

4. La traducció automàtica

Índex

4.1. Introducció

En aquest capítol oferim una panoràmica general sobre els sistemes de traducció automàtica orientada a les necessitats dels traductors humans. Durant les passades dècades els traductors sovint han tingut la percepció de que la traducció automàtica era una amenaça per a la seva professió. Es pensava que en un futur proper els sistemes de traducció automàtica aconseguirien un nivell de qualitat suficient com per fer innecessària la intervenció humana en el procés de traducció. Aquesta percepció exageradament optimista no era exclusiva dels professionals de la traducció, sinó que altres sectors de la societat també han cregut que el traductor humà era una espècie en extinció.

Tot i que és cert que la qualitat dels sistemes de traducció automàtica ha millorat moltíssim en els darrers anys, i donat que aquests sistemes estan actualment a l'abast de qualsevol usuari, el públic general és més conscient dels límits d'aquest tipus de sistemes. Aquest millor coneixement dels sistemes de traducció automàtica fa que es valori més la tasca del traductor humà i que es tingui consciència de quins tipus de treballs de traducció requereixen comptar amb un professional de la traducció.

La percepció que tenen els traductors humans també està canviant des d'una amenaça per a la professió cap a una eina de treball que pot ser d'utilitat en algunes situacions (parells de llengües, tipus de text i nivell de qualitat requerit concrets). En aquest sentit, la pràctica totalitat del sistemes de traducció assistida permeten una connexió amb sistemes de traducció automàtica. D'aquesta manera, quan l'eina de traducció assistida no troba cap coincidència amb un índex de similitud suficient a la memòria de traducció, es presenta el resultat de traduir el segment actual amb un sistema de traducció automàtica. El traductor podrà acceptar, editar o rebutjar aquesta proposta automàtica.

Per alguns parells de llengües i tipus de text és habitual treballar amb postedició de textos traduïts automàticament. En algunes situacions aquesta estratègia permet assolir uns nivells de productivitat molt elevats.

Per a què un traductor pugui treballar de manera efectiva amb un sistema de traducció automàtica és important que conegui a fons la tecnologia que hi ha al darrera d'aquests sistemes. D'aquesta manera el traductor podrà decidir en cada situació (parell de llengües, tipus de text i qualitat final requerida) quin ús fa dels sistemes de traducció automàtica. Aquest ús es pot concretar en tres:

- Cap ús, és a dir, no fer servir cap sistema de traducció automàtica.
- Combinació de traducció assistida i traducció automàtica dins d'un sistema de traducció assistida per ordinador.
- Postedició de traducció automàtica, és a dir, traduir automàticament tot el text i posteditar-lo posteriorment.

Aquest capítol està dividir en sis apartats, En primer lloc parlarem de la història de la traducció automàtica. Aquesta història és interessant pel fet que es passa d'un gran optimisme inicial a una situació més realista a partir de l'informe ALPAC, que marca les pautes de la investigació posterior, que a llarg termini va donar els seus fruits. Després veurem les diferents estratègies que es fan servir en els sistemes de traducció automàtica: directa, transferència, interlingua, estadística i basada en exemples, a més dels sistemes híbrids, que combinen diverses de les estratègies anteriors. Per a un professional de la traducció que fa servir un sistema de traducció automàtica és important conèixer l'estratègia en què es basa el sistema que està fent servir. D'aquesta manera podrà revisar amb més atenció aquells aspectes en què els sistema emprat tingui més probabilitats de produir errors.

Seguirà el capítol explicant els diferents usos que poden tenir els sistemes de traducció automàtica, que van des de tenir una idea de què parla un text, fins a produir una traducció amb una qualitat suficient per a ser publicada directament. Després analitzarem els sistemes de traducció automàtica que estan disponibles a Internet i les diferents maneres d'accedir a aquests sistemes. Ens fixarem especialment en les possibilitats d'accés directe i automàtic des d'alguna aplicació de traducció assistida. Veurem també els diferents nivells possibles de combinació de traducció automàtica i assistida, que va des del cas més simple, que consisteix a presentar el resultat de la traducció automàtica només quan no es troba res significatiu a la memòria de traducció; fins a la combinació de segments provinents de la memòria de traducció amb segments traduïts automàticament.

Per últim parlarem de la tasca de postedició de traducció automàtica, en què s'ha de fixar el traductor-posteditor i quines eines i estratègies pot fer servir per a dur a terme aquesta tasca d'una manera més efectiva.

En l'apartat *Per ampliar coneixements* presentarem una lectura complementària per saber més sobre la història de la traducció automàtica. També presentarem Moses, un *toolkit* per a la creació de sistemes de traducció estadístics. Per últim, proporcionarem un parell d'enllaços a reculls de sistemes de traducció automàtica.

4.2. Història de la traducció automàtica

La traducció automàtica és una de les primeres tasques computacionals que es van intentar desenvolupar un cop van estar disponibles els primers ordinadors. Segons assenyala Trujillo (1999) la història de la traducció automàtica ha estat influenciada per la política, la ciència i l'economia dels diferents períodes de la història moderna i distingeix una sèrie de períodes en la seva història. També Hutchins (2007) fa una seqüenciació semblant. En podem distingir doncs, els següents períodes en la història de la traducció automàtica:

- **Precursors o etapa anterior a l'aparició dels ordinadors.** Des del segle XVII diversos científics i filòsofs han proposat representacions del significat independents de la llengua per tal de superar les barreres lingüístiques. Leibniz i Descartes van proposar una sèrie de codis que relacionaven les paraules entre diferents llengües. El 1933 el francès Georges Artsrouni i el rus Petr Smirnov-Troyanski van patentar procediments mecànics per a dur a terme traduccions.

