com la predicció del mercat de valors i la criptoanàlisi d'un xifrat rotatori, que es va fer servir durant la Segona Guerra Mundial. La tècnica de modelització i estimació de Baum va demostrar que funcionava per a observacions discretes (és a dir, les que suposen un conjunt finit i, per tant, es regeixen per distribucions de probabilitat discretes) i més tard per a observacions aleatòries que es van modelar amb funcions de densitat de probabilitat còncaues. Tot i que aquesta tècnica funcionava bé, era limitada, de manera que Liporace, també de l'IDA, va relaxar la densitat per incloure una restricció de densitat simètrica el·líptica amb l'ajuda del teorema de representació de Fan. El procés doblement estocàstic de Baum va començar a aplicar-se en l'àmbit de la veu, al principi en sistemes d'identificació de veu. Com més gent feia servir la tècnica HMM, més evident es va fer la limitació de la forma de les funcions. L'HMM, sent una mesura de probabilitat, era susceptible de ser incorporada en un mar més ampli de descodificació de la veu que incloïa un model de llengua. L'ús d'una gramàtica d'estat finit el reconeixement de veu

de grans vocabularis representava una extensió consistent de la cadena de Markov que l'HMM va fer servir per considerar l'estructura de la llengua. Tot i que aquestes estructures (per diversos nivells de restriccions de la llengua) eren, en el millor dels casos, aproximacions al discurs eren eficients des del punt de vista informàtic i sovint suficients per obtenir resultats de rendiment raonables. La fusió del model ocult de Markov (amb l'avantatge de coherència estadística, especialment en la gestió de la variabilitat acústica) i la xarxa d'estats finits (amb la cerca i eficiència de càlcul, particularment en la gestió d'hipòtesis de seqüències de paraules) van constituir el desenvolupament tecnològic dels anys vuitanta.

Una altra teoria que es va introduir a finals dels anys vuitanta va ser la idea de les xarxes neuronals artificials (XNA). Les xarxes neuronals es van introduir per primer cop als anys cinquanta, però no van produir bons resultats inicialment. Amb l'arribada d'un model de processament distribuït en paral·lel (PDP) als anys vuitanta, que era una densa interconnexió d'elements computacions simples, i un mètode d'entrenament anomenat retropropagació de l'error, es va tornar a despertar l'interès envers la idea d'imitar el mecanisme de processament neuronal humà. Una forma particular del PDP, el perceptró multicapa que es mostra a continuació, és el que té més importància perquè té la capacitat d'aproximar qualsevol funció (d'entrada) a una precisió arbitrària, sempre que no es limités la complexitat de la configuració del processament. Els primers intents de fer servir xarxes neuronals per al reconeixement de veu es va centrar en tasques simples com el reconeixement de fonemes i de paraules, cosa que va funcionar. No obstant això, com el problema del reconeixement de la veu requereix inevitablement la gestió de la variació temporal, les xarxes neuronals no funcionaven.

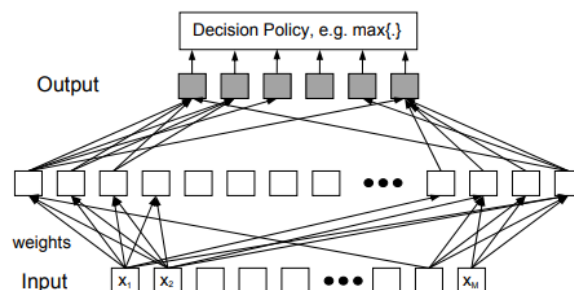


Figura 7. Perceptró multicapa

Als anys noranta, es van produir diverses innovacions en l'àmbit del reconeixement de patrons. El problema del reconeixement de patrons, que tradicionalment seguia el marc de Bayes i requeria l'estimació de les distribucions de les dades, es va transformar en un problema d'optimització que implicava la minimització de l'error de reconeixement empíric. Aquest canvi fonamental de paradigma es va deure al reconeixement del fet que les funcions de distribució per al senyal de veu no es podien elegir o definir amb precisió, i també que la teoria de decisió de Bayes es tornaria inaplicable en aquestes circumstàncies. Al cap i a la fi, l'objectiu d'un disseny de reconeixement hauria de ser aconseguir el menor error de reconeixement i no la millor adequació de la funció de distribució al conjunt de dades, com propugna el criteri de Bayes. El concepte de classificació mínima o error empíric va donar lloc a diverses tècniques com les màquines de vector de suport (SVM).

L'èxit dels mètodes estadístics va tornar a despertar l'interès de DARPA entre els 80 i els 90 i va donar lloc a nous sistemes de reconeixement de veu com el Sphinx de CMU, el sistema BYBLOS de BBN i el sistema DECIPHER de SRI. El sistema Sphinx de CMU va integrar amb èxit el model estadístic dels models ocults de Markov amb la força de cerca de la xarxa de l'anterior sistema HARPY. Així doncs, va ser capaç d'entrenar i integrar models de telèfons dependents del context en una xarxa de descodificació lèxica, i va aconseguir resultats notables en el reconeixement continu de veu de grans vocabularis.

A la dècada dels noranta es van fer grans progressos en el desenvolupament d'eines de programari que van permetre que es fessin grans investigacions a escala mundial. A mesura que els sistemes es tornaven més sofisticats, un sistema de programari de base ben estructurat era indispensable perquè la investigació i el desenvolupament continuessin incorporant nous conceptes i algoritmes. El sistema que va posar a disposició l'equip de la Universitat de Cambridge anomenat Hidden Markov Model Tool Kit es va convertir en una de les eines de programari més adoptades en la investigació del reconeixement automàtic de la parla.

La majoria d'investigacions dels sistemes de reconeixement de veu a partir dels anys vuitanta consideraven que el principal problema era transformar les ones de veu en paraules. Alguns investigadors també creien que el procés de reconeixement de veu

(*speech to text*) era un primer pas necessari en el procés que permetia a una màquina ser capaç d'entendre i respondre adequadament a la veu humana. En les avaluacions de la tecnologia de reconeixement de veu i la comprensió de tasques, es van treure dues conclusions sobre el procés de comunicació entre humans i màquines. La primera és que els possibles usuaris d'un sistema de reconeixement de veu haurien de pronunciar frases naturals que sovint no satisfien plenament les limitacions gramaticals del sistema de reconeixement (construccions no gramaticals, frases mal formulades, etc.) i que la llengua oral es pot veure afectada per l'anomenat soroll (interferències, sons externs, etc.). En segon lloc, com en la comunicació entre humans, les aplicacions de veu normalment requerien un diàleg entre l'usuari i la màquina per aconseguir la comprensió que es desitjava. El mètode de localització de paraules clau (i la seva aplicació en el sistema de processament de trucades per reconeixement de veu) es va introduir com a resposta al primer factor, mentre que el segon factor va despertar l'atenció de la comunitat d'investigadors de l'àrea de gestió del diàleg. A principis dels 1990 es van introduir moltes aplicacions i demostracions de sistemes que reconeixien la importància de la gestió dels diàlegs sobre l'exactitud del reconeixement de paraules d'un sistema amb l'objectiu de crear, amb el temps, una màquina que realment imités les capacitats de comunicació dels humans. Entre aquests sistemes es troben el Pegasus i Jupiter, que es van desenvolupar a l'institut de Tecnologia de Massachussets i el How May I Help You (HMIHY) que va desenvolupar Gorin a AT&T. El Pegasus era un sistema de conversa de veu que proporcionava informació sobre l'estat dels vols de les aerolínies a través d'una línia telefònica ordinària. El Jupiter era un sistema semblant que es centrava en l'accés a la informació meteorològica. Aquests sistemes van personificar l'eficàcia de la gestió del diàleg. Amb una gestió del diàleg gestionada correctament, aquests sistemes podien guiar l'usuari perquè proporcionés la informació necessària per processar una sol·licitud sense demanar explícitament detalls de la consulta. La gestió del diàleg, normalment també incorpora la confirmació incrustada de les frases reconegudes i la gestió d'errors perquè l'usuari reaccioni com si hi hagués un agent humà en lloc d'una màquina a l'altra banda de la línia telefònica. L'objectiu era dissenyar una màquina que comunicés en comptes de merament reconèixer enunciats orals (Rabiner, 2004).

El final dels anys noranta es caracteritza pel desplegament de les aplicacions reals habilitades per veu com la gestió automatitzada de trucades assistides per operador o les línies d'atenció al client. Tot i que els sistemes de reconeixement i comprensió de la veu automàtics encara no eren perfectes quant a l'exactitud de les paraules o les tasques, les aplicacions degudament desenvolupades van continuar aprofitant la tecnologia fins avui dia, i s'han transformat, com veurem en els apartats següents, en màquines que es poden confondre amb persones reals.

2.1.3. L'intel·ligència artificial (IA) i el processament de veu (speech to speech)

Si ara com ara parlem de sistemes de síntesi i reconeixement de veu, hem de parlar d'intel·ligència artificial (IA). Encara que sembli un sistema nou i modern, és una disciplina que té força anys d'història i, com sempre, és important saber d'on venim per saber on som. A continuació, farem una breu introducció a l'intel·ligència artificial i a tots els subelements que se'n deriven i que cal tenir presents a l'hora de parlar del processament de veu.

Les definicions següents no són d'ús universal, però sí importants per parlar de l'IA (Winston, 1984). La intel·ligència és la capacitat de fer allò que és correcte en el moment correcte, en un context en què no ser res (no canviar de comportament) seria pitjor. Per tant, la intel·ligència requereix capacitat de percebre contextos d'acció, capacitat d'actuar i capacitat d'associar determinats contextos a determinades accions.

Segons aquesta definició, les plantes són intel·ligents (Trewavas, 2005), i també un termòstat (McCarthy, 1983; Touretzky, 1988). Tots aquests poden percebre el context i respondre, per exemple, les plantes reaccionen a la direcció de la llum i els termòstats a la temperatura. A continuació, classifiquem un sistema com a cognitiu si és capaç de modificar la seva intel·ligència, cosa que les plantes i els termòstats (almenys, els mecànics) no poden fer. La intel·ligència, és un subconjunt estricte de la computació, és a dir, de transformació de la informació. Cal tenir en compte que la computació és un procés físic, no matemàtic: requereix temps, espai i energia. La intel·ligència és el subconjunt de la computació que transforma un context en acció.

Intel·ligència artificial (IA), per convenció, és un terme que es fa servir per descriure artefactes (normalment digitals) que amplien alguna de les capacitats relacionades amb la intel·ligència natural. Així, els sistemes automàtics de visió, el reconeixement de veu, el reconeixement de patrons i els sistemes de producció fixa (és a dir, que no aprenen) es consideren exemples d'IA (Russell i Norvig, 2009). De totes maneres, també es poden considerar formes de computació, tot i que allò que produeixen constitueix una acció en un sentit convencional. Amb tot, si acceptem el concepte de robòtica aplicada al cos, podríem ampliar la definició fins a considerar IA qualsevol artefacte que amplii les nostres capacitats de percepció i acció. Seria una definició inusual, però ens facilitaria entendre millor els canvis que l'IA proporciona a la nostra societat. El terme aprenentatge automàtic (AA) designa qualsevol mitjà de programació d'IA que requereixi no només una codificació manual, sinó també un component de generalització automatitzada de les dades que es presenten per mitjà de l'acumulació d'estadístiques sobre aquests (Murphy, 2012; Erickson et al., 2017). Sovint, tot i que no sempre, l'AA es limita a cercar regularitats en les dades associades a categories d'interès, cosa que inclou oportunitats apropiades per a determinades accions. Així mateix, l'AA es fa servir amb freqüència per deduir associacions i també per adquirir noves competències d'acció (Huang et al., 2016).

Hem de tenir en compte que l'AA implica un component programat manualment. La conceptualització o descobriment d'un algoritme mai no porta a l'aparició espontània d'una màquina capaç de sentir o actuar. Per definició, tota IA és un artefacte, resultat d'accions humanes deliberades. Perquè hi hagi aprenentatge s'ha de construir, primer, alguna cosa dissenyada per connectar una font de dades a una representació. Tots els sistemes intel·ligents tenen una arquitectura, una configuració del flux d'energia i informació, i gairebé tots inclouen emplaçaments on es reté la informació que s'anomena memòria. El disseny d'aquesta arquitectura es denomina enginyeria de sistemes. Aquesta és la fase en què s'ha de comprovar la seguretat i la validesa d'un sistema. En contra d'algunes afirmacions estrofolàries, però freqüents, la seguretat de l'IA no és un camp d'estudi nou. De fet, l'enginyeria de sistemes és anterior als ordinadors (Schlager, 1956) i sempre ha estat un component principal de l'aprenentatge de les ciències informàtiques. La integració de l'IA en el *software* es remunta a fa molts

anys, de manera que té con una llarga trajectòria de seguretat en els dissenys (Bryson, 2003; Chessell i Smith, 2013).

En termes tècnics, l'autonomia és la capacitat d'actuar com un individu (Armstrong i Read, 1995; Cooke, 1999). Per exemple, un país perd la seva autonomia si les seves institucions s'enfonsen, de manera que només les accions individuals dels seus ciutadans són eficaces, tant si les seves institucions tenen influències d'altres actors o govern com si no poden determinar el rumb del país per si mateixes. Cal dir que tots dos extrems són força inusuals. De fet, entre animals socials com els humans, l'autonomia mai no és absoluta (Gilbert et al., 2012). La nostra intel·ligència individual determina moltes de les nostres accions, però algunes cèl·lules poden tornar-se cancerígenes i atendre els seus propis objectius en detriment del nostre benestar general (Hanahan i Weinberg, 2011). D'una manera semblant, entenem que una família, un lloc de treball o un govern influeixen en les nostres accions. També estem subjectes a una influència social molt superior a la que en som conscients (Devos i Banaji, 2003). No obstant això, se'ns considera autònoms perquè, en certa manera, la nostra intel·ligència individual també influeix en el nostre comportament. Conseqüentment, un sistema tècnic capaç de percebre el món i seleccionar una acció específica per al context en què es trobi s'anomena 'autònom' encara que les seves accions estiguin determinades pels dissenyadors que van desenvolupar la seva intel·ligència i els qui l'operen. Els operaris poden influir en l'IA en configurar els paràmetres del seu funcionament que inclou quan i on funciona, si ho fa. Així doncs, els dissenyadors creen el sistema i determinen les seves capacitats, en especial, a quina informació pot accedir i quines accions pot efectuar. Fins i tot, si un dissenyador decideix introduir un component d'atzar en un sistema d'IA, com la dependència de l'entorn en què es troba o un generador de nombres aleatoris, aquesta inclusió és també una elecció deliberada.

2.1.3.1. Les bases de la intel·ligència artificial

Els inicis de l'IA i el concepte que coneixem com a màquines intel·ligents es remunta a l'antiga Grècia. Els artefactes 'intel·ligents' fa milers d'anys que apareixen en la literatura i des de llavors, els dispositius han demostrat que poden actuar amb un cert grau d'intel·ligència. Quan els ordinadors moderns es fan fer accessibles per a

tothom, després de la Segona Guerra Mundial, es van començar a desenvolupar programes que permetien fer tasques intel·lectualment difícils.

La mitologia grega, com hem mencionat anteriorment, va ser la primera que va incorporar el pensament de desenvolupar una màquina intel·ligent. Hefest, déu del foc i de la teckné (intel·ligència manual) i que era ferrer i Talos, gegant de bronze (i de fet, forjat per Hefest), van ser els primers a incorporar la idea dels robots intel·ligents. A més hi va haver d'altres intents en l'edat antiga que van provar de crear artefactes amb un comportament semblant al dels humans.

Ja a principis del segle IV aC, Aristòtil va inventar el sil·logisme, el primer sistema formal de raonament deductiu. Aquesta modalitat d'argumentació deductiva consisteix en l'afirmació de dues proposicions, relacionades entre elles i anomenades premisses, de les quals en resulta una altra, dita conclusió. Més tard, al segle XIII, es van crear els caps parlants, éssers considerats de naturalesa entre la mecànica i la màgia, que parlaven, aconsellaven els seus amos o predeien el futur. Un dels exemples més famosos és el cap amb forma d'home de Roger Bacon (1214-1294), fet de llautó i que podia respondre a preguntes sobre el futur o el d'Albert el Gran, amb forma de dona, entre d'altres que hi va haver. En aquest mateix segle Ramon Llull va inventar les màquines per descobrir veritats no matemàtiques a partir de combinacions. L'any 1206, l'inventor àrab Al-Jazari, va dissenyar el que es considera el primer robot hominoide. La màquina tenia l'aspecte de quatre músics a bord d'un petit vaixell que funcionava amb el flux de l'aigua i es feia servir per entretenir els invitats a les festes reials, com podem veure en la Figura 8. El seu mecanisme tenia un tambor programable amb clavilles que xocaven amb petites palanques que accionaven instruments de percussió. Es podien canviar els ritmes i els patrons que tocaven si es movien les clavilles.



Figura 8. Vaixell automàtic d'Al-Jazari

Al segle XV Gutenberg va inventar la impremta fent servir el tipus mòbil i va imprimir la coneguda com a 'Bíblia de Gutenberg', l'obra més important del món de la impremta. Entre els segles XV i XVI ja es van inventar els rellotges, les primeres màquines modernes de medicació, inicialment fent servir torns. Al segle XVI els fabricants de rellotges van estendre la seva producció i van crear fins i tot animals mecànics. N'és un exemple el lleó autòmat de Da Vinci, un regal que li va encomanar el papa Lleó X a l'artista per regalar-li al rei Francesc I de França.

D'altra banda, es diu que Rabbi Loew de Praga, l'any 1580, va inventar el Golem, un home fet de matèria inanimada que tenia un aspecte i comportament humans. A principis del segle XVII, Descartes va suggerir que els cossos d'animals eren màquines complexes, pensament que també van recolzar altres pensadors de l'època tot i que cadascun va aportar les seves variacions i interpretacions del mètode cartesià. L'any 1642 Pascal va crear la primera calculadora mecànica (Figura 9), una petita caixa de fusta que tenia a la tapa una filera de discos numerats, amb els forats per introduir els dits i fer-los girar. Cada disc tenia una finestreta, i hi havia tota una filera de finestretes sota la filera de discos: de dreta a esquerra s'alineaven les unitats, desenes, centenes, milers, etc. Quan una roda feia una volta completa, avançava l'altra roda situada a la seva esquerra.



Figura 9. Calculadora de Pascal

El 1651 Thomas Hobbes va publicar el Leviatan, un llibre sobre política en què desenvolupa una teoria del pensament mecanicista. Entre el 1662 i el 1666, Samuel Morland va idear altres calculadores aritmètiques i posteriorment, el 1673, Gottfried Leibniz va fer un pas endavant respecte als seus precursors ja que, en crear l'Step Reckoner (Figura 10) va millorar la calculadora pascalina. Mentre la calculadora de Pascal només sumava i restava la de Leibniz també multiplicava i dividia. A més, el matemàtic també va preveure el càlcul universal de raonament pel qual els arguments es podien decidir mecànicament.

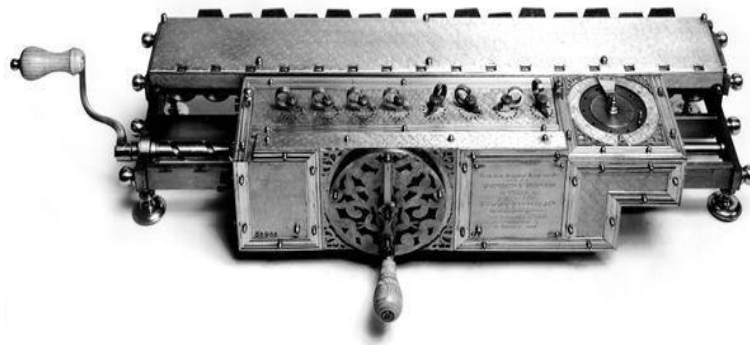


Figura 10. Màquina de Leibniz

Durant el segle XVIII es van inventar diversos artefactes mecànics com el cèlebre ànec amb aparell digestiu de Jacques de Vaucanson o la famosa farsa de El Turc (1769) de Von Kempelen, que ja hem mencionat anteriorment i que simulava ser un autòmat que jugava als escacs. Edgar Allen Poe va escriure a l'abril del 1836 que El Turc no podia ser una màquina perquè, si ho fos, no perdria.

Al segle XIX Joseph-Marie Jacquard va inventar el teler de Jacquard, la primera màquina programable. L'artefacte utilitzava targetes perforades per aconseguir teixir la roba seguint uns patrons, permetent que fins i tot els usuaris més inexperts poguessin elaborar complexos dissenys. No obstant això, hi va haver un intent de destrossar aquests tipus de màquines. Segons Bloy (2005), després del final de les guerres franceses es va fer cada cop més evident que Anglaterra patia greus problemes socials, econòmics i polítics. Hi va haver grans canvis com a resultat directe de les guerres franceses, però d'altres, es van deure al creixement i a canvis naturals. L'angoixa i el

descontentament que causaven aquests canvis es van manifestar en un seguit d'esdeveniments en el període 1811-1819. N'és un exemple l'apogeu del ludisme. Els luddites eren un col·lectiu que mirava de salvar els seus mitjans de vida i, per fer-ho, destrossaven les màquines industrials que es desenvolupaven per a les indústries tèxtils que 'els hi feien la competència' com per exemple, les màquines de cosir. No sembla que hi hagués cap mena de motivació política al darrere ni cap organització nacional: sembla que simplement atacaven perquè veien les màquines com la raó del declivi de la seva manera de viure.

Més endavant, l'any 1818 Mary Shelley publica Frankenstein. D'altra banda, Charles Babbage i Ada Lovelace dissenyen la Màquina Analítica (Figura 11), capaç de resoldre equacions diferencials.

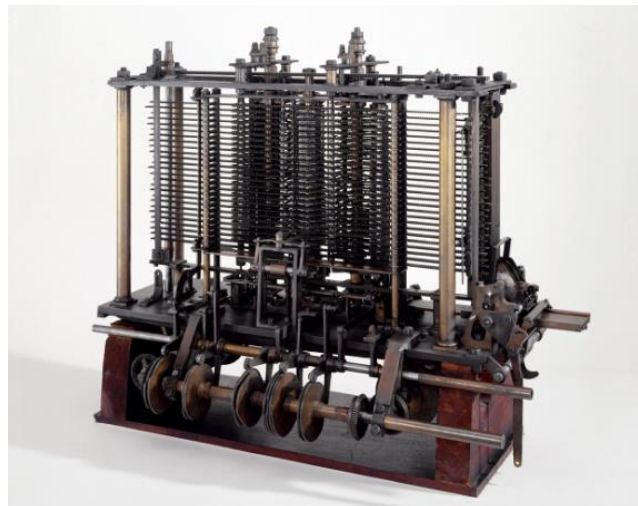


Figura 11. Màquina Analítica de Lovelace i Babbage.

Lovelace es va adonar que la màquina tenia més aplicacions a part de la pura calculació i va reconèixer tot el potencial de la màquina. A les notes, va incloure el que ara es considera el primer algorisme que es va intentar dur a terme en una màquina. És per això que se la considera la primera programadora.

George Boole, durant el primer quart del segle XIX, crea l'àlgebra de Boole, una branca de les matemàtiques amb propietats i regles similars, tot i que diferents, a les de l'àlgebra ordinària. Amb aquest invent, pretenia explicar les lleis fonamentals d'aquelles operacions de la ment humana per les quals es regeixen els raonaments. Posteriorment, aquesta àlgebra fou utilitzada per al disseny de circuits digitals. L'àlgebra de Boole té

una característica especial: les seves variables només poden adoptar dos valors, tradicionalment denominats cert i fals (normalment representats com a 1 i 0, respectivament). Així doncs, l'àlgebra de Boole té valors lògics binaris.

Durant la primera meitat del segle XX, Bertrand Russell i Alfred North Whitehead van publicar *Principia Mathematica*, un conjunt de tres llibres que va revolucionar la lògica formal. Russell, Ludwig Wittgenstein, i Rudolf Carnap van portar la filosofia a l'anàlisi lògica del coneixement. D'altra banda, l'any 1912, Torres i Quevedo van crear la màquina d'escacs 'Ajedrecista'. Fent servir electroimants sota el tauler d'escacs, la màquina jugava automàticament un final de rei i torre contra el rei d'un oponent humà. No jugava de manera massa precisa i no sempre arribava al mat en el nombre mínim de moviments, a causa de l'algorisme simple que avaluava les posicions. No obstant això, aconseguia la victòria en totes les ocasions. Al contrari que els falsos autòmats com per exemple El Turc que eren dispositius operats per humans ocults, per fer la falsa impressió de ser autòmats, el 'Ajedrecista' era capaç de jugar als escacs sense intervenció humana. Es podria considerar, així doncs, el primer joc d'ordinador de la història. L'any 1921 Karel Capek escriu l'obra teatral "R.U.R" en què apareix, per primera vegada en la història, el terme 'robot'. Aquesta paraula, l'havia ideada el seu germà a partir de la paraula txeca *robota* que significa 'esclau'. Al 1936, Alan Turing va proporcionar una formalització força influent dels conceptes d'algorisme i computació: la màquina de Turing. En el camp de la intel·ligència artificial, és conegut, sobretot, pel test de Turing (1950), un criteri segons el qual es pot distingir la intel·ligència d'una màquina si les seves respostes en la prova són indistingibles de les d'un ésser humà. A la Fira Mundial de Nova York del 1939, nord-americana Westinghouse va presentar Elektro, un artil·lugi creat per Joseph Barnett. Elektro era un robot de dos metres d'alçada, de 120 quilos i capaç de fer 26 moviments diferents. Barnett, enginyer de Westinghouse Electric Corporation, va fer servir la tecnologia punta per crear el primer hominoide que, a més, sabia parlar. Concretament, tenia gravades aproximadament 700 paraules i feia servir un reproductor de discos de 78 rpm. Entre altres capacitats, podia fumar, inflar globus, caminar o moure el cap i els braços. Els seus ulls fotoelèctrics podien distingir entre el llum vermell i el verd. Barnett va crear Elektro a la fàbrica de Westinghouse, a Mansfield, i es va basar en un seguit de reproductors de discos, cèl·lules fotovoltaïques, motors i relés telefònics. El robot va ser

un símbol de desenvolupament tecnològic i innovació després d'anys de crisi. Elektro, en aquesta ocasió, no va ser l'únic robot que hi va assistir. Van presentar Sparko, un gos-robot Terrier d'uns 30 quilos i era capaç de bordar, seure i fins i tot obeir a ordres simples.

L'any 1943, Warren McCulloch i Walter Pitts publiquen "Un càlcul lògic A Logical Calculus of the Ideas Immanent in Nervous Activity", on assenten els principis de les xarxes neuronals. Al 1943, en un article, Arturo Rosenblueth, Norbert Wiener i Julian Bigelow encunyen el terme "cibernètica" i el 1948 Weiner publica un llibre amb aquest mateix nom. Emil Post va demostrar que els sistemes de producció són un mecanisme computacional general i també va investigar sobre la integritat, la inconsistència i la teoria de la prova. George Polya va publicar el seu famós llibre sobre el pensament heurístic, "How to Solve It", al 1945 en què introdueix el terme 'heurística' en el pensament modern, cosa que va influenciar diversos experts en IA.

Vannevar Bush va publicar al juliol del 1945 "As We May Think" una visió clarivident del futur en què els ordinadors ajuden les persones en diverses tasques. Grey Walter va experimentar amb robots autònoms, unes tortugues anomenades Elsie i Elmer, a Bristol el 1948, i es va basar en la premissa que un número petit de cèl·lules podria donar lloc a comportaments complexos.

Turing va publicar al 1950 "Computing Machinery and Intelligence" (maquinària computacional i intel·ligència, en català) en què, com hem esmentat anteriorment, introdueix el seu test de Turing com una manera de fer operatiu un test de comportament intel·ligent. El mateix any, Claude Shannon va publicar una anàlisi detallada del joc d'escacs, un treball que descrivia la programació d'un ordinador per jugar a escacs, que es va convertir en la base de posteriors desenvolupaments. Per últim, el mateix any, l'escriptor Isaac Asimov va publicar les seves tres lleis de robòtica. Tots els robots dels seus relats i novel·les estan dissenyats per a complir-les. En el seu univers, les lleis són «formulacions matemàtiques impreses en els senders positrònics del cervell» dels robots (línies de codi del programa que regula el compliment de les lleis que es guarda en la seva memòria principal). Van aparèixer per primer cop en el relat "Runaround", el 1942 i estableixen tres normes. En primer lloc, un robot no farà mal a cap ésser humà o, per inacció, permetrà que un ésser humà pateixi cap mal. A

més, un robot ha de complir les ordres que li donin els éssers humans, a excepció de les que entressin en conflicte amb la primera llei. Per últim, un robot ha de protegir la seva pròpia existència sempre que aquesta protecció no entri en conflicte amb la primera o la segona llei. Els robots sempre han d'actuar respectant les lleis de la robòtica, no com una regla moral sinó perquè l'estructura del seu "cervell positrònic" està construïda al voltant d'elles. Aquestes lleis sorgeixen com a mesura de protecció dels humans. Segons Isaac Asimov, la concepció d'aquestes lleis volia contrarestar el temor que els robots es poguessin rebel·lar i atacar els seus creadors, i no són més que una aplicació de les lleis de les eines, no formulades per ser massa evidents: una eina ha de ser segura, ha de fer la tasca per la qual està fabricada i no s'ha de fer malbé (amb aquest ordre de prioritats).

2.1.3.2. La intel·ligència artificial a partir dels anys 50

Podem començar a parlar d'intel·ligència artificial quan es crea el terme, l'any 1956, quan John McCarthy va fer la primera conferència sobre l'assumpte i, com a conseqüència, va encunyar el terme. També es fa, en l'època, la primera demostració d'un programa que tenia comportaments intel·ligents: el Logic Theorist (LT), un sistema que van crear Allen Newell, Herbert A. Simon i Cliff Shaw que imitava el comportament humà per solucionar problemes matemàtics. L'any següent, ells mateixos, van crear el General Problem Solver (GPS) amb l'objectiu de construir una màquina capaç de resoldre problemes generals. Podia resoldre, en principi, qualsevol mena de problema simbòlic formal com per exemple, provar teoremes, resoldre problemes geomètrics, treballar amb la lògica proposicional i jugar a escacs, i es basava en l'anterior estudi de l'LT. El GPS va ser el primer programa en què es va separar el coneixement dels problemes de l'estratègia per resoldre'ls, i es va implementar en el llenguatge de programació IP (Information Processing Language). Tot i que el GPS va aconseguir resoldre problemes simples (com el de les Torres de Hanoi), no podia resoldre problemes del món real.

Entre el 1952 i el 1962, Arthur Samuel (IBM), escriu el primer programa de joc, concretament de dames, per aconseguir l'habilitat suficient per desafiar un campió mundial. Se'l considera, així doncs, el pioner en el terreny dels jocs informàtics. El programa de Samuel va ser un dels primers d'autoaprenentatge del món i, com a

conseqüència, una demostració força primerenca del concepte fonamental d'intel·ligència artificial (IA).

El 1958, McCarthy (MIT, Institut de Tecnologies de Massachusetts) es va inventar el llenguatge LISP, una família de llenguatges de programació d'ordinador de tipus multiparadigma. El mateix any, Herb Gelernter i Nathan Rochester (IBM) van descriure el teorema de geometria que aprofita un model semàntic del domini en forma de diagrama de casos 'típics'. També es va celebrar, al Regne Unit, la conferència de Teddington sobre la Mecanització dels processos de pensament, on es van presentar programes amb sentit comú de John McCarthy, el Pandemonium d'Oliver Selfridge i alguns models de programació heurística i intel·ligència artificial de Marvin Minsky. El 1959, John McCarthy i Marvin Minky funden el MIT AI Lab. A finals dels anys cinquanta, Margaret Masterman i companys seus de la Universitat de Cambridge dissenyen xarxes semàntiques per a la traducció automàtica.

Als principis dels seixanta, Ray Solomonoff assenta les bases d'una teoria matemàtica de l'IA i introdueix mètodes universals bayesians per a la inferència inductiva i la predicció. Licklider, d'altra banda, va escriure l'article Simbiosi home-ordinador, cosa que indica la necessitat de simplificar la interacció entre els ordinadors i els seus usuaris. James Slagle (MIT) va escriure en llenguatge LISP el primer programa d'integració simbòlica, el SAINT (per les seves sigles en anglès), que podia resoldre problemes de càlcul amb un nivell com el d'un universitari de primer any. El 1962 es funda la primera empresa de robots industrials: Unimation. El robot industrial de l'empresa, Unimate, va treballar en una línia d'assemblatge d'automòbils a General Motors. El 1963 Thomas Evans fa un programa al qual anomenat ANALOGY, en què demostra que els ordinadors poden resoldre els mateixos problemes d'analogia que hi ha a les proves de coeficient intel·lectual. La tesi d'Ivan Sutherland al MIT en Sketchpad presenta la idea dels gràfics interactius en la informàtica. A més, Edward A. Feigenbaum i Julian Feldman publiquen *Computers and Thought*, la primera col·lecció d'articles sobre intel·ligència artificial.

El 1964, la tesi de Danny Bobrow al MIT va mostrar que els ordinadors poden entendre el llenguatge natural suficientment bé per resoldre problemes de paraules d'àlgebra. La tesi de Bert Raphael, també al MIT, sobre el programa SIR va demostrar

el poder de la representació lògica del coneixement per als sistemes de preguntes i respostes. L'any següent, J. Alan Robinson va inventar un procediment de prova mecànica, el mètode de resolució, que permetia que els programes poguessin treballar eficientment amb la lògica formal com a llenguatge de representació. A més, com ja hem esmentat anteriorment, Joseph Weizenbaum (MIT) construeix ELIZA, un programa interactiu capaç de tenir un diàleg sobre qualsevol tema en anglès.

El 1966, Ross Quillian (Carnegie Institute of Technology) desenvolupa les xarxes semàntiques, grafs en què els nodes o vèrtexs etiquetats representen conceptes o característiques específiques, mentre que els arcs, també representats, representen vincles de diverses classes entre conceptes. També hi ha el primer taller sobre intel·ligència artificial a Edimburg. L'informe negatiu de la traducció automàtica tira per terra molts anys de feina en el processament del llenguatge natural (NLP). L'any 1967 el programa *dendral* (Edward Feigenbaum, Joshua Lederberg, Bruce Buchanan, Georgia Sutherland a Stanford) va fer una demostració de com interpretar espectres de massa en compostos químics, cosa que va suposar el primer programa basat en coneixement per al raonament científic. D'altra banda, Joel Moses (MIT) va demostrar el model del raonament simbòlic per integrar problemes en el programa Macsyma (PDF file), cosa que va suposar el primer programa basat en el coneixement per a les matemàtiques. Richard Greenblatt va crear al MIT un programa per jugar a escacs basat en el coneixement, MacHack.

A finals dels anys seixanta, Doug Engelbart va inventar el ratolí a SRI. L'Institut d'Investigació d'Stanford desenvolupa el robot Shakey, que combinava locomoció, percepció i resolució de problemes. Roger Schank (Stanford) va definir el model de dependència conceptual per entendre el llenguatge natural. Més tard les van fer servir Robert Wilensky i Wendy Lehnert per a les seves tesis doctorals sobre la comprensió de la història i després, Janet Kolodner per a la comprensió de la memòria. Yorick Wilks (Stanford) va desenvolupar la visió de coherència semàntica del llenguatge anomenada Semàntica de preferències, que es va incorporar al primer programa de traducció automàtica impulsat per la semàntica, i va ser la base per a les tesis de Boguraev i David Carter a Cambridge. L'any 1969 es va celebrar la primera Conferència Internacional Conjunta sobre Intel·ligència Artificial a Washington D.C. Per últim,

Marvin Minsky i Seymour Papert van publicar *Perceptrons*, on van demostrar els límits de les xarxes neuronals simples (de dues capes). Amb això, comença el que s'anomena l'hivern de l'IA dels anys setanta, un fracàs de confiança i finançament. No obstant això, el progrés va seguir. Els primers a provar-ho van ser McCarthy i Hayes, que van començar a fer l'assaig 'Alguns problemes filosòfics des del punt de vista de l'IA'.

A principis dels anys setanta, Seppo Linnainma publica el model invers de diferenciació automàtica, mètode que es va anomenar més endavant com retropropagació i que es fa servir força per entrenar xarxes neuronals artificials. D'altra banda, Jane Robinson i Don Walker van establir un grup de processament del llenguatge natural en el SRI. Jaime Carbonell desenvolupa SCHOLAR, un programa interactiu per a la instrucció assistida per ordinador basada en xarxes semàntiques com a representació del coneixement. Bill Woods, d'altra banda, va desenvolupar les xarxes de transició augmentada (ATNs) per representar la comprensió del llenguatge natural. El programa de doctorat de Patrick Winston, ARCH, al MIT, va aprendre conceptes a partir d'exemples del món dels blocs per a nens. Amb la seva tesi de doctorat, Terry Winograd (MIT), el 1971 va demostrar l'habilitat dels ordinadors d'entendre oracions en anglès gràcies al seu programa de comprensió de la llengua, SHRDLU, un braç-robot que feia instruccions escrites en anglès. El 1972, Alain Colmerauer crea el llenguatge de programació Prolog, i el 1973 el grup de Assembly Robots de la Universitat d'Edimburg construeix Freddy, el famós robot escocès, capaç d'utilitzar la visió per localitzar models. Un any més tard, Ted Shortliffe presenta MYCIN (Stanford) en què demostra el poder dels sistemes basats en regles per a la representació i interferència del coneixement en el domini de la diagnosi i la teràpia mèdica. De vegades, se l'anomena el primer sistema expert. Earl Sacerdoti va desenvolupar un dels primers programes de planificació, ABSTRIPS, i també fa tècniques de planificació jeràrquica. El 1975 Marvin Minsky publica un article a *Frames* com a representació del coneixement, en què reuneix diverses idees sobre esquemes i vincles semàntics. El programa d'aprenentatge Meta-Dendral va produir diversos resultats en química, els primers descobriments científics que es van fer per un ordinador en una revista arbitrada. Barbara Grosz (SRI) va establir límits en els enfocaments tradicionals de l'IA per a modelatge del discurs.

A mitjans dels anys setanta, Alan Kay i Adele Goldberg (Xerox PARC) van desenvolupar el llenguatge Smalltalk, amb què van establir el poder de la programació orientada als objectes i les interfícies orientades a les icones. David Marr i els seus companys del MIT descriuen 'l'esbós primari' i la seva funció en la percepció visual. El 1976, amb el programa AM, Doug Lenat (Stanford) mostra el mode de descobriment (cerca guiada de manera flexible per a conjectures interessants). Herbert A. Simon guanya el Premi Nobel d'Economia per la seva teoria de la racionalitat acotada, una de les pedres angulars de l'IA. El 1978, el SUMEX-AIM de Stanford, encapçalat per Ed Feigenbaum i Joshua Lederberg demostra el poder d'ARPAnet en la col·laboració científica. El programa MOLGEN que van escriure a Stanford Mark Stefik i Peter Friedland va desvelar que una representació del coneixement orientada als objectes es podia fer servir per planificar experiments de clonació genètica. El programa MYCIN, inicialment, la tesi doctoral de Ted Shortliffe a Stanford, va demostrar que tenia un rendiment de nivell elevat. La tesi doctoral de Bill VanMelle a Stanford mostra la generalitat de la representació del coneixement i l'estil de raonament de MYCIN en el programa EMYCIN. Jack Myers i Harry Pople, desenvolupen a la Universitat de Pittsburgh INTERNIST, un programa de diagnòstic mèdic basat en el coneixement clínic del Dr. Myers. Cordell Green, David Barstow i Elaine Kant, a Stanford desenvolupen el sistema CHI per a la programació automàtica. L'Stanford Cart, que va construir Hans Moravec, es converteix en el primer vehicle autònom controlat per ordinador en el moment en què atravesava, amb èxit, una sala plena de cadires i circumval·la el laboratori d'IA de Stanford. Drew McDermott i Jon Doyle, al MIT, i John McCarthy, a Stanford, comencen a publicar treballs sobre lògica no monòtona i aspectes formals d'aspectes formals del manteniment de la veritat.