- **Pioners o esforços inicials.** A partir de l'ús dels ordinadors per desxifrar missatges xifrats durant la segona guerra mundial es van començar a fer servir tècniques numèriques per abordar la tasca de traducció automàtica. Andrew Booth i Warren Weawer en 1946 i 1947 van fer els primers intents d'ús dels ordinadors per a la traducció. En aquesta època el terme que es feia servir era el de traducció mecànica (*mechanical* translation). En 1948 Andrew Booth va treballar en l'anàlisi morfològica per a un diccionari mecànic. El 1949 Warren Beawer va posar les bases per al tractament del problema de l'ambigüitat semàntica. A partir d'aquest moment comença la recerca en traducció automàtica tant als Estats Units com en altres països del món. En 1954 es du a terme la primera demostració pública d'un sistema de traducció automàtica rus-anglès. En 1952 es va organitzar la primera conferència sobre traducció automàtica al MIT. A la conferència van sorgir una sèrie d'idees respecte a la preedició i postedició, l'ús de micro-glossaries per evitar els problemes d'ambigüitat, i algun tipus d'anàlisi de l'estructura sintàctica. Es va proposar també l'organització d'una demostració pública per a poder atreure fons per a la investigació en traducció automàtica. D'aquesta manera la primera demostració pública d'un sistema de traducció automàtica es va dur a terme el 1952. La demostració consistia en la traducció de 49 oracions ben seleccionades del rus a l'anglès, fent servir un diccionari molt restringit de només 250 paraules i 6 regles gramaticals. L'interès científic real no va ser gaire notable, però va ser suficientment impressionant per a estimular la gran inversió que va iniciar per la recerca en traducció automàtica, especialment als Estats Units i per a inspirar l'inici d'aquesta recerca en altres països, especialment la URSS.

- **La dècada de les grans expectatives i la desil·lusió (1956-1966).** Durant aquesta dècada van aparèixer molts grups de recerca especialment en els Estats Units la Unió Soviètica, i en general els mètodes de recerca eren una barreja entre les aproximacions empíriques i teòriques. A mitjans dels anys 1960 hi havia grups de recerca en molts països europeus (Hongria, Txecoslovàquia, Bulgària, Bèlgica, Alemanya, França, etc.) i també a Xina, Mèxic i Japó. Molts d'aquests grups van tenir una vida curta, amb l'excepció d'un grup de la Universitat de Grenoble a França. De fet aquest període va ser important per a la recerca en traducció automàtica perquè a permetre fer recerca en àmbits que avui anomenaríem lingüística computacional o enginyeria del llenguatge.

- **L'informe ALPAC (1966)**. L'exagerat optimisme inicial arriba al seu final en els Estats Units amb aquest informe, que afirmava que la traducció automàtica no era efectiva a nivell de costos. S'elimina el finançament públic als Estats Units per a projectes relacionats amb la traducció automàtica i la recerca en aquesta àrea es continua principalment fora d'aquest país. L'informe oferia una sèrie de recomanacions i indicava que la investigació en aquesta àrea s'havia de centrar en:
  - Mètodes pràctics per a avaluar les traduccions
  - Mitjans per accelerar el procés de traducció humana
  - Avaluació de la qualitat i costos de diverses fonts de traduccions

- Investigació sobre la utilització de les traduccions, per evitar la producció de traduccions que després no es llegeixen.
- Estudi de les fonts de retards en el procés de traducció
- Avaluació de la velocitat relativa i costos de diversos tipus de traducció assistida per ordinador
- Adaptació a la traducció dels processos d'edició i producció existents
- Estudi del procés de traducció en la seva globalitat.
- Producció de material de referència per al traductor, incloent-hi l'adaptació de glossaris existents per a la consulta automàtica en sistemes de traducció automàtica

- **Els anys 1970 i els sistemes de traducció automàtica operatius.** En els Estats Units la recerca es centra en la traducció a l'anglès de textos científics russos, en canvi, en el Canadà i a Europa les necessitats eren ben diferents (anglès-francès al Canadà i entre les llengües de la Comunitat Europea). Al Canadà cal destacar el projecte TAUM (*Traduction Automatique de l'Université de Montreal*), que té com a resultat el sistema Météo per traduir previsions del temps, que s'ha estat utilitzant amb èxit des del 1976. Els experiments més innovadors d'aquesta època es centraven en l'estratègia d'interlingua. Cap a mitjans d'aquesta dècada es comença a dubtar d'aquesta aproximació i es pensa que l'aproximació de transferència oferia millors perspectives. Durant aquesta època també s'instal·len les primeres versions de Systran i la Comunitat Europea compra una versió anglès-francès el 1976.

- **El renaixement de principis dels anys 1980.** Cap a finals dels anys 1970 i els principis dels 1980 va créixer l'interès cap a la traducció automàtica. El 1982 s'inicia el projecte Eurotra de la Comunitat Europea, el mateix any s'inicia el projecte Mu al Japó. El principal rival de Systran és Logos, desenvolupat per a traduir manuals d'avions. El 1982 apareix la versió alemany-anglès de Systran i durant els anys 1980 van apareixent altres parells de llengües. Els primers sistemes de traducció per a ordinadors personals són els sistemes American Weidner (1981) i ALPS (1983).

- **Els finals dels 1980 i els principis dels 1990.** A finals dels 1980 apareix el parell alemany-anglès del sistema comercial METAL. Aquest sistema segueix una estratègia de transferència i aviat apareixen altres parells de llengües per a l'holandès, francès, castellà així com d'anglès i alemany. Comencen a aparèixer sistemes dissenyats per a llenguatges especialitzats, com el de la PAHO (Pan American Health Organization). Les grans companyies d'electrònica japoneses, com ara Fujitsu, Hitachi, NEC, Sharp, Toshiba comencen a comercialitzar programes d'ajuda a la traducció, especialment per a japonès-anglès i anglès-japonès. En aquest període també comença a treballar-se en traducció de veu i en la traducció automàtica amb aproximació estadística.

- **Els finals dels 1990 i MAT (*Machine Aided Translation*).** Apareixen sistemes de traducció potents per a ordinadors personals, la traducció automàtica a Internet i es generalitza molt l'ús de memòries de traducció i programes d'ajuda per a traductors. Creix l'interès per la traducció automàtica basada en exemples.