El 1980, es desenvolupen i es comercialitzen les màquines LISP. Es fa també la primera Conferència Nacional de l'Associació Americana d'Intel·ligència Artificial (AAAI) a Stanford. Danny Hillis dissenya la màquina de connexió, que fa servir la computació paral·lela per donar un nou poder a l'IA i la computació en general, i més tard funda Thinking Machines Corporation. L'any 1982 es comença el projecte de sistemes informàtics de cinquena generació, una iniciativa del Ministeri de Comerç Internacional i Indústria del Japó, per crear un "ordinador de cinquena generació" que se suposava que havia de fer molts càlculs fent servir el paral·lelisme massiu. James F.

Allen inventa el càlcul de l'interval, la primera formalització àmpliament utilitzada en esdeveniments temporals. A mitjans dels vuitanta, les xarxes neuronals es comencen a fer servir àmpliament amb l'algoritme de retropropagació, que havia publicat Linnainmaa el 1970 i que va aplicar a les xarxes neuronals Paul Werbos. També es mostra el programa de disseny AARON, que va crear Harold Cohen, a la Conferència Nacional AAAI de 1985, basat en més d'una dècada de desenvolupament i investigació. El 1986, l'equip d'Ernst Dickmanns de la Universitat de Bundeswehr de Munic construeix els primers automòbils-robot, que podien conduir fins a 90 km per hora per carreteres buides. Marvin Minsky publica *The Society of Mind*, una descripció teòrica de la ment com a una col·lecció d'agents cooperatius. Gairebé al mateix temps, Rodney Brooks introdueix l'arquitectura de subsunció i la robòtica basada en el comportament com un model modular més minimalista d'intel·ligència natural. El 1987, Alacritous Inc. llança la generació 2.0 de Alarcity, el primer sistema d'assessorament comercial i estratègic. El sistema es va basar en un sistema expert autodesenvolupat amb 3.000 regles sobre l'evolució dels mercats i les estratègies competitives, i en van ser coautors Alistair Davidson i Mary Chung. El 1989, Dean Pomerleau, a CMU, crea ALVINNN, un vehicle terrestre autònom creat amb xarxes neuronals.

Els anys noranta es caracteritzen perquè es van produir grans avenços en totes les àrees de l'IA, i hi va haver importants demostracions d'aprenentatge automàtic, raonament basat en casos, planificació, programació, raonament, comprensió i traducció del llenguatge natural, realitat virtual, jocs, etc. El projecte COG de Rod Broks (MIT, Figura 12) va significar un pas important en el desenvolupament de robots hominoides. Es va basar en la hipòtesi que per aconseguir una intel·ligència humana, l'artefacte havia de guanyar experiència interactuant amb humans, tal com fan els nadons. Això va fer que calguessin moltes interaccions amb el robot durant molt de temps.

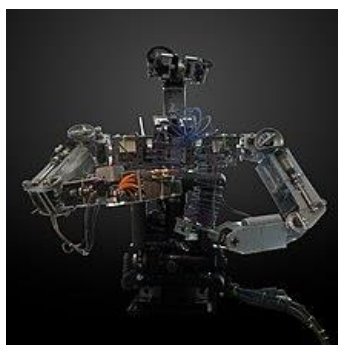


Figura 12. El robot COG

El TD-Gammon, un programa de backgammon que va escriure Gerry Tesauro, va demostrar que l'aprenentatge de reforç és suficientment poderós per poder crear un programa de joc d'alt nivell que podria competir i sortir victoriós en partides contra jugadors professionals. Carol Stoker i l'equip de robòtica de la NASA exploren la vida marina a l'Antàrtica amb un robot submarí operat de manera remota a través d'un enllaç satèl·lit des de Moffett Field (Califòrnia). L'any 1996, Deep Blue (Figura 13) una supercomputadora que havia desenvolupat IBM per jugar a escacs guanya el campió del món vigent, Gary Kaspárov, el 10 de febrer de 1996. No obstant això, Kaspárov va guanyar 3 i en va empatar 2 de les següents partides, i va derrotar Deep Blue per 4-2. La trobada va concloure el 17 de febrer de 1996. Una nova versió, anomenada Deeper Blue va jugar de nou contra Kaspárov el maig de 1997, i va guanyar la trobada a 6 partides per 3½-2½, cosa que el va convertir en la primera computadora que va derrotar un campió del món vigent, amb un ritme de joc de torneig estàndard.



Figura 13. El Deep Blue, primera màquina que va guanyar un humà

També el 1997, es fa el primer partit del projecte anomenat Robocup, que ofereix partits de futbol de 40 equips de robots i uns 5.000 espectadors. Aquest projecte pretenia promoure, a través de competències integrades en robots autònoms, la investigació i l'educació sobre l'IA. El 1998, Tiger Electronics presenta Furby, i es converteix en el primer intent de situar l'IA en un entorn domèstic. El 1999 Sony presenta un robot semblant millorat, AIBO, que es converteix en la primera 'mascota' autònoma fruit de la intel·ligència artificial. A finals dels anys noranta, els rastrejadors web i altres programes d'extracció d'informació basats en l'IA es tornen essencials per a l'ús generalitzat de la World Wide Web. També es fa una demostració d'una sala

intel·ligent al laboratori d'IA del MIT. Per últim, s'inicia un treball a Oxygen Architecture, que connecta ordinadors mòbils i estacionaris amb una xarxa adaptativa.

Ja al segle XXI, arribem a l'època actual de l'IA, en què la tecnologia evoluciona gairebé cada dia. No obstant això, només anomenarem alguns dels grans invents que s'ha produït gràcies als avenços que hi ha hagut. L'any 2000, es comercialitzen els *robotpets*, les joguines intel·ligents. A més, Cynthia Breazeal (MIT) publica la seva tesi sobre les màquines sociables, i crea KISMET, un robot que expressa emocions amb la seva cara. També, el 2002 s'inventa el Roomba d'iRobot, que aspira el terra i evita obstacles. El vehicle autònom de Stanford, Stanley, guanya la cursa de DARPA Grand Challenge, i els robots de la NASA Spirit i Opportunity naveguen de manera autònoma per la superfície de Mart. El 2005, Honda crea ASIMO, un robot hominoide artificialment intel·ligent que és capaç de caminar tan ràpidament com un ésser humà. El mateix any, la tecnologia de la recomanació basada en el seguiment de l'activitat web fa que la intel·ligència artificial també s'extrapoli al màrqueting, i a més, neix Blue Brain, un projecte que té com a objectiu estudiar l'estructura del cervell dels mamífers a partir d'una simulació del cervell molecularMENT. Aquest tipus d'enginyeria permetrà avançar en la comprensió del funcionament del cervell i estudiar les seves disfuncions. DARPA, per la seva part, llança l'*Urban Challenge* perquè els automòbils autònoms obeeixin les normes de trànsit i operen en un entorn urbà.

El 2010, Microsoft introdueix al mercat Kinect per a Xbox 360, el primer dispositiu de videojocs que rastreja el moviment del cos humà amb només una càmera 3D i detecció d'infrarojos, cosa que permet als usuaris de la consola jugar sense fils. L'aprenentatge automàtic per a aquesta tecnologia de captura del moviment el va desenvolupar el grup Computer Vision AI Microsoft Research de Cambridge. El 2013, el robot HRP-2 (de SCHAFT, subsidiari de Google), guanya a 15 equips en les proves de desafiament de robòtica de DARPA. L'HRP-2 va obtenir 27 de 31 punts en 8 tasques com per exemple conduir un vehicle, tirar les escombraries, pujar escales, travessar portes, tancar vàlvules i connectar una mànega. El mateix any NEIL, l'aprenent d'imatges infinit, es presenta a la Universitat Carnegie Mellon per comparar i analitzar constantment les relacions entre diferents imatges. El 2015 se signa una carta per prohibir el desenvolupament i ús d'armes autònomes que firmen Hawking, Musk,

Wozniak i 3.000 investigadors en robòtica i IA més. L'any següent, l'AlphaGo de Google DeepMind va guanyar Lee Sedol (el campió mundial) en el joc de taula xinès 'go'.

2.1.3.3. Situació actual: realitat i desafiaments

En els darrers deu anys l'IA ha fet un gir de 180 graus. No obstant això, hi ha qui creu que aquest canvi comença l'any 2007, amb l'aparició dels primers altaveus intel·ligents. Sigui com sigui, hem de tenir en compte que l'IA ens ha permès, durant molts anys, traslladar una gran quantitat de coneixement humà a format digital. Avui dia, no només la podem fer servir per funcionar com si fos un humà, sinó també per generar més coneixement i accions mitjançant la computació automatitzada (Bryson, 2019). Durant les dècades anteriors, l'interès principal de l'IA era crear artefactes intel·ligents amb forma i comportament humà, però ara com ara, superada ja aquesta fase, va molt més enllà. La capacitat de descobriment de l'IA no només s'ha expandit per l'increment massiu de les dades digitals i la capacitat de computació, sinó també per les innovacions en els algorismes. Avui dia busquem fotografies, vídeos i àudio (Barrett et al., 2016; Wu et al., 2016) i podem traduir, transcriure, llegir els llibres, interpretar emocions, falsificar signatures i manipular vídeos. Podem falsificar àudios o vídeos de retransmissions en directe, cosa que ens permet triar les paraules de les quals seran testimonis milions de persones, sobretot en el cas de persones famoses, com els polítics, sobre els quals hi ha una gran quantitat de dades per compondre models precisos (Thies et al., 2016; Suwajanakorn et al., 2017). Hi ha hagut exemples com els resultats de les eleccions presidencials dels EUA de 2016 i el referèndum del Brexit en què es diu que hi hauria hagut vots alterats mitjançant el encaminament de missatges de votants indecisos a través de les xarxes socials (Cadwalladr, 2017a, 2017b; ICO, 2018). L'IA és a la nostra disposició i ens pot beneficiar. Gràcies als avenços en robòtica, l'IA és, cada cop més, part del nostre dia a dia en forma de vehicles, drons i dispositius domèstics autònoms, inclosos els altaveus intel·ligents i les consoles de videojocs (Jia et al., 2016). Cada cop estem més envoltats de percepcions, anàlisis i accions automatitzades generalitzades. Si ens fixem amb els dispositius que hi ha ara com ara en el mercat, cal que recordem algunes definicions. Els robots són artefactes que senten i actuen en el món físic i en temps real, de manera que un telèfon intel·ligent podria ser un robot

(domèstic). A més d'amb micròfons, té diversos tipus de sensors que li permeten saber quan la seva orientació canvia o desapareix. Pot interactuar amb l'usuari i transmetre-li informació, instruccions incloses, a altres dispositius tal com poden fer alguns altres dispositius com consoles de videojocs i, el més important en el nostre cas, els assistents digitals domèstics: els altaveus intel·ligents, com Google Home i Echo, d'Amazon (Alexa).

L'IA és fonamental per a algunes de les empreses amb més èxit de la història en termes de capitalització del mercat: Apple, Alphabet, Microsoft i Amazon. Com les tecnologies de la informació i de la comunicació (TIC) en general, l'IA ha revolucionat l'accés de milers de milions de persones del món al coneixement. Aquest accés ha contribuït a una reducció massiva de la desigualtat mundial i la pobresa extrema, ja que ha permès que, per exemple, els grangers puguin saber quins són els preus justos i els cultius més renibles, i fins i tot que puguin tenir accés a unes prediccions meteorològiques precises (Aker i Mbiti, 2010). No obstant això, hi ha tecnòlegs i part del públic en general que plantegen algunes preocupacions. Smitch (2018), president de Microsoft, va afirmar que la tecnologia intel·ligent planteja assumptes que apel·len a la protecció dels drets humans fonamentals com el dret a la privacitat o la llibertat d'expressió, cosa que augmenta la responsabilitat de les empreses tecnològiques que creen els productes, però que també reclama una regulació governamental raonada i el desenvolupament d'unes normes sobre els bons usos. Durant segles, la substitució dels treballadors per mitjà de la tecnologia ha provocat temors de gran rellevància. No hi ha cap dubte que les noves tecnologies estan canviant les societats, les famílies i les vides, però tampoc que, històricament, la majoria dels canvis han estat positius (Pinker, 2012).

En general, l'esperança de vida és més elevada i la mortalitat infantil és més baixa que mai. Si ho mirem des d'aquest punt de vista, són bons indicadors de la satisfacció dels humans, ja que la baixa mortalitat infantil està molt relacionada amb l'estabilitat política (King i Zeng, 2001). No obstant això, hi ha disrupcions que ens porten al conflicte polític i, segons algunes hipòtesis, podrien estar relacionades amb l'apogeu de l' IA. La desigualtat de renda mostra una correlació molt elevada amb la polarització política (McCarty et al., 2016). La polarització política es defineix com la incapacitat dels partits polítics de cooperar en un sistema de govern democràtic, però els períodes

de polarització es caracteritzen per un increment de polítiques identitàries i extremismes. La polarització política i la desigualtat de renda estan correlacionades, però no se sap la causa ni quina serà la conseqüència atès que no se saben ben bé els factors subjacents (Stewart et al., 2018). Si se sap, però, que l'última vegada que els indicadors van ser tan elevats com ara va ser immediatament abans i després de la Primera Guerra Mundial, i van fer falta moltes dècades d'innovació de les polítiques, crisi financera mundial i la Segona Guerra Mundial perquè la desigualtat i la polarització es reduïssin i s'estabilitzessin en el període entre 1945 i 1978 (Scheidel, 2017), tot i que a alguns països com els Estats Units i el Regne Unit, només hi va haver la crisi financera.

Si reflexionem sobre l'impacte l'IA en el comportament individual, podem concloure que és el camp on n'és més evident l'impacte. Hem viscut llargs períodes d'espionatge domèstic, però ara com ara, les TIC ens permeten mantenir registres de llarga durada de qualsevol individu que produeixi dades emmagatzemables. No són només l'emmagatzematge i l'accessibilitat dels registres digitals el que canvia la nostra societat sinó també el fet que aquests registres es poden escrutar mitjançant algorismes de reconeixement de patrons. Hem perdut la premissa per defecte de l'anonimat per opacitat (Selinger i Hartzog, 2017). Fins a cert punt, tots som famosos: a qualsevol ens pot identificar un desconegut, ja sigui mitjançant un programa de reconeixement facial, per les dades de les nostres compres o pels nostres hàbits a les xarxes socials (Pasquale, 2015). A més, fins i tot es pot saber la nostra identitat, les nostres predisposicions polítiques o econòmiques i quines estratègies podrien ser eficaces per a modificar-les (Cadwalladr, 2017a, b). L'AA ens permet descobrir nous patrons i periodicitats que fins ara eren inconcebibles com, per exemple, que l'elecció d'unes determinades paraules o, fins i tot, la pressió que exercim quan escrivim a mà amb un llapis digital poden indicar un estat emocional (Bandyopadhyay i Hazra, 2017; Hancock et al., 2007), o que un patró d'ús de xarxes socials prediu categories de personalitat, preferències polítiques i, fins i tot, assoliments (Youyou et al., 2015). L'aprenentatge automàtic ha obert les portes a unes capacitats gairebé humanes i fins i tot sobrehumanes en la transcripció de la parla a partir de la veu i el reconeixement d'emocions a partir de la gravació d'àudio i de vídeo, entre d'altres (Valstar i Pantic, 2012; Thies et al., 2016; Deng et al., 2017).

D'altra banda, com millor sigui el model del qual disposem sobre la probabilitat que algú faci alguna cosa, menys informació necessitarem per deduir quin serà el seu següent pas (Bishop, 2006; Youyou et al., 2015). Aquest principi permet la falsificació a partir de l'elaboració d'un model sobre l'escriptura o la veu d'una persona, combinar-ho amb un text i produir una «predicció» o transcripció de com és de probable que la persona escrigui o pronuncii el text (Haines et al., 2016; Reed et al., 2016). Aquest mateix principi podria permetre a estratègics polítics identificar quins votants són susceptibles, si no de canviar de partit, almenys d'incrementar o reduir la probabilitat d'anar a votar i, consegüentment, dedicar recursos a convèncer-los que ho facin. (Cadwalladr, 2017).

L'IA ens permet conèixer, aprendre, descobrir i fer coses que semblaven impossibles fa cinquanta anys. Podem, fins i tot, passejar per una ciutat desconeguda on no sabem l'idioma, i ser capaços d'orientar-nos i comunicar-nos (Breslow et al., 2013). Amb tot, hem de ser conscients dels problemes que se'n deriven, dels quals ja n'hem parlat anteriorment: privacitat, llibertat i innovació. Tot i que no podem saber amb seguretat on ens portaran les tendències actuals, és d'esperar que, quan s'instauri algun tipus de repressió, l'IA i les TIC seran els mitjans que es faran servir. Es diu que a la Xina estan fent servir el reconeixement facial no només per identificar persones, sinó també per interpretar el seu estat d'ànim i concentració, tant en l'àmbit de la reeducació com en escoles. Els estudiants i fins i tot els professors poden ser castigats si aparenten no estar contents amb la seva educació (o reeducació). Els sistemes TIC capaços de detectar el grau de comprensió i atenció dels estudiants i d'informar els professors perquè adaptin la docència i els materials, també s'estan promovent a Occident i són una part essencial de la formació personalitzada amb IA fora de les aules convencionals. No obstant això, encara s'estan desenvolupant i, probablement, s'estan aplicant sistemes semblant en altres sectors professionals. Si permetem que aquesta tendència es mantingui, podem esperar que les nostres societats guanyin en seguretat (o, com a mínim, que hi hagi més pau als carrers) i homogeneïtat, però perdin capacitat d'innovació i diversitat. Més persones que mai tenen els mitjans i els recursos econòmics per viatjar d'un país a un altre, de manera que és probable que alguns països ofereixin una millor qualitat de vida, també en termes de governança i protecció individual, i atrauran les persones amoïnades per la llibertat personal. Podem esperar

també que, gràcies al poder combinat de la feina i la innovació d'aquests immigrants i de la població existent, aquests països puguin arribar a protegir-se i, fins i tot, protegir-ne d'altres (Bryson, 2018). Una tendència que pot amoïnar en la governança de l'IA és la tendència a adoptar l'alineament de valors com la solució a problemes de la ciència en general i d'ètica de l'IA en particular. La idea és que ha de garantir que la societat dirigeix i aprova el rumb que adopten la ciència i la tecnologia (Soares i Fallenstein, 2014). Tot i que això pot semblar força segur i democràtic, potser no ho és, ja que la ciència és un mecanisme basat en principis que permet a la societat percebre el seu context amb precisió. Al contrari, la governança és com eligeix una societat entre diferents vies d'acció possibles. El sentiment popular no pot decidir què és cert sobre la naturalesa, només determinar quines polítiques són més fàcils d'aplicar. Limitar la capacitat d'una societat a percebre només allò que vol saber equivaldria a encegar-la (Caliskan et al., 2017). D'una manera semblant, el sentiment popular influeix força en les polítiques que s'apliquen, però de cap manera les determina. Per tant, preguntar a la població què espera de l'IA és com preguntar-li quina pel·lícula de ciència-ficció preferiria que es fes realitat: ni hi ha cap garantia que s'elegeixi alguna cosa factible. Els ciutadans han de determinar, a través del govern, les seves prioritats econòmiques i polítiques, però el progrés gairebé mai no s'aconsegueix mitjançant un referèndum. Pel contrari, la governança normalment consisteix en negociacions informades entre un nombre limitat de negociadors experts, amb l'ajuda de persones que coneixen la matèria. No obstant això, més enllà dels problemes econòmics i de governança, hem de recordar que l'IA amplia i millora què significa ser humà i en particular, la nostra capacitat de resolució de problemes. Davant dels desafiaments globals de seguretat i sostenibilitat, les millores poden continuar aportant una ajuda significativa, sempre que desenvolupem mecanismes adequats per a regular-les. Amb un catàleg sensat de polítiques i organismes de regulació, hauríem de continuar ampliant i, també, limitant quan correspongui, el potencial d'aplicació de l'IA.