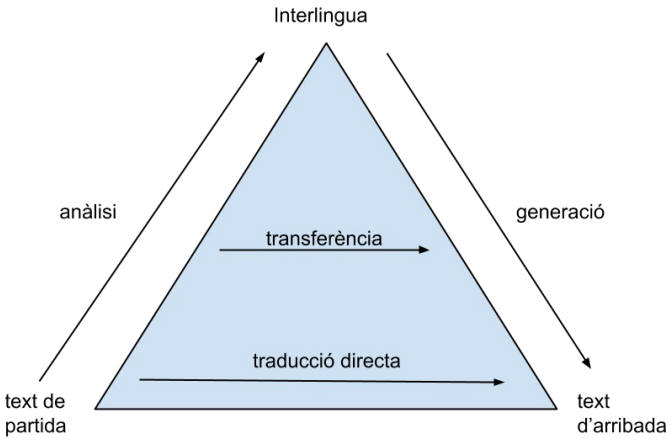
Actualment hi ha disponibles a Internet sistemes de traducció automàtica que ofereixen bona qualitat (com per exemple Google Translate (https://translate.google.com/) o Microsoft Bing Translator (http://www.bing.com/translator/)). Aquests sistemes, a més de poder-se consultar per Internet, ofereixen una API per a una connexió directa amb sistemes de traducció assistida (alguns l'ofereixen de forma gratuïta però amb un volum de text determinat, o bé com a servei de pagament). També es popularitza moltíssim l'ús de sistemes de traducció automàtica entre els traductors i apareixen sistemes TAO complets gratuïts i amb una llicència lliure (per exemple OmegaT – http://www.omegat.org/). També es popularitzen conjunts d'eines que permeten entrenar els teus propis sistemes de traducció automàtica estadística, com per exemple Moses (http://www.statmt.org/moses/).

4.3. Tipus de sistemes de traducció automàtica

Tradicionalment els sistemes de traducció automàtica s'han dividit en tres tipus:

- **Traducció directa**: en aquests sistemes la traducció es du a terme d'una manera directa a partir de la consulta a diccionaris bilingües i amb unes poques regles que permeten fer alguns canvis, com per exemple el canvi de l'ordre de les paraules. Els primers sistemes de traducció automàtica funcionaven d'aquesta manera i encara hi ha sistemes comercials que es basen en aquesta estratègia.
- **Transferència**: aquests sistemes suposen una anàlisi del text d'entrada que dóna com a resultat una estructura que es transfereix a partir d'unes regles a una estructura pròpia de la llengua de sortida. A partir de l'estructura de sortida es genera l'oració en la llengua d'arribada. Normalment aquests sistemes fan una anàlisi sintàctica, que pot ser completa o superficial i per aquest motiu es parla de **sistemes de transferència sintàctica**.
- **Interlingua**: en aquesta estratègia les oracions en la llengua de partida s'analitzen per a obtenir una representació independent de la llengua. A partir d'aquesta representació independent es genera l'oració en la llengua d'arribada.

Per representar les diferències entre aquestes tres estratègies sovint s'ha fet servir l'anomenat *triangle de Vauquois*, que rep el nom del matemàtic i informàtic francès Bernard Vauquois (1929-1985). A la següent imatge podem observar aquest triangle:



En l'eix vertical es representa l'esforç d'anàlisi i generació necessari per a cada estratègia i en l'eix horitzontal l'esforç en la transferència. En l'estratègia de traducció directa l'esforç d'anàlisi i generació és molt baix, ja que l'anàlisi i generació es limiten a la morfologia. En canvi l'esforç de transferència en aquesta estratègia és molt elevat. Per l'estratègia de transferència els esforços en l'anàlisi, generació i transferència són mitjans. A l'extrem, en l'estratègia d'interlingua els esforços d'anàlisi i generació són màxims, mentre que l'esforç en la transferència és nul, ja que la representació és independent de la llengua i coincideix tant pel text de partida com pel text d'arribada.

A banda de les estratègies presentades fins ara i representades en el triangle de Vauquois cal esmentar dues més:

- **Sistemes estadístics**: en aquesta estratègia les traduccions es generen a partir de models estadístics. Els paràmetres d'aquests models estadístics es calculen a partir de corpus textuals bilingües. Aquests sistemes, doncs, poden *aprendre* a traduir a partir de traduccions ja fetes.
- **Sistemes basats en exemples:** aquesta estratègia també fa servir corpus textuals bilingües per deduir com traduir noves oracions. En aquest cas no es generen models estadístics, sinó que l'estratègia es basa en la traducció per analogia. El sistema busca oracions ja traduïdes que puguin servir com a exemples per traduir les noves oracions.

Aquestes són les estratègies principals per a la construcció de sistemes de traducció automàtica. Aquestes estratègies es poden combinar per a produir sistemes **híbrids**. Per exemple, un sistema de transferència sintàctica es pot combinar amb un sistema estadístic. Quan l'anàlisi sintàctica no s'ha pogut produir satisfactòriament perquè la gramàtica no és capaç d'analitzar l'oració, o bé el sistema no disposa de regles de transferència per dur-la a terme, es pot donar pas a solucionar els fragments necessaris mitjançant el model estadístic. Alguns sistemes híbrids funcionen prenent la sortida de diversos sistemes de traducció automàtica i intenten determinar quina de les sortides és de més qualitat per agafar aquesta com la resultant, o bé agafar fragments amb bons índexs de qualitat per combinar-los. En aquests casos, la dificultat rau en saber determinar quina de les sortides o fragments són els de millor qualitat.

4.3.1. Traducció directa

En els sistemes de traducció directa la traducció es basa en la consulta a diccionaris bilingües per a determinar la traducció de les paraules o expressions multiparaula del text de partida. Els processos d'anàlisi i generació s'acostumen a limitar a l'anàlisi morfològica i la lematització. La lematització permetrà fer la cerca més fàcilment al diccionari bilingüe. Un cop es genera l'oració d'arribada s'apliquen una sèrie de regles per poder tractar alguns fenòmens, com per exemple, el canvi d'ordre de les paraules. Podem veure el procés de traducció directa amb el següent exemple:

Volem traduir l'oració catalana:

El nen menja un gelat gran.

L'anàlisi morfològica i lematització podria donar el següent resultat:

El *el* DA0MS0	**nen** *nen* NCMS000	**menja** *menjar* VMIP3S0	**un** *un* DI0MS0	**gelat** *gelat* NCMS000	**gran** *gran* AQ0CS0	**.** *.* Fp

Això ens permetria consultar el diccionari bilingüe (a partir dels lemes) i obtenir les traduccions de cada una de les paraules:

el the	nen boy	menjar eat	un an	gelat ice cream	gran big	. .

Les etiquetes morfosintàctiques de l'anàlisi ens permetran generar la forma correcta del verb *eat*, que en ser 3a persona singular del present d'indicatiu passa a *eats*.