2.1.3.4. El *Natural Language Processing* (NLP)

Per fer-nos a la idea, el processament del llenguatge natural (NLP, per les seves sigles en anglès), és un àmbit de la intel·ligència artificial que proporciona a les màquines la capacitat de llegir, entendre i derivar el significat de la llengua humana.

Tot allò que expressem, sigui per escrit o verbalment, conté una enorme quantitat d'informació. El tema que triem, el to o les paraules que fem servir: tot aporta alguna informació que podem extreure per interpretar-la i valorar-la. En teoria, podem entendre i fins i tot predir el comportament humà a partir d'aquesta informació. No obstant això, hi ha un problema: una persona pot generar centenars o milers de paraules en una mateixa oració, i cada oració té un nivell de complexitat. Si volem abastar més i analitzar diversos centenars, milers o milions de persones o oracions, la situació es pot complicar encara més.

Les dades que es generen a partir de converses, declaracions o fins i tot tuits són exemples de dades no estructurades. Les dades no estructurades no encaixen en l'estructura tradicional de files i columnes de les bases de dades relacionals, i representen la majoria de les dades disponibles en el món real. Aquest tipus de dades, estan molt desordenades i, per tant, són difícils de manipular. No obstant això, gràcies als avenços en disciplines com l'aprenentatge de les màquines, i com bé hem assenyalat en els apartats anteriors, ara com ara està havent una revolució. Avui dia ja no s'intenten interpretar textos o discursos basats en les paraules clau abans, sinó que es mira d'entendre el significat que hi ha darrere de les paraules (el contingut cognitiu). D'aquesta manera és possible detectar figures del discurs com la ironia o, fins i tot, fer una anàlisi dels sentiments. L'NLP és una disciplina que se centra en la interacció entre la ciència i el llenguatge humà, i s'està expandint a moltes indústries diferents. Ara com ara està en apogeu gràcies a les enormes millores en l'accés a les dades i l'augment de la potència de càlcul, que estan permetent que els professionals aconseguixin resultats força importants en àrees com la salut, els mitjans de comunicació, l'economia o els recursos humans, entre d'altres.

En termes simples, l'NPL representa l'ús automàtic del llenguatge natural humà com la parla o el text i, tot i que el concepte en si ja és fascinant, el valor real d'aquesta tecnologia prové dels casos d'ús. L'NPL pot ajudar a fer diverses tasques i es pot aplicar en diversos àmbits, que cada dia en són més. El processament del llenguatge natural permet el reconeixement i la predicció de malalties basant-se en els registres sanitaris electrònics i en la pròpia parla del pacient. Aquesta capacitat s'està explorant en condicions de salut que van desde malalties cardiovasculars fins a altres com la

depressió o l'esquizofrènia. Per exemple, Amazon Comprehend Medical és un servei que fa servir l'NLP per extreure les condicions de les malalties, els medicaments i els resultats dels tractaments a partir de les notes dels pacient en els informes dels assajos clínics i altres registres electrònics de salut. D'altra banda, les organitzacions poden determinar el que els clients diuen sobre un servei o producte si identifiquen i extreuen informació de fonts com les xarxes socials. Aquesta anàlisi del sentiment pot proporcionar molta informació sobre les eleccions dels clients i o sobre quins factors van determinar la seva elecció. A més, un treballador d'IBM va desenvolupar un assistent cognitiu que funciona com un motor de cerca personalitzat que aprèn tot sobre la persona que el fa servir com noms, cançons o qualsevol altra cosa i li pot recordar en el moment que ho necessiti. Un altre exemple és la gestió del correu electrònic d'empreses com Yahoo i Google, que filtren i classifiquen els correus gràcies a l'NLP, analitzant els correus que s'envien a través dels seus servidors i desviant el correu brossa fins i tot abans que arribin a la safata d'entrada. El grup d'NLP del MIT va desenvolupar un sistema capaç de determinar si una font d'informació és políticament parcial i correcta i detectar quines són fiables i quines no, per ajudar a identificar les notícies falses. Un altre exemple, i el més important per a aquest treball són les interfícies intel·ligents que funcionen per veu (com els altaveus d'Amazon i de Google). Aquest tipus de dispositius fan servir l'NLP per respondre a les indicacions vocals i fer les accions que calguin com trobar una botiga en particular, dir la predicció del temps, suggerir una ruta millor, encendre i apagar els llums de casa i, en el nostre cas, fer traduccions de certes paraules/ oracions. També es força interessant en l'àmbit financer, ja que permet tenir una visió sobre què passa i de què parla la gent. S'utilitza, així doncs, per rastrejar notícies, informes o comentaris sobre possibles fusions entre empreses i després s'incorpora a un algoritme de comerç per generar beneficis massius. L'NLP també s'està fent servir en fases de cerca i selecció de personal per identificar potencials treballadors. Amb la tecnologia IBM Watson NLP, LegalMotion va desenvolupar una plataforma per automatitzar les tasques sistemàtiques dels litigis i per ajudar els equips legals a estalviar temps, reduir els costos i canviar l'enfocament estratègic. Tot i això, l'NLP està particularment en apogeu en la indústria de la salut. Aquesta tecnologia ha permès la millora en la diagnosi de malalties i està també reduint els seus costos, mentre que les organitzacions d'atenció mèdica estan, cada cop més, adaptant-se als registres de salut electrònics. El fet que la documentació clínica es pugui

millorar significa que es pot entendre millor als pacients i se'n poden beneficiar, ja que es pot tenir una millor cura de la seva salut. L'objectiu d'això és millorar la seva experiència, i moltes organitzacions ja hi estan fent tot el possible. Empreses com Winterlight Labs ja han obtingut força millores en el tractament de la malaltia de l'Alzheimer mitjançant el monitoratge del deteriorament cognitiu a través de la parla i han creat les bases per assajos i estudis clínics per un gran ventall de trastorns del sistema nerviós central. La Universitat de Stanford, amb un enfocament semblant, va desenvolupar Woebot, un terapeuta de chatbot amb l'objectiu d'ajudar les persones amb ansietat i altres trastorns. No obstant això, hi ha força controvèrsia al voltant d'aquest tema. Fa uns anys, Microsoft va demostrar que analitzant grans mostres de consultes dels motors de cerca, podien identificar els usuaris d'Internet que patien càncer de pàncrees fins i tot abans de rebre un diagnòstic de la malaltia. Això, evidentment, va comportar controvèrsia perquè no se sabia com reaccionarien els pacients a aquest tipus de diagnòstic ni se sabia què passaria si es fes una anàlisi amb un fals positiu. Una cosa semblant li va passar a Google el 2009, en el cas Google Flu Trends que assegurava que era capaç de predir la grip, però més tard va desaparèixer per la seva falta de precisió i incompliment de les tasques que es predien. Així doncs, podem dir que l'NLP pot ser clau en l'àmbit clínic però queden encara problemes per resoldre a curt termini.

El problema principal al qual ens enfrontem quan parlem de l'NLP és que el llenguatge és molt difícil. El procés de comprendre i manipular el llenguatge és molt difícil i, per això, és normal que es facin servir diverses tècniques per afrontar els diferents desafiaments que es presenten abans d'unificar-les. Per posar en marxa aquestes tècniques, es fan servir llenguatges de programació com Python o R, però abans de summergir-se en les línies del codi (un tema força més complex) és important entendre els conceptes que hi ha al darrere. A continuació, explicaré quins són els algorismes que es fan servir més en l'NLP per definir el vocabulari de termes. En primer lloc, tenim el model *bag of words* (bossa de paraules, en català), que és un model que es fa servir força perquè permet comptar el nombre de paraules d'un text. Bàsicament, és un mètode que s'utilitza en el processament del llenguatge per representar documents ignorant l'ordre de les paraules i la gramàtica. Aquest enfocament no proporciona ni context ni significat. D'altra banda, tenim la tokenització, que és el procés de segmentar el text en oracions i paraules. És la tasca de tallar a trossos que anomenem *tokens* i a la

mateixa vegada descartar alguns caràcters com la puntuació. Tot i que pot semblar un procés força bàsic, en llengües com el català o l'anglès, en què separem les paraules amb espais, si pensem una mica conclourem que fins i tot en català, per exemple, els espais no són suficients per fer tokenitzacions correctes, ja que separar per espais pot trencar paraules que s'haurien de considerar com un únic *token*, com 'Nova York'. A més, la tokenització elimina la puntuació, cosa que fa més fàcil la segmentació però que també pot comportar alguns problemes en els textos biomèdics, per exemple, on hi ha molts guions, parèntesis i d'altres signes de puntuació. També tenim l'eliminació de les anomenades paraules buides. Aquest procés inclou desfer-se dels articles, pronoms i preposicions comuns que aporten poc o cap valor a l'objectiu de l'NLP i es filtren i s'exclouen del text que s'ha de processar per eliminar termes molt freqüents que no aporten cap mena d'informació al text. Les paraules buides es poden ignorar fàcilment fent una cerca de les paraules clau i alliberant espai de la base de dades per millorar el temps de processament. No existeix una llista universal de paraules buides: es poden preseleccionar algunes o directament fer una llista des de 0. És recomanable, però fer-ne una llista de predefinides i afegir-ne de noves gradualment. No obstant això, sembla que últimament es tendeix a no fer servir cap llista perquè, de vegades, eliminar les paraules buides pot eliminar informació pertinent i fins i tot modificar el context d'algunes frases (per exemple, si eliminem la paraula buida 'no'). Hi ha també el mètode *stemming* (de la paraula 'arrel' en anglès '*stem*') que és el procés de tallar el principi o el final de les paraules amb la intenció d'eliminar els afixos (tant els prefixos com els sufixos). El problema és que els sufixos poden crear noves paraules. Per solucionar-ho, és possible considerar certes llistes d'afixos (Python i R tenen algunes biblioteques que contenen afixos) i fer-ne un rastreig, però el problema no acaba aquí. Com els *stemmers* fan servir algorismes, és possible que el resultat del procés de *stemming* no sigui una paraula real o que fins i tot canviï el significat de la paraula i conseqüentment de l'oració. El *stemming* té algunes limitacions, però és important per corregir errors d'ortografia, són fàcils d'utilitzar i funcionen molt ràpidament (un dels aspectes més importants de l'NLP). Hem de tenir en compte que aquest procés es fa servir per millorar el rendiment i no pas la gramàtica. També hi ha el procés de lematització, que consisteix en, amb una forma flexionada (plural, femení, verb conjugat) trobar el lema (la paraula que s'accepta com a estàndard) corresponent, és a dir, el lema és la paraula com la trobaríem en un diccionari convencional. Tot i que aquest procés és força semblant al de *stemming*, la

lematització fa servir un enfocament diferent per arribar a l'arrel de la paraula. La lematització, com fa servir les formes del diccionari, necessita diccionaris detallats amb què l'algoritme pugui examinar i vincular les paraules al seu lema corresponent. La lematització també té en compte el context de la paraula per resoldre altres problemes com la desambiguació. Per això, aquest procés requereix molts més recursos que el procés de *stemming* sobre l'estructura del llenguatge i, com a conseqüència, exigeix més potència de càlcul que la creació o adaptació d'un algoritme de *stemming*. L'últim procés que descriurem és el *topic modeling*, que és un mètode per descobrir estructures ocultes en conjunts de textos o documents. En essència, agrupa textos per descobrir els temes basant-se en el contingut, processant les paraules individuals i assignant valors segons la seva distribució. Aquesta tècnica es basa en què cada document està compost per diversos temes i que cada tema agrupa un conjunt de paraules, cosa que fa que hàgim de descobrir els temes ocults per saber el significat dels nostres textos. Del món de les tècniques de *topic modeling*, la més utilitzada és la LDA, un algoritme que es va inventar fa uns vint anys i que funciona amb un mètode d'aprenentatge no supervisat que troba els significats dels documents. En els mètodes d'aprenentatge no supervisat com aquest, no hi ha cap variable que guiï el procés d'aprenentatge i les dades s'exploren mitjançant algoritmes per trobar patrons. L'LDA troba grups de paraules relacionades de dues maneres. En la primera, assigna a cada paraula un tema a l'atzar (l'usuari defineix el nombre de temes), de manera que l'algoritme assigna a tots els documents els temes perquè les paraules de cada document tinguin cadascuna el seu tema imaginari. De la segona manera, l'algoritme recorre cada paraula de manera interactiva i reassigna la paraula a un tema tenint en compte la probabilitat que la paraula pertangui a un tema i la probabilitat que el document estigui generat per un tema. Les probabilitats es calculen diverses vegades. El *topic modeling* és molt útil per classificar textos, crear sistemes de recomanació (llibres basats en lectures anteriors) o fins i tot per detectar tendències en les publicacions en línia.

Ara per ara, l'NLP està lluitant per detectar matisos en el significat del llenguatge, ja sigui per falta de context, errors d'ortografia o diferències dialectals. El març del 2016, Microsoft va llançar Tay, un *charbot* d'IA a Twitter com a prova. La idea era que com més usuaris interactuessin amb Tay, més intel·ligent seria. El resultat va ser que van haver d'eliminar Tay 16 hores després a causa dels seus comentaris racistes i

abusius. Mesos després, Microsoft va aprendre de l'experiència i va llançar Zo, la segona generació de *charbot* en anglès que no va cometre els mateixos errors que el seu avantpassat. Tot i que el futur sembla força ple de desafiaments i amenaces per a l'NLP, la disciplina s'està desenvolupant de forma molt ràpida i és probable que en els anys vinents moltes aplicacions que semblen impossibles, seran possibles.

2.2. La traducció automàtica

En aquest apartat farem un breu resum sobre què és la traducció automàtica i quins tipus hi ha per poder entendre, més endavant, en un exercici pràctic, quin tipus de sistema fa servir cada altaveu i també com funciona cadascun.

Entenem per traducció automàtica la transformació, fent servir un sistema informàtic, d'un text informatitzat escrit en la llengua de partida, en un altre text informatitzat escrit en la llengua d'arribada.

Els sistemes de traducció automàtica són aplicacions o serveis en línia que fan servir les tecnologies de l'aprenentatge automàtic per traduir grans quantitats de text des de i a qualsevol dels idiomes que suporten. El servei tradueix un text de partida d'una llengua a una altra llengua d'arribada. Tot i que els conceptes darrere de la tecnologia de la traducció automàtica i les interfícies per utilitzar-la són relativament simples, la ciència i les tecnologies que hi ha al darrere són extremadament complexes i reuneixen tecnologies d'avantguarda, particularment el *deep learning* (intel·ligència artificial), una gran quantitat de dades, lingüística, Cloud Computing i APIs Web.

Històricament, la tècnica d'aprenentatge de màquines primària que es feia servir a la indústria era la traducció automàtica estadística (TAE). La TAE utilitza l'anàlisi estadística avançada per estimar les millors traduccions possibles per unes paraules amb un context d'unes poques paraules. Tots els principals proveïdors de serveis de traducció fa des de mitjans dels anys 2000 que utilitzen la TAN. L'arribada de la traducció automàtica neuronal (TAN) va causar un canvi radical en la tecnologia de la traducció, i com a resultat, va proporcionar traduccions d'una qualitat molt més elevada. Aquesta tecnologia de traducció es va començar a fer comuna entre traductors i desenvolupadors a finals del 2016. Les tecnologies de traducció TAE i TAN tenen dos

elements en comú. Per una banda, totes dues requereixen grans quantitats de contingut traduït per entrenar els sistemes. A més, no actuen com a diccionaris bilingües, traduint paraules basades en una llista de traduccions potencials, sinó que tradueixen basant-se en el context de la paraula que es fa servir en una oració.

Ara com ara es podria considerar que conviuen, principalment, tres tipus de traducció automàtica (TA): la que utilitza informació lingüística (TA basada en regles), l'estadística (TAE) i la basada en xarxes neuronals (TAN). Parlarem de l'evolució i funcionament de cadascuna per separat i hem de tenir en compte que ara com ara n'hi ha dues que es fan servir especialment: la traducció automàtica estadística (TAE) i la més nova, la traducció automàtica neuronal (TAN).

La traducció basada en regles (TABR), si bé es continua investigant, és la que menys atenció acapara. Una de les possibles raons per les quals aquest tipus de traducció no té gaire presència en la investigació és perquè requereix una gran inversió de temps i recursos i, d'altra banda, cada cop més, s'estan obtenint millors resultats amb tècniques més modernes. Per desenvolupar una única combinació de llengües calen les gramàtiques de la llengua de partida d'arribada, així com amb diccionaris bilingües i regles de transferència. A més, el sistema no pot traduir estructures lingüístiques que no estiguin incloses en les seves gramàtiques o regles de transferència, ni paraules o expressions que no hi hagi en els diccionaris. Això fa que el manteniment d'aquests tipus de motors basats en regles sigui pràcticament continu per garantir que siguin capaços de traduir textos d'un nou domini, per exemple, o estructures gramaticals que en un primer moment no es van preveure. Si bé aquests tipus de sistemes tenen inconvenients, també presenten grans avantatges. En el cas de les llengües semblants entre si, com per exemple el castellà i el català, s'ha demostrat que en alguns escenaris poden obtenir millors resultats que els motors estadístics.