Ara, en el conjunt de regles tenim una que diu:

```
NC* AQ* -> AQ* NC*
```

Que afecta a *icecream big* i que fa que es canviï l'ordre per *big ice cream*, resultant en la frase traduïda

The boy eats a big ice cream.

L'explicació exposada amb aquest exemple és una simplificació. Queden molts aspectes a tractar com per exemple l'ambigüitat, que fa que una paraula del text de partida es pugui traduir per més d'una paraula del text d'arribada, així com el tractament de fenòmens lingüístics més complexos.

Aquesta estratègia és la que empraven els primers sistemes de traducció automàtica i encara es fa sevir en alguns sistemes comercials, com per exemple:

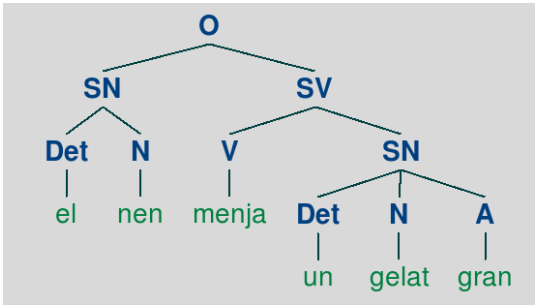
- Systran (http://www.systransoft.com/)
- Logos (http://logos-os.dfki.de/). Aquest sistema s'ha alliberat recentment i ha passat a ser Open Logos.

4.3.2. Sistemes de transferència

En els sistemes de transferència es genera una representació de l'oració de partida, generalment en forma d'anàlisi sintàctica, tot i que també pot incloure informació semàntica. Aquesta representació es transforma en una representació en la llengua d'arribada a partir d'un conjunt de regles. Un cop es té la representació en la llengua d'arribada es fa la transferència lèxica, és a dir, es tradueixen les paraules mitjançant un diccionari. Per últim, es generaran les formes adequades de les paraules en la llengua d'arribada.

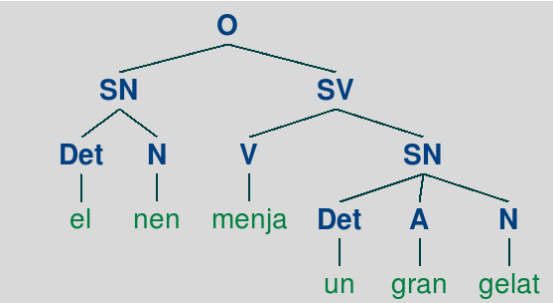
Si continuem amb el mateix exemple de l'apartat anterior, on volem traduir l'oració catalana *El nen menja un gelat gran.* a l'anglès es duran a terme els següents passos (considerem que la transferència és sintàctica):

Anàlisi de l'oració de partida:

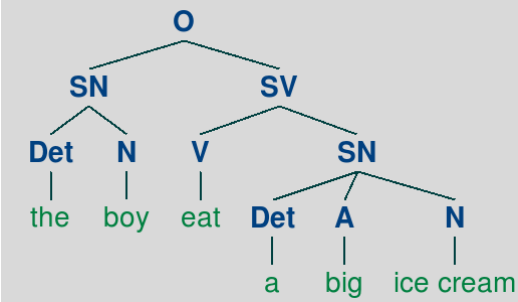


Les regles de transferència es referiran a parts de l'anàlisi i faran per exemple:

I donat que no es pot aplicar cap altra transformació, la representació de l'oració d'arribada queda de la següent manera:



Que un cop es fa la transferència lèxica l'arbre queda:



I que en fer la generació morfològica es posa el verb (*eat*) en tercera persona singular (*eats*) resultant en la frase traduïda:

The boy eats a big ice cream.

De nou això és una simplificació del funcionament d'aquests tipus de sistemes, però expressa els fonaments bàsics del seu funcionament.

4.3.3. Sistemes d'interlingua

En aquesta estratègia el text original es transforma en una representació abstracta, anomenada interlingua, que és independent de la llengua i el text traduït es genera directament a partir d'aquesta generació abstracta. Aquesta estratègia té una sèrie d'avantatges, entre els que es poden destacar:

- Necessita menys component per relacionar cada llengua de partida amb cada llengua d'arribada.
- Es requereixen menys components per tal d'afegir una nova llengua.
- Els components d'anàlisi i generació poden ser desenvolupats amb coneixement d'una sola llengua.
- Es poden desenvolupar sistemes de traducció automàtica per a parells de llengües molt diferents (per exemple anglès i àrab)

I té un inconvenient principal i important:

- La dificultat de definir la representació abstracta independent de la llengua, l'interlingua. La dificultat és encara més gran si es pretén desenvolupar sistemes per a dominis no restringits o amplis.

Veiem ara amb més detall el fet que en els sistemes d'Interlingua es necessiten menys components per desenvolupar sistemes de traducció automàtica per diversos parells de llengües. Posem per cas que volem

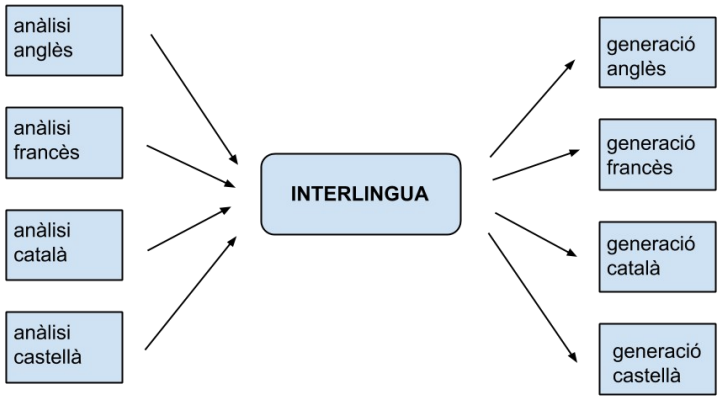
desenvolupar un sistema de traducció automàtica per als següents parells de llengua (i en totes dues direccions): anglès, francès, català i castellà. És a dir, disposem dels següents parells de llengües:

- anglès-francès
- francès-anglès
- anglès-català
- català-anglès
- anglès-castellà
- castellà-anglès
- francès-català
- català-francès
- català-castellà
- castellà-català

En els sistemes interlingua, si considerem que els mòduls d'anàlisi i generació no són reversibles, necessitaríem els següents mòduls:

- Sistema d'anàlisi per a l'anglès
- Sistema de generació per a l'anglès
- Sistema d'anàlisi per al francès
- Sistema generació per al francès
- Sistema d'anàlisi per al català
- Sistema de generació per al català
- Sistema d'anàlisi per al castellà
- Sistema de generació per al castellà

És a dir, necessitem un total de 8 mòduls. Aquests mòduls es poden observar a la següent figura:



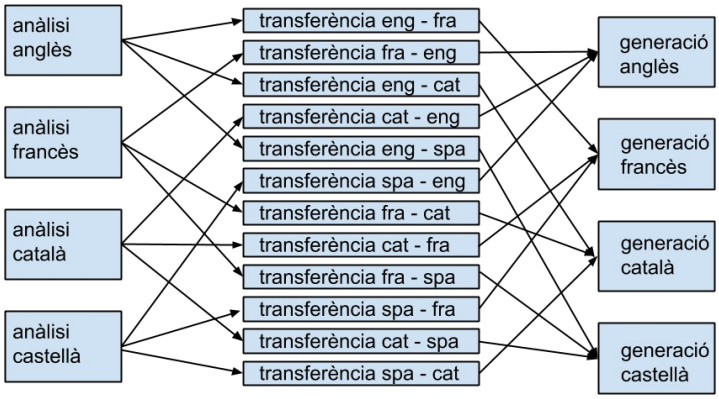
Si volem afegir una nova llengua, per exemple, l'italià, només caldrà afegir dos mòduls més, el mòdul d'anàlisi de l'italià i el mòdul de generació de l'italià i en tindríem un total de 10. En general, en els sistemes d'interlingua, si volem tenir un sistema de traducció automàtica per a *n* llengües necessitarem un total de *2n* mòduls (considerant que els mòduls d'anàlisi i generació no són reversibles).

La situació per als sistemes de transferència és molt diferent. Seguint el mateix exemple, per a disposar del sistema de traducció automàtica necessitaríem:

- Sistema d'anàlisi per al anglès
- Sistema de generació per al anglès
- Sistema d'anàlisi per al francès
- Sistema de generació per al francès
- Sistema d'anàlisi per al català

- Sistema de generació per al català
- Sistema d'anàlisi per al castellà
- Sistema de generació per al castellà
- Sistema de transferència anglès-francès
- Sistema de transferència francès-anglès
- Sistema de transferència anglès-català
- Sistema de transferència català-anglès
- Sistema de transferència anglès-castellà
- Sistema de transferència castellà-anglès
- Sistema de transferència francès-català
- Sistema de transferència català-francès
- Sistema de transferència francès-castellà
- Sistema de transferència castellà-francès
- Sistema de transferència català-castellà
- Sistema de transferència castellà-català

És a dir, per a *n* llengües necessitem un total de *n* sistemes d'anàlisi, *n* de generació i *n(n-1)* de transferència, el que fa un total de *2n+n(n-1)*, és a dir, *n(n+1)* o el que és el mateix, n^2+n mòduls, que en el nostre exemple és de 20 mòduls. Aquesta situació la podem observar a la següent figura:



Si ara volem afegir una llengua més, l'italià per exemple, haurem d'afegir 1 mòdul d'anàlisi, 1 de generació i 8 de transferència (anglès-italià, italià-anglès, francès-italià, italià-francès, català-italià, italià-català, castellà-italià i italià-castellà), passant el sistema a tenir 30 mòduls.

Si la tecnologia que fem servir permet que els mòduls d'anàlisi, generació i transferència siguin reversibles el nombre de mòduls necessaris, tant per a l'interlingua com per a la transferència es redueix notablement, quedant de la següent manera:

Interlingua (anglès, francès, català i castellà):

- Sistema d'anàlisi i generació per a l'anglès
- Sistema d'anàlisi i generació per al francès
- Sistema d'anàlisi i generació per al català
- Sistema d'anàlisi i generació per al castellà

És a dir, un total de 4 mòduls, és a dir, d'*n* mòduls. Si volem afegir una nova llengua, només serà necessari afegir un nou mòdul.

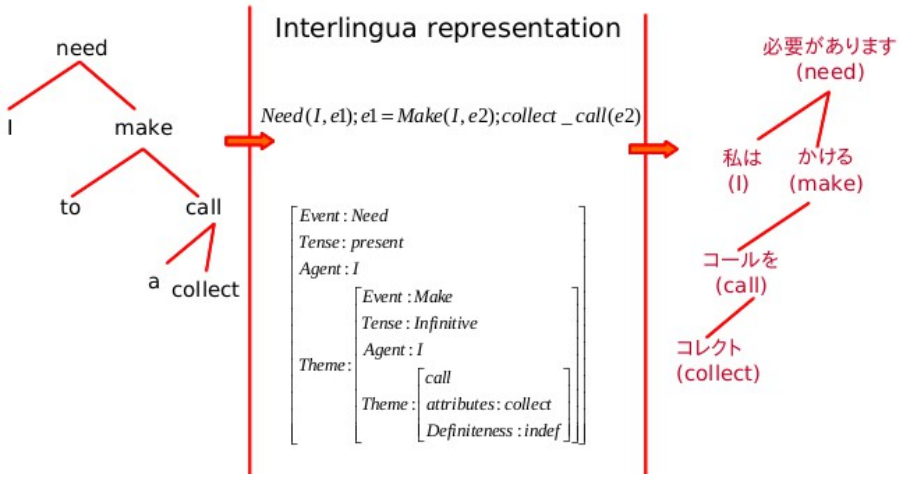
Transferència (anglès, francès, català i castellà):

- Sistema d'anàlisi i generació per a l'anglès
- Sistema d'anàlisi i generació per al francès
- Sistema d'anàlisi i generació per al català
- Sistema d'anàlisi i generació per al castellà
- Sistema de transferència anglès ↔francès
- Sistema de transferència anglès ↔català
- Sistema de transferència anglès ↔castellà
- Sistema de transferència francès ↔català
- Sistema de transferència francès ↔castellà
- Sistema de transferència català ↔castellà

És a dir, un total de 10 mòduls (4 d'anàlisi-generació i 6 de transferència). Si ara afegim una nova llengua s'afegiria un nou sistema d'anàlisi-generació i 4 de transferència, fent un total de 15. És a dir, necessitem n sistemes d'anàlisi-generació i $\frac{n(n-1)}{2}$ sistemes de transferència, o el que és el mateix, un total de $\frac{(n^2+n)}{2}$ mòduls.