D'altra banda, els enfocaments dels sistemes de TA basats en corpus (Hearne, Way, 2011:1) es fonamenten en l'ús de corpus paral·lels. Això és, els textos es tradueixen automàticament fent servir la informació obtinguda d'exemples de traduccions anteriors fetes per humans. Les dues estratègies principals de TA que engloba aquest paradigma són la TA basada en exemples (EBMT, per les sigles en anglès) i la TA estadística (TAE). Mentre que els sistemes EBMT tradueixen només

fent servir la noció de "similitud" amb oracions vistes prèviament, els sistemes TAE no funcionen així. Segons Hearne i Way (2011) la TA estadística fa servir dos processos diferents i separats: l'entrenament i el *decoding*. La fase d'entrenament consisteix a extraure un model estadístic de traducció d'un corpus paral·lel, i un model estadístic de la llengua de destí a partir d'un (normalment més gran) corpus monolingüe (Brown et al. 1990, 1993). El model de traducció comprèn un diccionari bilingüe en el qual cada possible traducció d'una determinada paraula o frase d'origen té una probabilitat associada. No obstant això, el model no sembla a un diccionari convencional perquè en aquest sistema moltes de les entrades representen traduccions que són improbables però no impossibles, i les probabilitats associades ho reflecteixen. El model lingüístic comprèn una base de dades de seqüències de paraules en la llengua de destí (normalment d'entre 1 i 7 paraules de longitud), cadascuna de les quals també està associada a una probabilitat. Durant l'entrenament es pot extraure informació addicional més enllà dels models lingüístics i models de traducció, com per exemple models de longitud relativa de les frases, reordenació de les paraules, importància relativa de la traducció respecte al model lingüístic, etc. Aquests models es fan servir més tard en el *decoding*, el procés que realment produeix una traducció. El procés de descodificació tracta essencialment la traducció com un problema de cerca: amb la frase a traduir, se cerquen totes les traduccions possibles permeses pel model de traducció, i tots els reordenaments possibles perquè se li assigni la que tingui una probabilitat global més gran segons els models de traducció i llengua. Els sistemes de TAE, en vegada de centrar-se en el millor procés per generar una única traducció òptima per a una oració en llengua de partida, se centra a generar milers de traduccions hipotètiques per a la cadena d'entrada, i després en esbrinar quina és més probable. La traducció automàtica estadística, a partir d'una gran quantitat de text en la llengua de partida i en la llengua d'arribada, alineat i no alineat, cerca la millor traducció possible de cadascun dels elements de l'oració, des del segment més gran possible (frase completa) fins al més petit (paraula). La qualitat de la traducció depèn del corpus de treball (corpus monolingües i corpus bilingües), el temps de resposta (quantitat d'opcions valorades, quantitat de text processat). La majoria d'aquests sistemes de TAE duen a terme processos de segmentació, *truecasing* i neteja de corpus. Segmentar significa separar amb espais les paraules dels signes de puntuació. Aïllar la puntuació permet incrementar les probabilitats d'obtenir coincidències amb els textos que es tradueixen

automàticament. El procés de *truecasing*, en canvi, consisteix a determinar el *casing* més probable de cada paraula, majúscules o minúscules. La neteja consisteix en la suspensió de les frases llargues i alineades malament dels corpus amb l'objectiu de minimitzar els problemes en la fase d'entrenament. Després el sistema processa les dades lingüístiques proporcionades en la fase d'entrenament en la qual, a partir de l'anàlisi de coocurrències de paraules i segments en les dues llengües, s'infereixen de manera automàtica correspondències de traducció. El resultat de l'entrenament és el model de traducció, format per una taula de frases, un model de llengua i, ocasionalment, una taula de reordenament. Com la consulta de les taules pot ser lenta, els models es binartizen perquè es carreguin més ràpidament. Finalment, l'optimització (o *tuning*) és un procés que determina automàticament els valors òptims d'un seguit de paràmetres perquè el motor generi les millors traduccions possibles. En el procés d'optimització, es tradueixen automàticament milers de frases d'un subconjunt dels models (anomenat *development* o *turning set*), es comparen amb les traduccions humanes de referència i s'ajusten automàticament els valors de cada paràmetre per millorar la qualitat del motor, amb mètriques automàtiques com per exemple el BLEU (Papineni, Roukos, Ward, i Zhu, 2002).

Les millores contínues en la traducció són importants. No obstant això, les millores de rendiment de la TAE es van començar a estancar a mitjans de la dècada del 2010. És aquest en el moment que comença l'apogeu de la TAN, que inicia una nova dècada de millora de la qualitat de la traducció. Les traduccions de xarxes neuronals es diferencien fonamentalment en la forma en què es fan en comparació amb la TAE. Microsoft, a la seva pàgina oficial, mostra una imatge dels passos que segueixen les traduccions a través de les xarxes neuronals per traduir una oració (Figura 14).

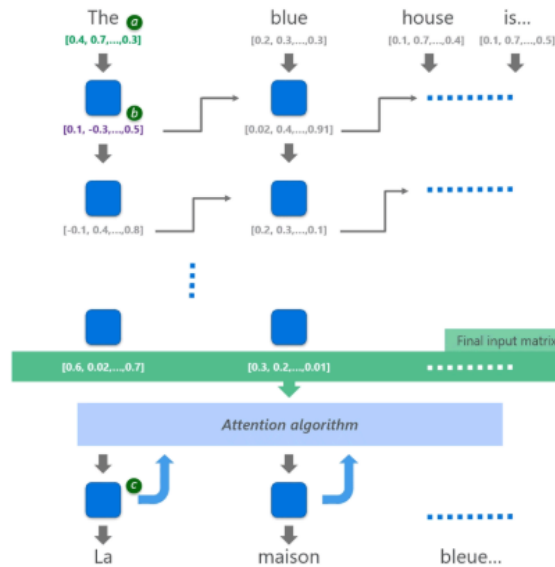


Figura 14. Gràfica de funcionament de la TAN.

La traducció agafarà, així doncs, el context de la frase completa i no només unes quantes paraules com faria la TAE, i produirà traduccions més fluides i semblants a les que podrien traduir els éssers humans. Pel que fa a l'entrenament de les xarxes neuronals, cada paraula està codificada amb un vector de 500 dimensions (a) que representa les seves característiques úniques en un parell de llengües en particular (per exemple, anglès i xinès). Basant-se en els parells de llengua que es fan servir per a l'entrenament, la xarxa neuronal autodefinirà quines han de ser les seves dimensions. Podrien codificar conceptes simples com el gènere (femení, masculí, neutre), el nivell de cortesia (argot, casual, escrit, formal, etc.), el tipus de paraula (verb, substantiu, etc.), però també qualsevol altra característica no obvia derivada de les dades d'entrenament. Els passos que segueix la TAN són els següents. En primer lloc, cada paraula, o més específicament el vector de dimensió 500 que la representa, passa per una primera capa de 'neurones' que la codificarà en un vector de 1000 dimensions (b) que representi la paraula en el context de les altres paraules de l'oració. Un cop totes les paraules s'han codificat una vegada en aquests vectors de la dimensió 1000, el procés es repeteix diverses vegades, cosa que fa que per cada capa d'aquesta representació de 1000 dimensions de cada paraula s'ajusti millor al context de l'oració completa. La matriu de sortida és la que farà servir la capa de conversió (per exemple, un algoritme de *software*) que utilitzarà tant aquesta matriu de sortida com el resultat de les paraules traduïdes prèviament per definir quina paraula, tenint en compte el text en llengua de

sortida, s'hauria de traduir a continuació. Farà servir també aquests càlculs per eliminar paraules no necessàries en la llengua d'arribada. La capa de descodificació (traducció), tradueix la paraula seleccionada (o més específicament, el vector de dimensió 1000 que representa la paraula en el context de l'oració completa) en el seu equivalent de llengua d'arribada més adient. El resultat d'aquesta última capa (c) torna a la capa de conversió per calcular quina és la següent paraula de l'oració en llengua de partida que s'ha de traduir.

Gràcies a aquest sistema, la producció final és, en la majoria dels casos, més fluida i més propera a una traducció humana que mai.

3. Marc pràctic

En aquest apartat faré, en primer lloc, un breu introducció al que són els altaveus intel·ligents. No obstant això, no faré una extensa explicació sobre aspectes tècnics, ja que ja hem esmentat tots els que són rellevants per dur a terme la prova en els anteriors apartats. A continuació, també faré una petita descripció del funcionament i les característiques de tots dos altaveus, tant de Google Home com d'Echo Dot, comentaré els aspectes més importants per realitzar la prova principal i, per fer paleses algunes de les funcions, faré alguna petita prova. Després d'això, passaré a fer la pràctica. En el meu cas, m'he basat en una prova que va fer una gran empresa de traducció dels Estats Units (que explicaré més tard). En total, faré la prova amb 50 unitats lingüístiques entre les quals hi ha paraules, sintagmes, oracions amb diferents entonacions, refranys, referents culturals i oracions cèlebres de pel·lícules i de la història. Explicaré, també, quines paraules/oracions he triat i per què, i faré una anàlisi de les respostes que proporcionen cadascun dels altaveus. Traduiré frases del castellà a l'anglès i al francès, i més endavant ja explicaré els motius de la meva elecció. Separaré, en una primera instància, les explicacions de cadascun dels altaveus i, finalment, faré una comparativa dels diferents resultats que obtingui en la qual tindrè en compte, des del punt de vista d'un posteditor humà, quines són les traduccions 'millorables'. Per fer-ho, assignaré a cadascuna de les traduccions que obtingui una puntuació del 0-10, cosa que em facilitarà, a més, fer un gràfic de resultats al final i poder extreure conclusions basades en fets.

3.1. Els altaveus intel·ligents

És important, quan parlem dels altaveus intel·ligents, saber de què parlem. Un altaveu intel·ligent és un dispositiu que, amb un assistent virtual integrat, ofereix accions interactives i es pot activar sense tocar cap botó amb el que s'anomena una *'hot word'*. Un altaveu intel·ligent també es podria definir com un dispositiu de reproducció d'àudio sense fils i intel·ligent que fa servir diversos tipus de connectivitat per a funcions addicionals. El seu objectiu original era lliurar, sense fils, contingut d'àudio des de la biblioteca multimèdia de l'usuari. Com qualsevol producte electrònic, els altaveus intel·ligents varien força quant a les seves característiques, disseny i qualitat.

Aquests dispositius, normalment, formen part del conjunt de productes existents d'una empresa. Alguns exemples són Google Home i Amazon Echo. Normalment, els altaveus intel·ligents inclouen connectivitat WiFi i Bluetooth. Els productes van des de les caixes d'expansió habilitades per a WiFi fins als altaveus intel·ligents RGB LED que es poden estendre per tota la casa para obtenir un so estèreo distribuït o localitzat. Alguns altaveus intel·ligents tenen una estació base amb un altaveu levitant magnètic que es pot col·locar per casa. Aquestes estacions base sovint presenten no només el component de l'estació base WiFi sinó també un *subwoofer* per al so de baixa freqüència, cosa que la converteix en una estació de greus. Alguns altaveus intel·ligents poden, gràcies a la connexió WiFi o el Bluetooth, fins i tot controlar una casa sencera, i d'altres, que tenen pantalles incorporades, poden mostrar respostes visuals. Un altaveu intel·ligent amb una pantalla tàctil se l'anomena pantalla intel·ligent. És un dispositiu Bluetooth intel·ligent que integra una interfície d'usuari conversacional amb pantalles de visualització para augmentar la interacció de la veu amb imatges i vídeo. Tal com els altaveus intel·ligents, ofereixen controls per a dispositius domèstics intel·ligents, aplicacions de transmissió de característiques i navegadors web amb controls tàctils per seleccionar el contingut. Hi ha diversos dispositius al mercat, amb els seus respectius assistents virtuals, de diferents marques. Podem trobar el dispositiu de Yandex, Yandex Station i Irbis A, amb l'assistent Alice incorporat; el d'Alibaba Group, Tmall Genie, que incorpora l'assistent AliGenie; la gamma Amazon Echo series, Amazon Tap i Sonos One, que incorporen Alexa; el HomePod d'Apple, que fa servir l'assistent Siri; els dispositius de Baidu Xiaoyu i RavenH, que funcionen amb l'assistent DuerOS Open Platform; Wave i Firends, de Naver & Line Corporation, amb l'assistent Clova; els Google Home series, de Google, amb l'assistent amb el mateix nom; Galaxy Home, de Samsung Electronics que funciona amb l'assistent Bixby; el de Microsoft, Harman Kardon INVOKE, que incorpora Cortana; i, per últim, tenim alguns en desenvolupament com el de Tencent i alguns europeus com Magenta de Deutsche Telekom. No obstant això, i com ja hem assenyalat amb anterioritat, ens centrarem en el funcionament de la TA DE l'assistent Google Assistant, de Google, gràcies al dispositiu Google Home Mini; i d'altra banda amb el funcionament de la TA d'Alexa, d'Amazon, a través de l'Echo Dot.

3.1.1. Echo Dot Alexa

Alexa és l'assistent virtual desenvolupat per Amazon i que podem controlar directament per veu. El 2014 va aparèixer el primer dispositiu i altaveu Echo d'Amazon que incorporava l'assistent de veu Alexa. El seu nom no és una casualitat ni una preferència dels desenvolupadors, sinó que es va triar minuciosament perquè l'assistent pogués reconèixer millor el seu propi nom perquè inclou una 'X'. El seu funcionament és força fàcil i ja familiar per a molts dels usuaris: pronunciem el seu nom i, tot seguit, l'acció que volem que faci, la pregunta que volem que ens respongui o l'aplicació que volem fer servir. Tot això, directament sense haver de prémer cap botó, directament amb la nostra veu.

Podem preguntar gairebé qualsevol cosa a Alexa, i ella intentarà contestar-nos de la millor manera possible. Hi ha preguntes que ens contestarà directament gràcies a les *skills* que té integrades predefinidament i d'altres que cercarà a Internet. Les *skills* són complements o aplicacions com les que podem instal·lar en els nostres telèfons o tauletes i n'hi ha de tota mena: des de jocs fins a diaris digitals. Algunes de les *skills* més comunes avui dia són les que permeten automatitzar la casa i realitzar funcions com obrir i tancar els llums, pujar o baixar les persones, etc.

En el nostre cas, ens centrarem en les llengües que incorpora, que ara com ara són el castellà, l'anglès, el francès, l'alemany, l'italià i el japonès. De fet, per saber-ho, només cal preguntar-li 'Alexa, ¿qué idiomas hablas?' i ens respon directament els idiomes en què està disponible. Com podem veure, en aquest punt m'he trobat un inconvenient per fer el meu treball perquè Alexa no parla el català. Per no canviar massa d'idioma i ja que tinc dos idiomes nadius, he decidit configurar l'altaveu en castellà per fer les proves de traducció. La traducció integrada amb què compta Alexa és la traducció automàtica neuronal de Microsoft Translator, i crida especialment l'atenció que, quan rebem la resposta de les traduccions, Alexa ens parla amb accent nadiu.

Per traduir correctament la veu que rep en l'idioma origen a la llengua d'arribada, el sistema passa per quatre fases. En la primera, el reconeixement automàtic de veu transforma l'àudio en text. En segon lloc, la tecnologia de Microsoft TrueText que normalitza el text perquè sigui més apropiat per a la traducció. Després, es tradueix el text a través de la TAN però seguint els models que es desenvolupen especialment per

traduir converses de veu. Per últim, es converteix el text en àudio per produir la traducció en àudio.

L'esquema del procés seria aquest:

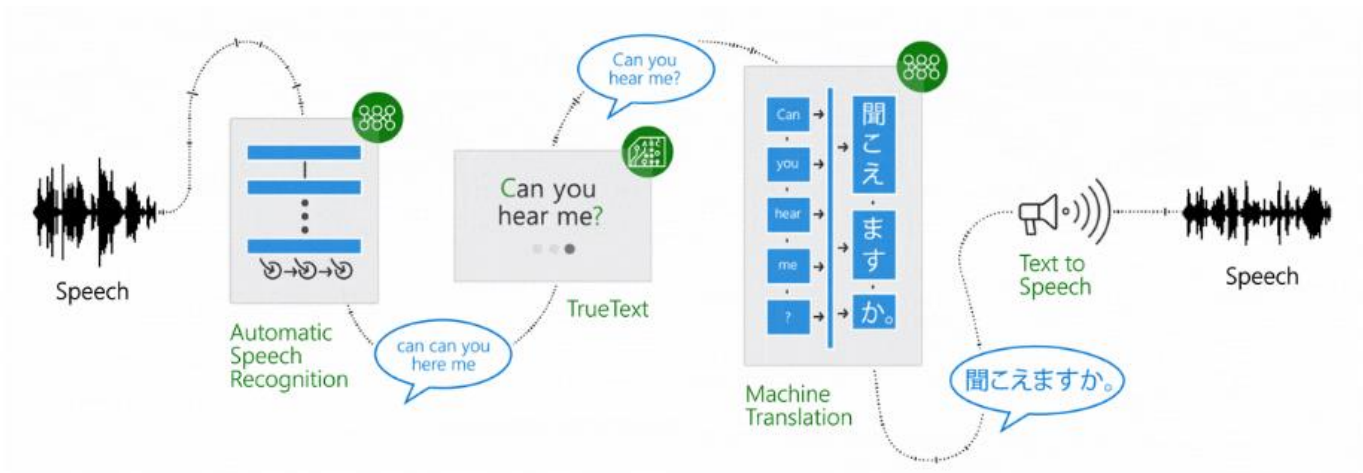


Figura 18. Procés de la traducció *speech-to-speech* amb la tecnologia de Microsoft.

Aquest sistema de traducció fa servir el sistema de reconeixement automàtic de veu (ASR) que es fa utilitzant un sistema de xarxa neural entrenat en l'anàlisi de milers d'hores de veu d'àudio entrant. Aquest model s'entrena amb interaccions humà-humà en lloc de humà-màquina, de manera que produeix un reconeixement de veu optimitzat per a converses normals. Para aconseguir-ho, calen moltes dades, així com una xarxa neuronal molt més gran que l'ASR tradicional de persona-màquina.

Cal explicar, d'altra banda, què és el TrueText de Microsoft. Com a éssers humans que conversen amb altres éssers humans, no parlem tan perfectament, clarament o curiosament com creiem que ho fem. Amb la tecnologia TrueText, el text literal es transforma per reflectir més estrictament la intenció de l'usuari, ja que elimina les paraules que sobren como "um", "ah", "i", "com", quequeses i repeticions. El text també es fa més llegible i traduïble mitjançant l'addició de pauses d'oració, puntuació adequada i l'ús de majúscules. Per aconseguir aquests resultats, Microsoft ha desenvolupat TrueText. A continuació, podem veure un diagrama on es mostra la transformació de TrueText per normalitzar el text literal.



Figura 19. Esquema dels sistema TrueText.

L'equip de Microsoft continua treballant en el seu desenvolupament per fer-ne una eina cada cop millor. Alexa, per la seva part, té aquesta *skill* integrada amb la qual realitzarà les seves traduccions en la prova de traducció que farem més endavant. Un dels missatges més clars que Amazon subrtalla és que Alexa en castellà no és una mera traducció l'anglès, sinó una adaptació a la cultura i les expressions espanyoles. Per això han creat un equip local a Espanya que treballa perquè Alexa entengui els diferents accents i respongui els diferents comandaments.

3. 1. 2. Google Home

Google, com a la majoria dels seus productes, fa servir el seu nom comercial per anomenar tant el dispositiu com per a l'assistent virtual (Google Assistant). El Google Home permet, entre d'altres, escoltar música, reproduir vídeos, escoltar les notícies, etc. Com en l'Echo Dot, també és possible automatitzar la casa i controlar-la per veu. El primer dispositiu, Google Home, es va llançar als Estats Units al novembre del 2016 i entre el 2017 i el 2019 es van produir els llançaments mundials. Des de llavors, amb l'evolució en aquest tipus de tecnologia, s'han desenvolupat una gran quantitat d'actualitzacions i noves funcions. Quant als idiomes que parla i, per tant pot traduir, m'he trobat en el mateix problema que amb l'Echo Dot d'Alexa: no parla el català. En canvi, es pot configurar i traduir a les llengües que es poden veure en aquest quadre:

País o regió	Llengües admeses
Austràlia	Anglès
Àustria	Alemany
Bèlgica	Neerlandès i alemany
Brasil	Portuguès
Canadà	Anglès i francès
Dinamarca	Danès
França	Francès
Alemanya	Alemany
Índia	Anglès (Índia), anglès (EUA) i hindi
Indonèsia	Indonesi
Irlanda	Anglès (Regne Unit)
Itàlia	Italià
Japó	Japonès
Països Baixos	Neerlandès
Noruega	Noruec
Singapur	Anglès (Singapur) i anglès (EUA)
Espanya	Castellà
Suècia	Suec
Suïssa	Alemany i italià
Taiwan	Mandarí (Taiwan)
Tailàndia	Tailandès
Regne Unit	Anglès
Estats Units	Anglès

La característica més rellevant que té integrat el dispositiu de Google Home és el mode d'interpret a temps real. A diferència de quan li preguntem per la definició d'una paraula o frase, l'assistent, en aquest cas, escolta tot allò que diem fins que acabem de parlar i reproduceix el missatge en la llengua d'arribada. Per activar aquesta funció, cal que diguem al nostre assistent “Hey Google, quiero que seas mi intérprete de inglés” (o qualsevol altre idioma). Després d'això, l'altaveu emetrà un soroll cada cop que

comenci a escoltar què diem i quan deixem de parlar, començarà a reproduir en la llengua d'arribada tot allò que hem dit.

Quant a la traducció per veu que fan servir aquests dispositius per traduir, Google Home fa servir la TAN de Google. Segueix un procés força semblant al que hem esmentat amb Microsoft, però tots dos sistemes es diferencien perquè Google fa servir el seu sistema propi de TAN i, després de processar totes les paraules que li demanem que tradueixi en la llengua en què tenim configurat l'altaveu, ens torna el missatge oral en la llengua d'arribada.