Tot i aquest avantatge respecte al nombre de mòduls necessaris, els sistemes Interlingua s'enfronten a la gran dificultat que suposa el disseny d'aquesta representació intermèdia independent de la llengua, és a dir, la definició de la interlingua. A més, és molt més complex dissenyar els sistemes d'anàlisi cap a aquesta interlingua i els de generació des de l'interlingua fins a la llengua d'arribada.

És difícil trobar sistemes interlingua operatius i disponibles de forma oberta i per tant no podem mostrar un exemple real de traducció amb aquests tipus de sistemes. Tot i això, presentarem un exemple extret dels materials docents dels professors Dr. Srinivas Bangalore i Juan Carlos Niebles[1]



Si volem traduir la frase anglesa *I need to make a collect call* (Necessito fer una trucada a cobrament a destinació) al japonès, el primer que es fa és analitzar l'oració anglesa per arribar a la representació interlingua que veiem en la part del mig de la imatge. A partir d'aquesta representació es podrà general l'oració japonesa.

1 Corresponents al curs *COS401/TRA301 - Introduction to Machine Translation* de la Universitat de Princeton (http://www.cs.princeton.edu/courses/archive/spring09/cos401/slides/Interlingua-based_MT.ppt)

Alguns sistemes d'interlingua, com el descrit en Schubert (1988) fan servir l'esperanto com a representació intermèdia, ja que és una llengua amb una sintaxi extremadament regular i una semàntica molt clara. Cal no confondre aquesta estratègia de fer servir una altra llengua com a interlingua, amb el fet de fer servir una llengua com a *pivot* o *llengua pont*. Les llengües pivot es poden fer servir en qualsevol de les estratègies de traducció automàtica i consisteix a traduir d'una llengua A a una B a través d'una llengua P. L'avantatge de fer servir llengües pivot és que es requereixen la construcció de menys mòduls, però per contra en cada retraducció es van introduint errors que perjudiquen a la qualitat final.

4.3.4. Sistemes estadístics

La *traducció automàtica estadística* (SMT-*Statistical Machine Translation*) és una estratègia de traducció automàtica que es basa en l'ús de models estadístics amb models obtinguts a partir de l'anàlisi de corpus bilingües. És a dir, el sistema és capaç de traduir a partir d'una sèrie de paràmetres que han estat calculats a partir d'una gran quantitat de textos originals i les seves traduccions.

Tot i que les primeres idees daten de 1949, aquesta estratègia no pren embranzida fins a principis dels anys 1990.

En general, els sistemes de traducció automàtica estadística treballen amb dues probabilitats:

- La probabilitat de què una paraula o conjunt de paraules de la llengua de partida (SL-*source language*) es tradueixi per una paraula o conjunt de paraules de la llengua d'arribada (TL-*target language*).
- La probabilitat de què una cadena en la llengua d'arribada sigui una oració vàlida en aquesta llengua.

Veurem la notació que es fa servir habitualment, que està influenciada pels primers exemples que consideraven un sistema de traducció automàtica entre el francès i l'anglès. El que ens interessa és calcular la probabilitat de que una frase en la llengua d'arribada (*e*: ja que es considera l'anglès com a llengua d'arribada) sigui la traducció d'una frase en la llengua de partida (*f*: ja que es considera el francès com a llengua de partida), això s'escriu com $p(e|f)$. Aquesta mateixa probabilitat es pot calcular com:

$$p(e|f) = p(f|e)p(e)$$

On:

- $p(f|e)$ és el *model de traducció* que expressa la probabilitat de què la cadena en la llengua de partida sigui la traducció de la cadena en la llengua d'arribada.
- $p(e)$ és el *model de llengua* i expressa la probabilitat de què la cadena en la llengua d'arribada sigui una cadena vàlida en aquesta llengua.

Aquesta descomposició és interessant perquè divideix el problema en dos subproblemes. La quantitat de possibles traduccions que es poden deduir dels paràmetres del sistema estadístic és enorme, i el que caldrà serà trobar la traducció que tingui una probabilitat més gran, és a dir:

$$\arg\max p(e|f) = \arg \max p(f|e)p(e)$$

La cerca de la millor traducció, és a dir, la que maximitza aquesta probabilitat no es pot fer calculant la probabilitat de tots els possibles candidats de traducció, ja que la quantitat de candidats és enorme i el càlcul prendria molta estona. Per aquest motiu, la cerca del millor candidat es fa a partir de diverses tècniques heurístiques per reduir l'espai de cerca però mantenint al mateix temps un nivell de qualitat acceptable. El mòdul encarregat de fer aquesta cerca és el *descodificador* i assegura que la traducció triada tingui una alta

probabilitat, però no pot assegurar que sigui la més probable de totes.

Els sistemes de traducció automàtica estadístics tenen una sèrie d'avantatges, entre els que es poden destacar:

- No és necessari el desenvolupament de regles lingüístiques, que en general requereixen un gran esforç humà, cosa que encareix notablement el desenvolupament.
- Els sistemes de traducció automàtica estadística poden entrenar-se i fer-se servir per a un gran ventall de llengües, ja que no contenen mòduls específics per a un determinat parell de llengües.
- Hi ha disponibles recursos lingüístics en format electrònic per a moltes llengües (no per a totes, però) i molts d'aquests recursos són multilingües (almenys entre alguns parells de llengües). Hi ha disponibles algorismes eficients per a l'alineament automàtic de textos i les seves traduccions

I com a inconvenients:

- El darrer avantatge que hem comentat és un inconvenient per a certs parells de llengües. Quan no estan disponibles els recursos necessaris per a entrenar un sistema estadístic per a un determinat parell de llengües, l'elaboració d'aquests recursos pot ser tan costosa com la generació de sistemes de traducció basats en regles.
- En certa manera els resultats són inesperats, podent obtenir tant traduccions de molt bona qualitat com traduccions realment decebedores.
- Aquests sistemes de traducció no funcionen bé per a llengües que tinguin un ordre de paraules molt diferent.

Els sistemes de traducció automàtica estadística es poden classificar en els següents tipus principals:

- Traducció a partir de paraules
- Traducció a partir de frases
- Traducció a partir de sintaxi
- Models de traducció factoritzats

Veiem breument cada un d'aquests tipus.