3.2. Prova de traducció

En aquest apartat, passarem a fer ja la prova de traducció amb els diferents sistemes que he mencionat anteriorment en els dos altaveus intel·ligents en qüestió: Echo Dot i Google Home. Com ja he explicat, cap dels dispositius té l'opció ni de configurar-se en català ni de traduir al català. Així doncs, he configurat tots dos altaveus en castellà. Com la meua segona tercera llengua és l'anglès, faré la traducció ES > EN en primer lloc, i com també sé francès, faré una segona prova amb totes les unitats lingüístiques que explicaré a continuació ES > FR.

Per fer la prova, he seleccionat 6 categories de 10 proves cadascuna que repetirem dues vegades, una per a la prova a l'anglès, i l'altra per a la prova en francès. En la primera prova, els altaveus traduiran 10 substantius simples de 5 àmbits diferents: ciències, tecnologia, llengua, dret i arts. En la segona, 10 sintagmes nominals dels mateixos 5 àmbits. En la tercera, formularé 10 frases amb diferents temps verbals i entonacions. En quart lloc, faré una prova d'oracions útils (per poder fer servir, per exemple, per defensar-nos quan viatgem a un país on no coneixem la llengua). En cinquè lloc, provaré la traducció de 5 refranys i de 5 referents culturals. Per últim, i inspirada en la prova que va fer One Hour Translation, provaré la traducció de 5 frases cèlebres de la història i de 5 frases mítiques de pel·lícules.

Després d'això, puntuaré les traduccions que n'obtingui de l'1 al 3, sent el 3 la millor nota i l'1 la pitjor. Per últim, i després de comprovar les unitats lingüístiques amb què faré la prova des d'un punt de vista humà, elaboraré un gràfic per esclarir el

rendiment de la TA de cadascun dels altaveus. Finalment, faré un comentari dels resultats obtinguts.

3.2.1. Prova 1: substantius simples de diferents especialitats

Després de fer les primeres proves aleatòries, he conclòs que Alexa entén que vols traduir una oració/ paraula quan li dius “Alexa, di [...] en anglés” (o una altra llengua) o, si estàs traduint una sola paraula, també li pots dir “Alexa, traduce [...] al anglés” (o una altra llengua). Així doncs, abans de traduir les unitats lingüístiques que li demanaré que tradueixi a continuació, hem de tenir en compte que sempre ‘invocarem’ l’Alexa d’aquesta manera.

D’altra banda, Google Home entén millor quan li demanes “Hey Google, ¿cómo se dice [...] en anglés?” (o una altra llengua) o també “Hey Google, traduce [...] al anglés” (o una altra llengua, de manera que sempre li demanarem que tradueixi mitjançant aquesta ‘fórmula’.

Com hem indicat abans, en aquesta primera prova traduirem 10 paraules simples de 5 temàtiques diferents (ciències, tecnologia, llengua, dret i arts). Així doncs, n’hi haurà dues de cada àmbit.

3.2.1.1. Del castellà a l’anglès

En aquesta prova indicaré, mitjançant una taula, la paraula en castellà, el resultat que ofereix Alexa en anglès, el resultat que ofereix l’Assistent de Google en anglès i el resultat que un traductor humà (jo) hauria posat.

Text de partida	Alexa	Google	Suggeriment humà
Invernadero	Greenhouse	Greenhouse	Greenhouse
Intoxicación	Intoxication	Poisoning	Poisoning
Teclado	Keyboard	Keyboard	Keyboard
Inalámbrico	Cordless	Wireless	Wireless
Paralingüística	For linguistics	Paralinguistics	Paralinguistics
Metáfora	Metaphore	Metaphore	Metaphore

Sentencia	Sentence	Judgment	Judgment
Querella	Quarell	Complaint	Complaint
Aglutinante	Binder	Binder	Binder
Caballete	Easel	Easel	Easel

3.2.1.2. Del castellà al francès

En aquesta prova, traduirem les mateixes paraules dels mateixos àmbits, però en vegada de traduir-les a l'anglès, en aquest cas, les traduirem al francès. Una altra vegada, crearé una taula on inclouré les paraules en castellà, l'opció d'Alexa, l'opció de Google i l'opció d'una traductora humana (jo).

Text de partida	Alexa	Google	Suggeriment humà
Invernadero	Serre	Serre	Serre
Intoxicación	Intoxication	Empoisonnement	Intoxication
Teclado	Clavier	Clavier	Clavier
Inalámbrico	Sans fil	Sans fil	Sans fil
Paralingüística	Pour la linguistique	Paralinguistique	Paralinguistique
Metáfora	Métaphore	Métaphore	Métaphore
Sentencia	Phrase	Jugement	Jugement
Querella	Coup d'arrêt	Plainte	Plainte
Aglutinante	Relier	Classeur	Liant
Caballete	Chevalet	Chevalet	Chevalet

El principal problema d'aquesta prova ha estat que l'altaveu Alexa, no ha reconegut la majoria de paraules a la primera i les he hagut de repetir diverses vegades. Podem veure, fins i tot, que 'paralingüística' és una paraula que no ha entès cap de les vegades. He hagut de repetir diverses vegades alguns mots perquè els processés i emetés una resposta. A més, Alexa, de primeres, triga força més que Google a reaccionar. Per la part de Google, considero que ha fet molt bones traduccions de manera general i que, per tant, i com ja explicaré més endavant, podria servir-nos per traduir paraules simples

a diferents llengües. Cal esmentar, també, que no hi ha hagut cap àmbit específic en què els traductors hagin experimentat més problemes que en altres: s’han defensat força bé en tots.

3.2.2. Prova 2: sintagmes de diferents especialitats

En aquesta prova, continuaré fent servir les mateixes formulacions de pregunta per als altaveus. Com ja he esmentat, aquesta prova és semblant a l’anterior però en vegada d’amb mots simples, faré la prova amb sintagmes de dues o més paraules. Els àmbits continuen sent els mateixos i per tant, només canviaran la composició de les unitats lingüístiques.

3.2.2.1. Del castellà a l’anglès

L’estructura d’aquesta prova i en la majoria de les següents serà aquesta: text de partida (en castellà), suggeriment d’Alexa en anglès (en aquest cas), suggeriment de Google en anglès, i traducció suggerida per una traductora humana (jo).

Text de partida	Alexa	Google	Suggeriment humà
Células madre	Stem cells	Mother cells	Stem cells
Ácido desoxirribonucleico	Deoxyribonucleic acid	Deoxyribonucleic acid	Deoxyribonucleic acid
Comercio electrónico	E-commerce	E-commerce	E-commerce
Página de inicio	Home page	Home page	Home page
Sintagma nominal	Nominal sintagma	Noun phrase	Noun phrase
Análisis paradigmático	Paradigmatic analysis	Paradigmatic analysis	Paradigmatic analysis
Recurso de apelación	Appeal	Appeal	Appeal
Norma jurídica	Legal rule	Legal rule	Precept
Cuento de hadas	Fairytale	Fairytale	Fairytale
Pintura flamenca	Flemish painting	Flemenco painting	Flemish painting

3.2.2.2. Del castellà al francès

En aquest apartat, procediré a fer exactament el mateix que en l'anterior, però en vegada de traduir a l'anglès, ho faré al francès.

Text de partida	Alexa	Google	Suggeriment humà
Células madre	Cellules souche	Cellules souche	Cellules souche
Ácido desoxirribonucleico	Acide désoxyribonucléique	Acide désoxyribonucléique	Acide désoxyribonucléique
Comercio electrónico	Commerce électronique	E-commerce	Commerce électronique
Página de inicio	Page d'accueil	Page d'accueil	Page d'accueil
Sintagma nominal	Sintagma nominal	Phrase nominal	Phrase nominal
Análisis paradigmático	Analyse paradigmatiche	Analyse paradigmatiche	Analyse paradigmatiche
Recurso de apelación	Appel	Charm	Appel
Norma jurídica	Règle juridique	Norme legal	Règle de droit
Cuento de hadas	Conte de fées	Conte de fées	Conte de fées
Pintura flamenca	Peinture flamande	Peinture flamenco	Peinture flamande

En aquesta prova, tots dos traductors m'han sorprès d'una manera grata, si bé és cert que han comès alguns errors greus, sobretot en anglès. He de subratllar, sobretot, que el contingut de les proves l'he triat jo i, per tant, la resposta dels altaveus està condicionada i no té un valor científic, però sí empíric. Tot i que, en aquest cas, he mirat de triar sintagmes que no fossin massa complicats perquè sóc conscient que en les proves que he fet no hi ha context, he posat algunes paraules amb trampa i, de vegades, hi ha caigut. Amb el terme 'pintura flamenca', el traductor de Google no ha entès cap de les dues vegades que es referia a l'estil de pintura i no pas a la dansa. A més, ha traduït una paraula que, tot i que és específica de l'àmbit de la ciència, és força comuna en la llengua ('cèl·lules mare') per '*mother cells*', una traducció massa literal. Contràriament, el traductor d'Alexa ha fet unes traduccions força bones en aquesta prova tret d'alguns sintagmes que, en pronunciar-los, no ha entès correctament.

3.2.3. Prova 3: oracions

La prova 3 consisteix, com hem esmentat anteriorment, en traduir oracions amb ple significat. Aquestes frases són frases comunes de la llengua, tot i que algunes, són subordinades per saber fins quin punt poden traduir els altaveus frases completes. A més, hi ha algunes una mica més difícils per provar, així, l'efectivitat de la traducció automàtica que fan servir cadascun dels dispositius en situacions de complexitat més alta.

3.2.3.1. Del castellà a l'anglès

En aquest cas, continuarem fent servir la mateixa estructura que en les proves anteriors, però eliminarem una columna: la del suggeriment humà. Això es deu al fet que en aquesta prova hi ha més marge d'error o, si més no, de diferència perquè les oracions es poden formular i dir de moltes maneres diferents. A més, hem de tenir en compte que són oracions descontextualitzades i poden tenir diverses interpretacions.

Text de partida	Alexa	Google
Te prometo que queríamos volver pronto	I promise we wanted to come back soon	I promise you we wanted to come back soon
Te lo iba a decir	I was going to tell you	I was going to tell you
Puede que aun no lo haya visto	I may not have seen it yet	He may not have seen it yet
Tengo ganas de volver a la normalidad	I'm looking forward to getting back to normal	I want to go back to normal
El esfuerzo se valora, pero no es suficiente	Effort is valued, but it is not enough	Effort is valued, but it is not enough
Se lo voy a decir en cuanto llegue	I'm going to tell you as soon as I get there	I will tell you as soon as I arrive
No me importa para nada	I don't care at all	I do not care at all
No puede ser verdad	It can't be true	It cannot be true
Me encanta pasear por la playa	I love walking on the beach	I love to walk on the beach
Avísame cuando llegues	Let me know when you get	Let me know when you

	there	arrive
--	-------	--------

3.2.3.2. Del castellà al francès

Com del castellà a l'anglès, el procés serà el mateix: no crearé la columna de suggeriment humà perquè les oracions poden tenir interpretacions i significats diferents depenent del punt de vista en què es mirin. Com estan descontextualitzades, em limitaré a veure què fan els traductors i, més tard, les puntuaré segons el seu significat i fluïdesa.

Text de partida	Alexa	Google
Te prometo que queríamos volver pronto	Je te promets qu'on voulait revenir bientôt	Je vous promets que nous voulions revenir bientôt
Te lo iba a decir	J'allais te le dire	J'allais te dire
Puede que aun no lo haya visto	Je ne l'ai peut-être pas encore vu	Vous ne l'avez peut-être pas encore vu
Tengo ganas de volver a la normalidad	J'ai hâte de revenir à la normale	Je veux revenir à la normale
El esfuerzo se valora, pero no es suficiente	L'effort est valorisé, mais il ne suffit pas	L'effort est apprécié, mais il ne suffit pas
Se lo voy a decir en cuanto llegue	Je te le dirai dès que j'y serai	Je vous le diré de mon arrivée
No me importa para nada	Je m'en fiche	Je ne me soucie pas du tout
No puede ser verdad	Ce n'est pas vrai	CA ne peut pas être vrai
Me encanta pasear por la playa	J'adore marcher sur la plage	J'adore marcher sur la plage
Avísame cuando llegues	Faites-moi savoir quand vous y arrivez.	Faites le moi savoir quand vous arrivez

En aquest apartat, cap dels dos traductors ha comès cap error greu. Això, a parer meu, significa dues coses. En primer lloc, tot i que no és cap sorpresa, que tots dos motors de traducció automàtica neuronal estan molt entrenats i que, per tant, fan la seva utilitat de traduir d'una llengua a una altra. A més, això significa, consegüentment que

són perfectament útils per traduir frases simples. Així doncs, podríem dir que serveixen per saber de què tracten alguns textos d'idiomes que no coneixem, o si més no, per tenir-ne una idea. Cal destacar que, tot i que de vegades els resultats de la traducció han estat diferents, no significa que cap dels dos estigui malament perquè, com he esmentat abans, són oracions descontextualitzades. Si hagués de comentar cap cosa, seria la traducció de 'No puede ser verdad' al francès perquè hi ha un 'CA' en majúscula que no té gaire sentit.

3.2.4. Prova 4: preguntes simples

En aquesta prova, avaluaré la capacitat que els traductors automàtics tenen incorporats els dispositius Echo Dot i Google Home de salvar-nos en situacions de comunicació oral i permetre que ens comuniquem de manera clara o, si més no, fer-nos entendre. En aquest cas, les frases seran més curtes i simples que en les proves anteriors atès que estem avaluant un nivell bàsic i elemental de traducció, però amb l'afegit que formularem una pregunta i no una oració enunciativa.

3.2.4.1. Del castellà a l'anglès

Seguirem l'estructura amb text de partida, traducció de l'Alexa a l'anglès i traducció de Google a l'anglès, però en aquest cas, partint de l'una hipòtesi que pot funcionar, innovaré una mica quant a com traduir amb l'altaveu de Google. Com he esmentat en parlar de les seves característiques, si li diem a Google "Hey Google, sé mi intérprete de inglés" (o un altre idioma dels que té disponibles), podem fer que cada cop que se sent un petit soroll i diem una frase en veu alta en castellà (en aquest cas), quan fem una pausa el traductor reproduïx en la llengua que li hem dit la frase/ paraula que li hem dit. Així doncs, com Alexa encara no incorpora aquesta opció, continuaré amb el mateix mètode, però amb Google provaré el mode intèrpret.

Text de partida	Alexa	Google
¿Cómo estás?	How are you?	How are you?
¿Qué hora es?	What time is it?	What time is it?
¿Dónde está la puerta?	Where's the door?	Where is the door?
¿Cuántos años tienes?	How old are you?	How old are you?

¿De dónde eres?	Where are you from?	Where are you from?
¿Podrías ayudarme?	Could you help me?	Could you help me?
¿Tienes hambre?	Are you hungry?	Are you hungry?
¿Estás bien?	Are you ok?	You're good?
¿Dónde vives?	Where do you live?	Where do you live?
¿A qué te dedicas?	What do you do?	What is your job?

3.2.4.2. Del castellà al francès

Continuarem fent servir la mateixa metodologia que en l'anterior: el mode intèrpret per a Google i el mateix procés de sempre per l'Alexa, però en aquest cas traduirem les frases al francès.

Text de partida	Alexa	Google
¿Cómo estás?	Comment vas-tu?	Comment ça va?
¿Qué hora es?	Quelle heure est-il?	Quelle heure il est?
¿Dónde está la puerta?	Où est la porte?	Où la porte est?
¿Cuántos años tienes?	Quel âge as-tu?	Quelle âge as-tu?
¿De dónde eres?	D'ou venez-vous?	D'ou viens tu?
¿Podrías ayudarme?	Pourriez-vous m'aider?	Pouvez-vous m'aider?
¿Tienes hambre?	Tu as faim?	Tu as faim?
¿Estás bien?	Tu vas bien?	Vous êtes dué?
¿Dónde vives?	Où habite-tu?	Où vis tu?
¿A qué te dedicas?	Que fais-tu?	Que fait-vous dans la vie?

En aquest cas, tots dos traductors han fet traduccions perfectes. Tot i que de vegades la traducció no ha estat la mateixa en un que en l'altre, han aconseguit plasmar perfectament les preguntes i, fins i tot, l'entonació, cosa que m'ha cridat força l'atenció. D'altra banda, m'agradaria comentar que el mode intèrpret de Google, en aquest cas, ha funcionat a la perfecció. Considero que és molt útil quan volem traduir coses que sabem que Google pot fer bé, com en aquest cas. És fàcil de fer servir, ràpid i dona molt bons resultats. Alexa, per la seva part, tot i que ha fet bones traduccions, continua amb el

problema que de vegades és difícil interactuar amb ella ja que molts cops no entèn que dius i, quan ho entèn, com ja he esmentat abans, li costa una mica produir la traducció.

3.2.5. Prova 5: refranys

En la prova 5 hem volgut afegir un esglao de dificultat. Aquesta, consisteix a traduir refranys del castellà a l'anglès i al francès. En el món de la traducció es diu sovint que és de les coses més difícils de traduir, ja que n'hi ha milers i sobretot, perquè depenent de l'àrea geogràfica dins d'un mateix país hi poden haver diverses variants. També, sovint cal buscar un equivalent més o menys universal que pugui tenir sentit i s'entengui a la majoria del territori al qual volem adreçar la nostra traducció. A més, és important destacar també que els refranys no son oracions que es puguin traduir literalment sinó que tenen una interpretació semàntica i, per tant, s'han d'interpretar i no cercar-ne el significat literal. Així doncs, he fet una recerca de refranys en castellà i els seus respectius equivalents en anglès i en francès. D'aquesta manera, podem considerar que aquest tipus d'expressions no es tradueixen, sinó que es transcreen o, si més no, es transformen.

3.2.5.1. Del castellà a l'anglès

En aquest cas sí que he inclòs la meva traducció en anglès perquè he fet una recerca sobre equivalències semàntiques de refranys en la llengua d'arribada. Així doncs, l'estructura torna a ser l'original: text de partida, traducció a l'anglès d'Alexa, traducció a l'anglès de Google i traducció humana.

Text de partida	Alexa	Google	Suggeriment humà
El que mucho abarca poco aprieta	The one that encompasses little squeeze	Jack of all trades, master of none	If you run after two hares you will catch neither
Quien calla otorga	Who's silent grants	Silence gives consent	Silence gives consent
La cabra siempre tira al monte	The goat always throws into the bush	The goat always throws the mountain	You can't teach an old dog new tricks
A caballo regalado	Don't look at his tooth	Don't give a tooth to a	Never look a gift

no le mires el diente	on a gift horse	gift horse	horse in the mouth
Matar dos pájaros de un tiro	Killing two birds with one stone	Kill two birds with one shot	Kill two birds with one stone
Dime con quién andas y te diré quién eres	Tell me who you're hanging out with and I'll tell you who you are	Tell me who are your friends and I'll tell you who you are	A man is known by the company he keeps
Buscar tres pies al gato	Look for three feet for the cat	Find three feet to the cat	Flogging a dead horse
No es oro todo lo que reluce	It's not gold all that shines	Not all that glitters is gold	All that glitters is not gold
Tirar la casa por la ventana	Throwing the house out the window	Throw the house out the window	Push the boat out
Llamar al pan pan y al vino vino	Call bread and wine wine	Call bread bread and wine wine	To call a spade a spade

3.2.5.2. Del castellà al francès

En aquest apartat, farem el mateix que en l'anterior. Jo, traductora humana, he buscat els equivalents en francès per als mateixos refranys que tot just he traduït a l'anglès. Així doncs, tornem a tenir 4 columnes: text original, traducció al francès d'Alexa, traducció al francès de Google i la traducció humana.