Traducció a partir de paraules *(word-based translation*)

Aquest és el model més senzill i es basa únicament en la traducció lèxica, és a dir, en la traducció de paraules de manera aïllada. Per a dur a terme la traducció amb aquesta estratègia necessitem un diccionari bilingüe entre la llengua de partida i la llengua d'arribada. Imaginem-nos que volem traduir la mateixa frase que hem fet servir en altres exemples:

El nen menja un gelat gran.

En un diccionari convencional tindríem la següent informació sobre la traducció de la paraula *nen*.

nen: child, boy

Per al nostre sistema de traducció el que necessitem és un diccionari que a més de les traduccions, aporti la probabilitat associada a cada una de les traduccions. Aquesta probabilitat es pot calcular a partir d'un corpus paral·lel. Si en el nostre corpus paral·lel la paraula *nen* apareix 160 vegades en la part catalana, i apareix traduïda com a *boy* 100 vegades i com a *child* 60 vegades podem calcular les probabilitats de la següent manera:

Paraula català	Freqüència	Traducció anglesa	Freqüència	Probabilitat
nen	160	boy	100	100/160=0,625
nen	160	child	60	60/160=0,375

Per poder traduir la frase de l'exemple, necessitaríem disposar del diccionari probabilístic per a cada una de les paraules.

El	nen	menja	un	gelat	gran
the – 0,8 a – 0,1 this – 0,1	boy – 0,625 child – 0,375	eats - 0,8 consumes – 0,1 ingests – 0,1	a - 0,3 an – 0,2 one – 0,3 the – 0,1 this – 0,1	ice_cream - 0,6 frozen – 0,3 ice – 0,1	big - 0,4 large – 0,4 old – 0,2

El nombre de possibles traduccions segons aquesta taula és de 3x2x3x5x3x3=810 i per cada una d'elles podem calcular una probabilitat (veiem alguns exemples):

The boy eats a ice_cream big	0,8 x 0,625 x 0,8 x 0,3 x 0,6 x 0,4 = 0,0288
The boy eats a ice_cream large	0,8 x 0,625 x 0,8 x 0,3 x 0,6 x 0,4 = 0,0288
This child ingests this ice old	0,1 x 0,375 x 0,1 x 0,1 x 0,1 x 0,2 = 0,0000075

Com veiem el sistema assigna una probabilitat més alta a les traduccions que semblen més correctes (els dos primers exemples tenen la mateixa probabilitat). Ara bé, queden alguns aspectes per resoldre. El primer que veurem serà els possibles canvis d'ordre de les paraules. Els models que s'apliquen permeten certs canvis en els ordres de les paraules, d'aquesta manera s'acabarien avaluant moltes més opcions, com per exemple:

Boy the eats a ice_cream big
The boy eats a big ice_cream
The boy eats a large ice_cream

Entre moltíssimes d'altres. Ara, a partir del model de llengua d'arribada, el que indica si una oració és probable en la llengua d'arribada. Podem simular aquest model fent consultes a la interfície del British National Corpus (http://www.natcorp.ox.ac.uk/) i obtindriem els següents resultats:

Combinació	Freqüència al corpus
the boy	3296
boy the	27
big ice_cream	2
large ice_cream	0

Combinant les diferents probabilitats (els del model de traducció i del model de llengua) el descodificador probablement triaria com a traducció més probable:

The boy eats a big ice cream.

Altres problemes que cal solucionar és el tema de la *fertilitat*, que es dóna perquè en una determinada llengua

una paraula pot ser traduïda per més d'una paraula en una altra llengua. Quan una paraula de la llengua de partida es tradueix per més d'una paraula en la llengua d'arribada, això no suposa cap problema, ja que es soluciona a nivell de diccionari (observeu gelat – ice cream). Quan passa al revés, la cosa canvia i és més complicat. Diverses paraules en la llengua de partida poden traduir-se per una única paraula en la d'arribada, i fins i tot algunes paraules de l'original no apareixen a la traducció. La solució passa per introduir paraules nul·les a la traducció.

L'estratègia de traduir per paraules pràcticament no es fa servir avui dia i es fan servir molt més la traducció a partir de frases, que veurem a continuació.

Traducció a partir de frases (*phrase-based translation*)

Aquesta estratègia de traducció automàtica estadística és la que es fa servir més actualment i la que obté millors resultats i es basa en la traducció de petites seqüències de paraules. Aquestes seqüències no tenen perquè estar motivades lingüísticament, sinó que són seqüències arbitràries. Per explicar aquesta estratègia farem servir la mateixa frase de l'exemple anterior, la traducció de la frase catalana *El nen menja un gelat gran* a l'anglès. Ara la nostra taula de traducció (la mateixa funció que el diccionari bilingüe en el cas de traducció per paraules) tindria un aspecte com el següent (l'expressarem en el format Moses, un sistema de traducció automàtica estadística que explicarem breument més endavant). Fixem-nos que la taula està construïda per combinació de 3 paraules, 2 paraules i 1 paraula. És un model de trigrams (on l'ordre màxim dels n-grames, o combinacions de paraules, és 3).

```
el nen menja ||| the boy eats ||| 0.60 ||| |||
el nen menja ||| this boy eats ||| 0.1 ||| |||
el nen menja ||| the child eats ||| 0.3 ||| |||
nen menja un ||| boy eats a ||| 0.6 ||| |||
nen menja un ||| boy eats an ||| 0.4 ||| |||
menja un gelat ||| eats an ice cream ||| 1 ||| |||
un gelat gran ||| a big ice cream ||| 0.75 ||| |||
un gelat gran ||| a large ice cream ||| 0.25 ||| |||
el nen ||| the boy ||| 0.5 ||| |||
el nen ||| the child ||| 0.3 ||| |||
el nen ||| a boy ||| 0.1 ||| |||
el nen ||| a child ||| 0.1 ||| |||
nen menja ||| boy eats ||| 0.6 ||| |||
nen menja ||| child eats ||| 0.4 ||| |||
menja un ||| eats a ||| 0.5 ||| |||
menja un ||| eats an ||| 0.4 ||| |||
menja un ||| eats the ||| 0.1 ||| |||
un gelat ||| an ice cream ||| 0.85 ||| |||
un gelat ||| one ice cream ||| 0.15 ||| |||
gelat gran ||| big ice cream ||| 0.5 ||| |||
gelat gran ||| large ice cream ||| 0.3 ||| |||
gelat gran ||| great ice cream ||| 0.2 ||| |||
el ||| the ||| 0.8 ||| |||
el ||| a ||| 0.1 ||| |||
el ||| this ||| 0.1 ||| |||
nen ||| boy ||| 0.625 ||| |||
nen ||| child ||| 0.375 ||| |||
menja ||| eats ||| 0.9 ||| |||
menja ||| consumes ||| 0.1 ||| |||
menja ||| ingests ||| 0.1 ||| |||
un ||| a ||| 0.3 ||| |||
un ||| an ||| 0.2 ||| |||
un ||| one ||| 0.3 ||| |||
un ||| the ||| 0.1 ||| |||
un ||| this ||| 0.1 ||| |||
gelat ||| ice cream ||| 0.6 ||| |||
gelat ||| frozen ||| 0.3 ||| |||
gelat ||| ice ||| 0.1 ||| |||
gran ||| big ||| 0.4 ||| |||
gran ||| large ||| 0.4 ||| |||
gran ||| old ||| 0.2 ||| |||
```