Text de partida	Alexa	Google	Suggeriment humà
El que mucho abarca poco aprieta	Celui qui englobe peu de compression	Jack de tous les métiers maître de rien	Qui trop embrasse mal étireint
Quien calla otorga	Qui est des subventions silencieuses	Le silence donne son consentement	Qui ne dit mot, consent
La cabra siempre tira al monte	La chèvre jette toujours dans la	La chèvre jette toujours la montagne	L'eau va toujours à la rivière

	brousse		
A caballo regalado no le mires el diente	Ne regarde pas sa dent sur un cheval-cadeau.	Ne donnez pas de dent à un cheval cadeau	A cheval donné, on ne doit en la bouche regarder
Matar dos pájaros de un tiro	Tuer deux oiseaux avec une pierre	Tuer deux oiseaux d'un seul coup	D'une pierre, deus cops
Dime con quién andas y te diré quién eres	Dis-moi avec qui tu traînes et je te dirai qui tu es	Dis-moi qui sont tes amis et je te dirai qui tu es	Dis-moi qui tu hantes, et je dirai qui tu es
Buscar tres pies al gato	Recherchez trois pieds pour le chat	Trouvez trois pieds au chat	Chercher midi à quatorze heures
No es oro todo lo que reluce	Ce n'est pas de l'or tout ce qui brille	Tout ce qui brille n'est pas or	Ce n'est pas la vache qui beugle le plus fort qui donne le plus de lait
Tirar la casa por la ventana	Jeter la maison par la fenêtre	Jetez la maison par la fenêtre	Jeter l'argent par les fenêtres
Llamar al pan pan y al vino vino	Appeler le pain et le vin	Appeler pain pain et vin vin	Appeler un chat chat

Personalment, diria que aquesta prova ha estat un fracàs tant per a Google com per a Alexa. L'únic que se'n podria salvar són les traduccions de Google dels refranys en anglès, i no totes. D'Alexa, seria més fàcil comptar les vegades que ha fet una bona traducció dels refranys que les que s'ha equivocat, sobretot en francès. Tots dos traductors han traduït de manera literal la majoria dels refranys que els hi he preguntat en castellà. No obstant això, hi ha alguns dels quals m'ha sorprès el resultat, com el de 'Quien mucho abarca poco aprieta', que Google ha traduït perfectament a l'anglès i al francès. En un cas concret, però, l'Alexa ha traduït 'Matar dos pájaros de un tiro' millor que Google, ja que no ha parlat de trets sinó de pedres, que és l'equivalent en anglès i en francès. Tanmateix, el resultat ha estat tan malament com m'esperava. Traduir refranys és força difícil, i moltes vegades requereix cercar durant hores, trucar un amic nadiu per preguntar-li, etc. cosa que, ara per ara, només podem fer els humans.

3.2.6. Prova 6: frases cèlebres

En aquesta prova, la sisena i última, m'agradaria estendre l'explicació una mica més. Una de les empreses de traducció més grans del món, OneHourTranslation (OHT), va dur a terme, l'any 2016, un estudi per esbrinar com de precisos són els assistents virtuals a l'hora de traduir fent servir la seva traducció automàtica neuronal (TAN) integrada. OHT va seleccionar 60 oracions de pel·lícules i personatges famosos, des del Padrino i el Mag d'Oz fins a Neil Armstrong, el primer home que va arribar a la lluna, passant pels presidents dels EUA Franklin Delano Roosevelt i John Fitzgerald Kennedy i figures històriques com Leonardo da Vinci i Isop. Les frases les van traduir Google Assistant, l'Alexa d'Amazon i l'assistent d'Apple, Siri, de l'anglès al francès, castellà, xinès i alemany, i després les van lliurar a cinc traductors professionals perquè les avaluessin en una escala de l'1 al 6. Google Assistant va obtenir la puntuació més alta en tres de les quatre llengües amb què es va fer l'estudi: anglès al francès, anglès a l'alemany i anglès al castellà, i va quedar segon lloc en la traducció de l'anglès al xinès. L'Alexa d'Amazon, va quedar primera en la categoria de traducció de l'anglès al xinès. La Siri d'Apple va ser la segona en la categoria de l'anglès al francès i de l'anglès al castellà, i tercera en la categoria de l'anglès a l'alemany i de l'anglès al xinès.

Així doncs, aquesta part del treball està basada en aquest estudi, en la traducció de frases cèlebres de figures famoses i oracions mítiques de pel·lícules. Podríem dir que aquest és l'apartat més específic i complex, ja que no només es tracta de traduir una oració o paraula o, fins i tot, com hem vist en la secció anterior, un refrany, sinó d'una frase que un cop va dir una persona i va esdevenir famosa. Sembla que, tenint en compte l'estudi que va fer OHT, Google Home en sortirà victoriós, però cal provar-ho per comprovar-ho.

3.2.6.2. Del castellà a l'anglès

En aquesta secció, em limitaré a veure què diuen els altaveus, sense cap suggeriment de traductora humana de manera que, més tard, per fer-ne l'avaluació, buscaré en detall quines haurien de ser les traduccions correctes de cadascuna de les frases. He cercat, com ara veurem, 5 frases cèlebres de persones famoses diferents i després, 5 frases més de pel·lícules que son de les més importants del món filmogràfic.

Text de partida	Alexa	Google
Pienso, luego existo	I think, then I exist	I think, therefore I exist
No hay caminos para la paz, la paz es el camino	There are no paths to peace, peace is the way	There are no way for peace, peace is the way
Solo sé que no sé nada	I just know I don't know anything	I only know that I know nothing
Ojo por ojo el mundo acabará ciego	Eye for an eye the world will end up blind	An eye for an eye and the world will go blind
Es tan corto el amor y tan largo el olvido	It's so short love and so long forgetfulness	Love is so short and forgetting so long
Mantén cerca a tus amigos, pero más cerca a tus enemigos	Keep your friends close, but closer to your enemies	Keep your friends close but your enemies closer
No existen preguntas sin respuesta, solo preguntas mal formuladas	There are no unanswered questions, only mis-asked questions	There are no unanswered questions, only poorly formulated questions
Puede que no sea muy listo, pero sé qué es el amor	I may not be very smart, but I know what love is	I may not be very smart, but I know what love is
Que la fuerza te acompañe	May the force be with you	May the force be with you
No quiero necesitarte porque no puedo tenerte	I don't want to need you because I can't have you	I don't need to need you because I can't have you

3.2.6.2. Del castellà al francès

Per acabar amb la pràctica, traduiré les mateixes oracions cèlebres al francès, tenint en compte que les traduccions ja estan encunyades perquè són oracions que són història. Així doncs, seguint l'estructura habitual, demanaré als dos altaveus intel·ligents que les tradueixin i, finalment, traurem les conclusions i veurem si els resultats coincideixen amb els resultats de l'estudi de OHT.

Text de partida	Alexa	Google
Pienso, luego existo	Je pense alors que j'existe	Je pense, donc se suis

No hay caminos para la paz, la paz es el camino	Il n'y a pas de chemin vers la paix, la paix est la voie	Il n'y a pas de moyen pour la paix, la paix est le chemin
Solo sé que no sé nada	Je sais juste que je ne sais rien	Je sais seulement que je ne sais rien
Ojo por ojo el mundo acabará ciego	Oeil pour un oeil le monde finira aveugle	Un oeil pour un oeil le monde finira aveugle
Es tan corto el amor y tan largo el olvido	C'est si peu d'amour et si longtemps d'oubli	L'amour est si court et l'oubli si long
Mantén cerca a tus amigos, pero más cerca a tus enemigos	Gardez vos amis proches, mais plus près de vos ennemis	Gardez vos amis proches, mais vos ennemis plus proches
No existen preguntas sin respuesta, solo preguntas mal formuladas	Il n'y a pas de questions sans réponse, seulement des questions mal posées	Il n'y a pas de questions sans réponse, seulement des questions mal formulées
Puede que no sea muy listo, pero sé qué es el amor	Je ne suis peut-être pas très intelligent, mais je sais ce qu'est l'amour	Je ne suis peut-être pas très intelligent, mais je sais ce qu'est l'amour
Que la fuerza te acompañe	Que la force soit avec vous	Que la force t'accompagne
No quiero necesitarte porque no puedo tenerte	Je ne veux pas avoir besoin de toi parce que je ne peux pas t'avoir	Je ne veux pas avoir besoin de toi car je ne peux pas t'avoir

Aquestes frases, tot i que no són les més difícils, tenen un problema afegit i és que tenen una traducció o són d'una manera concreta en la llengua original i per tant, no les podem canviar. Ningú no s'imagina la frase d'Star Wars d'una altra manera que no sigui 'Que la fuerza te acompañe' i ens sonaria força estrany sentir-ho d'una altra manera. És per això que és important saber si els dispositius són capaços de traduir tenint en compte aquests factors. El resultat, no obstant això, és força bo. La majoria de frases s'han traduït perfectament tot i que amb algunes variacions, però generalment, tant en anglès com en francès els resultats son molt fidels a la frase que correspon, si bé és cert que els cinèfils s'adonarien que hi ha variacions. En general, ha traduït millor

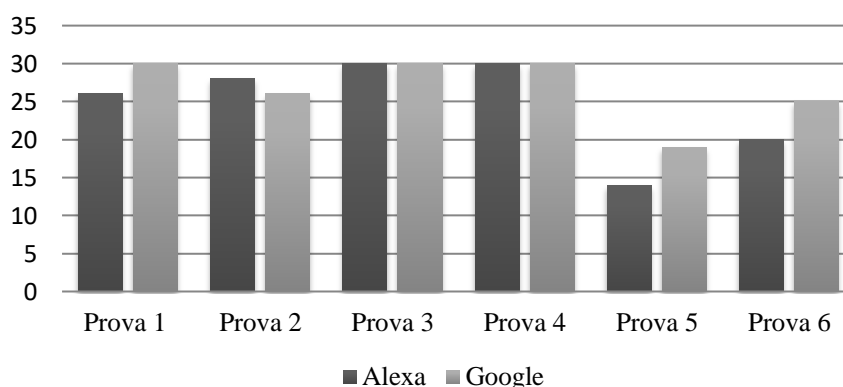
Google Home que Alexa a tots dos idiomes, però entrarem a parlar d'això a continuació amb l'explicació de les puntuacions que donaré a cada traducció i amb els gràfics.

3.2.7. Resultats

A continuació il·lustraré, amb gràfics i tenint en compte la puntuació que li he donat a cada traducció que ha fet l'Alexa a l'anglès i al francès, i el Google a l'anglès i al francès, quin dels dos funciona millor segons les diferents proves que he fet. Com ja he explicat anteriorment, puntuaré de l'1 al 3 cada resultat, sent el 3 la millor nota i l'1, la pitjor, de manera que 30 serà la millor puntuació i 10, la pitjor.

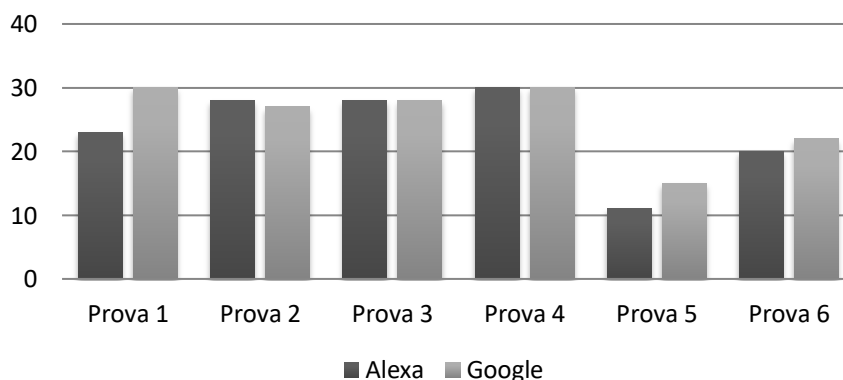
En aquest primer gràfic, mostraré els resultats que he obtingut de les proves de traducció a ES > EN, cosa que il·lustrarà el rendiment de cadascun dels traductors integrats dels altaveus.

Resultats traducció ES > EN



En el segon i últim gràfic, il·lustraré els resultats de les proves ES > FR.

Resultats traducció ES > FR



Amb aquests resultats podem concloure, de forma general, que el traductor que fa servir Google Home és, la gran majoria de les vegades, força millor que el que incorpora Alexa. Tant en francès com en anglès, Google Home ha proporcionat traduccions molt bones en les diferents proves. No obstant això, entrarem en detall a continuació prova per prova.

Quant a la prova 1, tant Alexa com Google Home han estat capaços de traduir gairebé tot a la perfecció, si bé és cert que Alexa ha tingut problemes en algunes paraules (no casualment, en les mateixes en francès i en anglès). A la prova 2, Google Home ha tret una pitjor nota que Alexa en anglès, però Google ha tornat a guanyar l'Alexa en la prova al francès. De totes maneres, el resultat de tots dos dispositius en aquesta prova és gairebé perfecte. Però per a perfectes, han estat els resultats de la prova 3, en què tant en francès com en anglès, Google Home i Alexa han oferit unes traduccions impecables. En la prova 4, com en la 3, s'han obtingut uns resultats immillorables, ja que tots dos dispositius han aconseguit la millor puntuació possible en anglès i, en francès. En la prova 5 és on han tingut més errades tant Google com Alexa. En anglès, però, Google ha aconseguit una puntuació per sobre dels 15 punts, que per la dificultat de la prova està força bé, mentre que Alexa ha obtingut una puntuació molt baixa. De fet, Alexa, en la traducció al francès d'aquesta prova ha obtingut la pitjor nota de tota la pràctica, només un punt per sobre de la nota mínima, mentre que Google, tot i que no ha aconseguit una nota espectacular, ha guanyat Alexa una altra vegada. En la prova 6 i última, Google ha tornat a guanyar Alexa en totes dues proves de traducció (a l'anglès i al francès). Tot i que en aquesta els resultats han estat força igualats perquè, fins i tot traduint de manera literal els resultats tenien sentit, hi ha frases encunyades que Google ha detectat i Alexa no.

Així doncs, podem concloure que el traductor integrat que fa servir Google Home obté millors resultats que el que té integrat l'Echo Dot d'Amazon, tant per traduir a l'anglès com per traduir al francès.

4. Conclusions

D'aquest treball podem extreure diverses conclusions. En primer lloc, com he demostrat en la meua pràctica, ara com ara el traductor integrat de Google Home funciona millor que el que té integrat l'Echo Dot, de Microsoft. No obstant això, hi ha moltes coses que calen tenir en compte. En primer lloc, com he explicat, la tecnologia canvia cada dia. De fet, mentre feia aquest treball m'he adonat que és força fàcil explicar què ha passat fins a arribar on som ara, però que és força difícil explicar en el punt on estem. Vivim en un món tecnològic, on cada dia hi ha avenços. Alguns surten bé i d'altres malament, però tant uns com els altres son igual d'importants perquè la tecnologia i, en general el mètode científic, funciona amb una orientació d'assaig i error.

D'altra banda, m'agradaria destacar, també relatiu amb la prova que he fet, que és evident que la traducció que incorporen aquests dispositius funciona, tradueix. També hem vist, però, que no transcrea. No obstant això, els desenvolupadors d'aquests dispositius, quan els comercialitzen a altres països, sí que fan un exercici de transcreació. Mentre feia les 'proves de so' abans de començar les pràctiques amb els altaveus, vaig provar de preguntar a l'Alexa "Alexa, ¿cuál es tu grupo favorito de música?" i em va respondre "No hay nada mejor que Mecano". Tot i que això pot no significar res, jo, com a localitzadora i traductora, considero que és important que es facin aquest tipus de transcreacions. Hagués pogut quedar igual de bé que l'Alexa digués "No hay nada mejor que The Beatles", però no ho fa.

Una cosa que m'ha sobtat i que, a la vegada, m'ha posat inconvenients a l'hora de fer la pràctica ha estat que els dispositius no estan disponibles en català. Potser m'ha sobtat més que no estigués disponible amb el Google Home, ja que té la majoria del seu web i productes localitzats. Tot i això, i per sort, com sóc bilingüe he pogut fer la pràctica de totes maneres, però no com esperava fer-la.

També m'ha sorprès gratament la nova *skill* de Google, el mode intèrpret. Quan feia les proves prèvies a la pràctica vaig descobrir que diu exactament el mateix que si ho escrivim al Google Translate, però és cert que dona una resposta molt més ràpida i eficaç. És per això que vaig incorporar una petita prova en la pràctica fent servir aquest mode.

La qualitat de traducció dels assistents virtuals ha estat força alta, de manera que podríem constatar que els altaveus intel·ligents sí serveixen per traduir frases curtes. De fet, se m'acudeixen diverses aplicacions en l'àmbit de la traducció, com per exemple i com he explicat abans, directament preguntar als assistents la definició d'una paraula o la traducció d'una paraula o frase. A més, ara com ara s'estan desenvolupant milers de *skills* que es poden instal·lar en els altaveus per ampliar-ne la funcionalitat. De fet, fent una cerca ràpida vaig trobar que hi ha empreses de traducció que ja n'estan desenvolupant algunes. A més, tot això em fa pensar una cosa: si tanta atenció s'està posant en desenvolupar aquest tipus d'assistent en la majoria dels àmbits (*marketing*, traducció, cuina, tecnologia de la llar, etc.) hi ha d'haver un motiu al darrere.

D'altra banda, cal destacar que la TAN, paral·lelament, evoluciona cada dia. És ja evident que hi ha moltes empreses interessades a desenvolupar i entrenar motors de traducció, i també és evident que cada dia són més útils, fan menys errors i van millor. De fet, cada dia s'està optant més, a les empreses, per fer servir la traducció automàtica i la postedició humana. En la majoria dels casos, estalvia temps i diners.

També vull indicar que aquest és un experiment empíric i no científic perquè, com ja he mencionat durant el treball, l'elecció de les paraules/ oracions ha estat deliberada cosa que fa que el treball i els resultats que he obtingut com a conseqüència siguin fruit d'una prova i no una evidència científica com hagués estat si hagués fet servir dades aleatòries.

Per últim, m'agradaria apuntar que aquest treball m'ha servit no només per aportar el meu gra de sorra en l'àmbit, sinó també per ampliar els meus coneixements en les branques de la computació, intel·ligència artificial, xarxes neuronals, *deep learning*, etc. És un fet que la tecnologia és aquí, en diferents formes i en diferents àmbits, i ja no podem mirar a un altre costat perquè, com dirien en anglès, és un *elephant in the room*.

5. Bibliografia

AAAI (2020). A Brief History of AI. Recuperat de: <https://aitopics.org/misc/brief-history> (Darrera consulta: 28 de maig de 2020)

Anyoha, R. (2017). The History of Artificial Intelligence. Harvard University, Estats Units. Recuperat de: <http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/> (Darrera consulta: 24 de maig de 2020)

Baker, J. K. (1979). Stochastic Modeling for Automatic Speech Understanding. Academic Press, New York, Nova York.

Baum, L.E. (1972) An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes, Inequalities, vol. 3, pàg. 1-8

Brown, P., Cocke J. et al. (1990). A statistical approach to machine translation. Thomas J. Watson Research Center, Nova York, Estats Units. Recuperat de: <https://www.aclweb.org/anthology/J90-2002.pdf> (Darrera consulta: 25 de maig de 2020)

Costa-Jussà, M., Fonollosa A.R. (2016). Character-based Neural Machine Translation. Centre de recerca TALP, Universitat Politècnica de Barcelona, Barcelona. Recuperat de: <https://arxiv.org/pdf/1603.00810.pdf> (Darrera consulta 20 de maig del 2020)

Dudley, H. (1939). The Vocoder, Bell Labs Record, Vol. 17, pàg. 122-126.

Fewster, G. (2017) The mind-blowing AI announcement from Google that you probably missed. FreeCodeCamp, Estats Units. Recuperat de: <https://www.freecodecamp.org/news/the-mind-blowing-ai-announcement-from-google-that-you-probably-missed-2ffd31334805/#.msj1mdvbh> (Darrera consulta: 28 de maig de 2020)

Flanagan, J. (1972). Speech Analysis, Synthesis, and Perception. Springer-Verlag, Berlin-Heidelberg-New York.

Flanagan, J., Rabiner L. (1973). Speech Synthesis. Dowden, Hutchinson & Ross, Inc., Pennsylvania, Estats Units.

Gorin, A. L., Parker, B. A. et al. (1996) How May I Help You?, Proc. Interactive Voice Technology for Telecommunications Applications (IVTTA), pàg. 57-60

Hof, R. (2014). A Chinese Internet Giant Starts to Dream. MIT Technology Review. Estats Units. Recuperat de: <https://www.technologyreview.com/2014/08/14/171789/a-chinese-internet-giant-starts-to-dream/> (Darrera consulta: 10 de maig del 2020)

Janghee, C. (2018). Mental Models and Home Virtual Assistants (HVAs). Michigan State University, Estats Units. Recuperat de: <https://dl.acm.org/doi/pdf/10.1145/3170427.3180286> (Darrera consulta: 20 de maig de 2020)

Juang, B. H., Rabiner L. R. (2004). Automatic Speech Recognition - A Brief History of the Technology Development. Georgia Institute of Technology, Atlanta, Estats Units. Recuperat de: https://folk.idi.ntnu.no/gamback/teaching/TDT4275/literature/juang_rabiner04.pdf (Darrera consulta: 27 d'abril del 2020)

Klatt D., Klatt L. (1990). Analysis, Synthesis, and Perception of Voice Quality Variations Among Female and Male Listeners. Journal of the Acoustical Society of America, Estats Units. JASA vol. 87 (2): 820-857.

Lee, C.H., Rabiner, L.R et al. (1990). Acoustic modeling for large vocabulary speech recognition, Computer Speech & Language, 4: 1237-1265.

Lee, K. F. (1988) Large-vocabulary speaker-independent continuous speech recognition: The Sphinx system, Tesi de doctorat, Carnegie Mellon University, Estats Units.

Lippmann, R. P. (1990). Review of Neural Networks for Speech Recognition, Readings in Speech Recognition, A. Waibel and K. F. Lee, Editors, Morgan Kaufmann Publishers, pàg. 374-392

Lopez, D. (2019). Your Guide to Natural Language Processing (NLP). Towards Data Science. Recuperat de: <https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1> (Darrera consulta: 11 de maig 2020)

Lowerre, B. (1986). The HARPY Speech Understanding System, Trends in Speech Recognition, W. Lea, Editor, Speech Science Publications, Estats Units.

Moreno, A., Redondo, T. (2016). Text Analytics: the convergence of Big Data and Artificial Intelligence. Universidad Autónoma de Madrid and Instituto de Ingeniería del Conocimiento, Madrid. Recuperat de: <https://dialnet.unirioja.es/servlet/articulo?codigo=5573981> (Darrera consulta: 27 de maig de 2020)

MTS Staff Writer (2019). A One Hour Translation (OHT) Survey Finds out How Good Google Assistant, Amazon's Alexa and Apple's Siri Are in Translating Some of the Most Popular Phrases of All Times. Recuperat de: <https://martechseries.com/content/content-marketing/one-hour-translation-oht-survey-finds-good-google-assistant-amazons-alexa-apples-siri-translating-popular-phrases-times/> (Darrera consulta: 22 de maig de 2020)

O'Brien, H. L., Cairns, P. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Canada-Nova York-Halle. Recuperat de: <https://www.sciencedirect.com/science/article/pii/S1071581918300041> (Darrera consulta: 27 de maig de 2020)

Pyae, A., Scifleet, P. (2018). Investigating differences between native english and non-native english speakers in interacting with a voice user interface: a case of Google Home. Proceedings of the 30th Australian Conference on Computer-Human Interaction. Recuperat de: <https://dl.acm.org/doi/10.1145/3292147.3292236> (Darrera Consulta: 16 de maig de 2020)

Quach, K. (2016). Google's neural network learns to translate languages it hasn't been trained on. The Register, Estats Units. Recuperat de:

https://www.theregister.com/2016/11/17/googles_neural_net_translates_languages_not_trained_on (Darrera consulta: 21 de maig del 2020)

Quoc, V., Schuster, M. (2016). A Neural Network for Machine Translation, at Production Scale. Google Brain Team. Recuperat de: <https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html> (Darrera consulta: 18 de maig de 2020)

Rabiner, L. R., Levinson S. E. et al. (1979). Speaker Independent Recognition of Isolated Words Using Clustering Techniques, IEEE Trans. Acoustics, Speech and Signal Proc., vol. 27, pàg. 336-349.

Robinette, P., Howard, A. M., & Wagner, A. R. (2017). Effect of Robot Performance on Human–Robot Trust in Time-Critical Situations. IEEE Transactions on Human-Machine Systems, 47(4), 425-436.

Samuel, A. L. (1959). Samuel, "Some Studies in Machine Learning Using the Game of Checkers," in IBM Journal of Research and Development, vol. 3, no. 3, pp. 210-229

Shaikh, S., Cruz, I. (2019). 'Alexa, Do You Know Anything?' The Impact of an Intelligent Assistant on Team Interactions and Creative Performance Under Time Scarcity. University of Southern California, Estats Units. Recuperat de: <https://arxiv.org/ftp/arxiv/papers/1912/1912.12914.pdf> (Darrera consulta: 30 de maig de 2020)

Saddler, H. J., Piercy, A. et al. (2020). Intelligent automated assistant. Apple Inc. Recuperat de: <https://patents.google.com/patent/US10553215B2/en> (Darrera consulta: 19 de maig de 2020)

Sokolov, A., Savchenko, A. V. (2019). Voice command recognition in intelligent systems using deep neural networks," IEEE 17th World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herlany, Eslovàquia. Recuperat de: <https://ieeexplore.ieee.org/abstract/document/8782755> (Darrera consulta: 24 de maig de 2020)

Varona A. (1997). Antecedentes y desarrollo de los sistemas actuales de reconocimiento automático del habla. Universidad del País Vasco, Bilbao. Recuperat de: <http://hedatuz.euskomedia.org/6564/1/04321346.pdf> (Darrera consulta: 8 de març del 2020)

Wheatstone, C., (1879). The Scientific Papers of Sir Charles Wheatstone, Londres: Taylor and Francis.

Wilpon, J. G., Roe D. B., (1992) AT&T Telephone Network Applications of Speech Recognition, Proc. COST232 Workshop, Roma.

Wilpon, J. G., Rabiner, L. R. et al. (1990) Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models, IEEE Trans. On Acoustics, Speech and Signal Processing, vol. 38, No. 11, pp. 1870-1878.

Wu, Y., Schuster M. et al (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. Estats Units. Recuperat de: <https://arxiv.org/abs/1609.08144> (Darrera consulta: 25 de maig de 2020)

6. Annexos

Prova 1

Text de partida	Alexa	Google	Suggeriment humà
Invernadero	Greenhouse 3	Greenhouse 3	Greenhouse
Intoxicación	Intoxication 3	Poisoning 3	Poisoning
Teclado	Keyboard 3	Keyboard 3	Keyboard
Inalámbrico	Cordless 3	Wireless 3	Wireless
Paralingüística	For linguistics 1	Paralinguistics 3	Paralinguistics
Metáfora	Metaphore 3	Metaphore 3	Metaphore
Sentencia	Sentence 2	Judgment 3	Judgment
Querella	Quarell 1	Complaint 3	Complaint
Aglutinante	Binder 1	Binder 3	Binder
Caballete	Easel 3	Easel 3	Easel

Text de partida	Alexa	Google	Suggeriment humà
Invernadero	Serre 3	Serre 3	Serre
Intoxicación	Intoxication 3	Empoisonnement 3	Intoxication
Teclado	Clavier 3	Clavier 3	Clavier
Inalámbrico	Sans fil 3	Sans fil 3	Sans fil
Paralingüística	Pour la linguistique 1	Paralinguistique 3	Paralinguistique
Metáfora	Métaphore 3	Métaphore 3	Métaphore
Sentencia	Phrase 2	Jugement 3	Jugement
Querella	Coup d'arrêt 2	Plainte 3	Plainte
Aglutinante	Relier 2	Classeur 3	Liant
Caballete	Chevalet 3	Chevalet 3	Chevalet

Prova 2

Text de partida	Alexa	Google	Suggeriment humà
Células madre	Stem cells 3	Mother cells 1	Stem cells
Ácido desoxirribonucleico	Deoxyribonucleic acid 3	Deoxyribonucleic acid 3	Deoxyribonucleic acid
Comercio electrónico	E-commerce 3	E-commerce 3	E-commerce
Página de inicio	Home page 3	Home page 3	Home page
Sintagma nominal	Nominal sintagma 1	Noun phrase 3	Noun phrase
Análisis paradigmático	Paradigmatic analysis 3	Paradigmatic analysis 3	Paradigmatic analysis
Recurso de apelación	Appeal 3	Appeal 3	Appeal
Norma jurídica	Legal rule 3	Legal rule 3	Precept
Cuento de hadas	Fairytale 3	Fairytale 3	Fairytale
Pintura flamenca	Flemish painting 3	Flemenco painting 1	Flemish painting

Text de partida	Alexa	Google	Suggeriment humà
Células madre	Cellules souche 3	Cellules souche 3	Cellules souche
Ácido desoxirribonucleico	Acide désoxyribonucléique 3	Acide désoxyribonucléique 3	Acide désoxyribonucléique
Comercio electrónico	Commerce électronique 3	E-commerce 3	Commerce électronique
Página de inicio	Page d'accueil 3	Page d'accueil 3	Page d'accueil
Sintagma nominal	Sintagma nominal 1	Phrase nominal 3	Phrase nominal
Análisis paradigmático	Analyse paradigmatique 3	Analyse paradigmatique 3	Analyse paradigmatique
Recurso de apelación	Appel 3	Charm 2	Appel

Norma jurídica	Règle juridique 3	Norme legal 3	Règle de droit
Cuento de hadas	Conte de fées 3	Conte de fées 3	Conte de fées
Pintura flamenca	Peinture flamande 3	Peinture flamenco 1	Peinture flamande

Prova 3

Text de partida	Alexa	Google
Te prometo que queríamos volver pronto	I promise we wanted to come back soon 3	I promise you we wanted to come back soon 3
Te lo iba a decir	I was going to tell you 3	I was going to tell you 3
Puede que aun no lo haya visto	I may not have seen it yet 3	He may not have seen it yet 3
Tengo ganas de volver a la normalidad	I'm looking forward to getting back to normal 3	I want to go back to normal 3
El esfuerzo se valora, pero no es suficiente	Effort is valued, but it is not enough 3	Effort is valued, but it is not enough 3
Se lo voy a decir en cuanto llegue	I'm going to tell you as soon as I get there 3	I will tell you as soon as I arrive 3
No me importa para nada	I don't care at all 3	I do not care at all 3
No puede ser verdad	It can't be true 3	It cannot be true 3
Me encanta pasear por la playa	I love walking on the beach 3	I love to walk on the beach 3
Avísame cuando llegues	Let me know when you get there 3	Let me know when you arrive 3

Text de partida	Alexa	Google
Te prometo que queríamos volver pronto	Je te promets qu'on voulait revenir bientôt 3	Je vous promets que nous voulions revenir bientôt 3
Te lo iba a decir	J'allais te le dire 3	J'allais te dire 3
Puede que aun no lo haya	Je ne l'ai peut-être pas	Vous ne l'avez peut-être

visto	encore vu 3	pas encore vu 3
Tengo ganas de volver a la normalidad	J'ai hâte de revenir à la normale 3	Je veux revenir à la normale 3
El esfuerzo se valora, pero no es suficiente	L'effort est valorisé, mais il ne suffit pas 3	L'effort est apprécié, mais il ne suffit pas 3
Se lo voy a decir en cuanto llegue	Je te le dirai dès que j'y serai 3	Je vous le diré de mon arrivée 3
No me importa para nada	Je m'en fiche 3	Je ne me soucie pas du tout 3
No puede ser verdad	Ce n'est pas vrai 2	CA ne peut pas être vrai 2
Me encanta pasear por la playa	J'adore marcher sur la plage 3	J'adore marcher sur la plage 3
Avísame cuando llegues	Faites-moi savoir quand vous y arrivez 3	Faites le moi savoir quand vous arrivez 3

Prova 4

Text de partida	Alexa	Google
¿Cómo estás?	How are you? 3	How are you? 3
¿Qué hora es?	What time is it? 3	What time is it? 3
¿Dónde está la puerta?	Where's the door? 3	Where is the door? 3
¿Cuántos años tienes?	How old are you? 3	How old are you? 3
¿De dónde eres?	Where are you from? 3	Where are you from? 3
¿Podrías ayudarme?	Could you help me? 3	Could you help me? 3
¿Tienes hambre?	Are you hungry? 3	Are you hungry? 3
¿Estás bien?	Are you ok? 3	You're good? 3
¿Dónde vives?	Where do you live? 3	Where do you live? 3
¿A qué te dedicas?	What do you do? 3	What is your job? 3

Text de partida	Alexa	Google
¿Cómo estás?	Comment vas-tu? 3	Comment ça va? 3

¿Qué hora es?	Quelle heure est-il? 3	Quelle heure il est? 3
¿Dónde está la puerta?	Où est la porte? 3	Où la porte est? 3
¿Cuántos años tienes?	Quel âge as-tu? 3	Quelle âge as-tu? 3
¿De dónde eres?	D'ou venez-vous? 3	D'ou viens tu? 3
¿Podrías ayudarme?	Pourriez-vous m'aider? 3	Pouvez-vous m'aider? 3
¿Tienes hambre?	Tu as faim? 3	Tu as faim? 3
¿Estás bien?	Tu vas bien? 3	Vous êtes dué? 3
¿Dónde vives?	Où habite-tu? 3	Où vis tu? 3
¿A qué te dedicas?	Que fais-tu? 3	Que fait-vous dans la vie? 3

Prova 5

Text de partida	Alexa	Google	Suggeriment humà
El que mucho abarca poco aprieta	The one that encompasses little squeeze 1	Jack of all trades, master of none 3	If you run after two hares you will catch neither
Quien calla otorga	Who's silent grants 1	Silence gives consent 3	Silence gives consent
La cabra siempre tira al monte	The goat always throws into the bush 1	The goat always throws the mountain 1	You can't teach an old dog new tricks
A caballo regalado no le mires el diente	Don't look at his tooth on a gift horse 1	Don't give a tooth to a gift horse 1	Never look a gift horse in the mouth
Matar dos pájaros de un tiro	Killing two birds with one stone 3	Kill two birds with one shot 2	Kill two birds with one stone
Dime con quién andas y te diré quién eres	Tell me who you're hanging out with and I'll tell you who you are 2	Tell me who are your friends and I'll tell you who you are 3	A man is known by the company he keeps
Buscar tres pies al gato	Look for three feet for the cat 1	Find three feet to the cat 1	Flogging a dead horse
No es oro todo lo	It's not gold all that	Not all that glitters is	All that glitters is

que reluce	shines 2	gold 3	not gold
Tirar la casa por la ventana	Throwing the house out the window 1	Throw the house out the window 1	Push the boat out
Llamar al pan pan y al vino vino	Call bread and wine wine 1	Call bread bread and wine wine 1	To call a spade a spade

Text de partida	Alexa	Google	Suggeriment humà
El que mucho abarca poco aprieta	Celui qui englobe peu de compression 1	Jack de tous les métiers maître de rien 3	Qui trop embrasse mal étreint
Quien calla otorga	Qui est des subventions silencieuses 1	Le silence donne son consentement 3	Qui ne dit mot, consent
La cabra siempre tira al monte	La chèvre jette toujours dans la brousse 1	La chèvre jette toujours la montagne 1	L'eau va toujours à la rivière
A caballo regalado no le mires el diente	Ne regarde pas sa dent sur un cheval-cadeau 1	Ne donnez pas de dent à un cheval cadeau 1	A cheval donné, on ne doit en la bouche regarder
Matar dos pájaros de un tiro	Tuer deux oiseaux avec une pierre 2	Tuer deux oiseaux d'un seul coup 1	D'une pierre, deux cops
Dime con quién andas y te diré quién eres	Dis-moi avec qui tu traînes et je te dirai qui tu es 1	Dis-moi qui sont tes amis et je te dirai qui tu es 2	Dis-moi qui tu hantes, et je dirai qui tu es
Buscar tres pies al gato	Recherchez trois pieds pour le chat 1	Trouvez trois pieds au chat 1	Chercher midi à quatorze heures
No es oro todo lo que reluce	Ce n'est pas de l'or tout ce qui brille 1	Tout ce qui brille n'est pas or 1	Ce n'est pas la vache qui beugle le plus fort qui donne le plus de lait
Tirar la casa por la	Jeter la maison par la	Jetez la maison par la	Jeter l'argent par les

ventana	fenêtre 1	fenêtre 1	fenêtres
Llamar al pan pan y al vino vino	Appeler le pain et le vin 1	Appeler pain pain et vin vin 1	Appeler un chat chat

Prova 6

Text de partida	Alexa	Google
Pienso, luego existo	I think, then I exist 2	I think, therefore I exist 2
No hay caminos para la paz, la paz es el camino	There are no paths to peace, peace is the way 2	There are no way for peace, peace is the way 2
Solo sé que no sé nada	I just know I don't know anything 2	I only know that I know nothing 3
Ojo por ojo el mundo acabará ciego	Eye for an eye the world will end up blind 2	An eye for an eye and the world will go blind 3
Es tan corto el amor y tan largo el olvido	It's so short love and so long forgetfulness 1	Love is so short and forgetting so long 3
Mantén cerca a tus amigos, pero más cerca a tus enemigos	Keep your friends close, but closer to your enemies 1	Keep your friends close but your enemies closer 3
No existen preguntas sin respuesta, solo preguntas mal formuladas	There are no unanswered questions, only mis-asked questions 2	There are no unanswered questions, only poorly formulated questions 2
Puede que no sea muy listo, pero sé qué es el amor	I may not be very smart, but I know what love is 2	I may not be very smart, but I know what love is 2
Que la fuerza te acompañe	May the force be with you 3	May the force be with you 3
No quiero necesitarte porque no puedo tenerte	I don't want to need you because I can't have you 3	I don't need to need you because I can't have you 2

Text de partida	Alexa	Google
Pienso, luego existo	Je pense alors que j'existe 2	Je pense, donc se suis 2
No hay caminos para la paz, la paz es el camino	Il n'y a pas de chemin vers la paix, la paix est la voie 2	Il n'y a pas de moyen pour la paix, la paix est le chemin 2
Solo sé que no sé nada	Je sais juste que je ne sais rien 2	Je sais seulement que je ne sais rien 3

Ojo por ojo el mundo acabará ciego	Oeil pour un oeil le monde finira aveugle 2	Un oeil pour un oeil le monde finira aveugle 3
Es tan corto el amor y tan largo el olvido	C'est si peu d'amour et si longtemps d'oubli 2	L'amour est si court et l'oubli si long 3
Mantén cerca a tus amigos, pero más cerca a tus enemigos	Gardez vos amis proches, mais plus près de vos ennemis 2	Gardez vos amis proches, mais vos ennemis plus proches 3
No existen preguntas sin respuesta, solo preguntas mal formuladas	Il n'y a pas de questions sans réponse, seulement des questions mal posées 2	Il n'y a pas de questions sans réponse, seulement des questions mal formulées 2
Puede que no sea muy listo, pero sé qué es el amor	Je ne suis peut-être pas très intelligent, mais je sais ce qu'est l'amour 2	Je ne suis peut-être pas très intelligent, mais je sais ce qu'est l'amour 2
Que la fuerza te acompañe	Que la force soit avec vous 3	Que la force t'accompagne 3
No quiero necesitarte porque no puedo tenerte	Je ne veux pas avoir besoin de toi parce que je ne peux pas t'avoir 1	Je ne veux pas avoir besoin de toi car je ne peux pas t'avoir 2