A part d'aquest component, també necessitem un model de la llengua d'arribada (en aquest cas l'anglès) que

ens doni la probabilitat de què una frase sigui una frase correcta de la llengua d'arribada. Suposarem que disposem d'un corpus de l'anglès que està format per 9 oracions i 79 paraules (evidentment, en els casos reals els corpus contenen centenars de milers d'oracions i milions de paraules). Veiem el corpus:

```
the baby was a boy
she made the boy brush his teeth every night
he remained a child in practical matters as long as he lived
what did you eat for dinner last night?
I didn't eat yet, so I gladly accept your invitation
this dog doesn't eat certain kinds of meat
the girl eats an ice cream
my father eats a big ice cream
we also saw abundance of large whales
had a great time at the party
```

Calculem el model de llenguatge amb algunes de les eines disponibles i obtenim un fitxer que té el següent aspecte (el format final pot dependre de l'eina emprada; mostrem únicament un fragment):

```
\1-grams:
-1.614897 the -0.011136
-2.012837 baby -0.011136
-1.836746 boy -0.011136
-2.012837 child -0.011136
-2.012837 an -0.011136
-1.836746 ice -0.312166
-1.836746 cream -0.011136
-2.012837 my -0.011136
-2.012837 father -0.011136
-2.012837 big -0.011136
-2.012837 we -0.011136
-2.012837 also -0.011136
-2.012837 saw -0.011136
-2.012837 abundance -0.011136
-2.012837 large -0.011136
-2.012837 whales -0.011136
-2.012837 had -0.011136
-2.012837 great -0.011136
-2.012837 time -0.011136
-2.012837 at -0.011136
-2.012837 party 0.000000
....
-0.507687 <unk>
\2-grams:
-2.198657 the baby 0.000000
-2.198657 the boy 0.000000
-2.198657 the girl 0.000000
-2.198657 the party 0.000000
-1.596597 baby was 0.000000
-1.596597 was a 0.000000
-2.198657 a boy 0.000000
-2.198657 a child 0.000000
-2.198657 a big 0.000000
-2.198657 a great 0.000000
....
```

Amb tota aquesta informació podem calcular les traduccions més probables de la nostra frase d'exemple (*el nen menja un gelat gran*). En aquest punt és important establir quins pesos donem a cada un dels models (el de traducció i el de llengua). Si donem un pes de 3 al model de traducció i un de 1 al de llengua obtenim la traducció satisfactòria *the boy eats a big ice cream* en primera posició. El programa també ens pot donar una llista de traduccions endreçades per probabilitat i indicant els valors de probabilitat segons el model de llengua i el model de traducció. Veiem aquesta informació per a les 10 millors traduccions:

```
0 ||| the boy eats a big ice cream  ||| LM= -24.4035 Distortion0= 0 WordPenalty0= -7
PhraseDictionaryMemory0= -0.798508 ||| -26.799
0 ||| the boy eats a big ice cream  ||| LM= -24.4035 Distortion0= 0 WordPenalty0= -7
PhraseDictionaryMemory0= -1.02165 ||| -27.4685
0 ||| the boy eats a big ice cream  ||| LM= -24.4035 Distortion0= 0 WordPenalty0= -7
PhraseDictionaryMemory0= -1.08619 ||| -27.6621
0 ||| the boy eats a big ice cream  ||| LM= -24.4035 Distortion0= 0 WordPenalty0= -7
PhraseDictionaryMemory0= -1.08619 ||| -27.6621
0 ||| the boy eats a big ice cream  ||| LM= -24.4035 Distortion0= 0 WordPenalty0= -7
PhraseDictionaryMemory0= -1.42712 ||| -28.6849
0 ||| the boy eats an ice cream old  ||| LM= -21.92 Distortion0= 0 WordPenalty0= -7
PhraseDictionaryMemory0= -2.28278 ||| -28.7683
0 ||| the boy eats an ice cream old  ||| LM= -21.92 Distortion0= 0 WordPenalty0= -7
PhraseDictionaryMemory0= -2.30259 ||| -28.8277
0 ||| the boy eats an ice cream old  ||| LM= -21.92 Distortion0= 0 WordPenalty0= -7
PhraseDictionaryMemory0= -2.30259 ||| -28.8277
0 ||| the boy eats a frozen old  ||| LM= -18.7077 Distortion0= 0 WordPenalty0= -6
PhraseDictionaryMemory0= -3.54738 ||| -29.3498
0 ||| the boy eats an ice cream old  ||| LM= -21.92 Distortion0= 0 WordPenalty0= -7
PhraseDictionaryMemory0= -2.50593 ||| -29.4377
```

Els pesos dels diferents models, i altres paràmetres del sistema es poden calcular automàticament. Aquest procés es coneix com a *ajust* (*tuning*). Una part del corpus paral·lel disponible es reserva per a aquesta etapa (un petit fragment és suficient). Llavors, un cop calculats els models, el sistema tradueix el corpus diverses vegades amb diferents combinacions dels diferents paràmetres. Com que es tracta d'un corpus paral·lel i la traducció és coneguda, el sistema podrà determinar quina és la millor combinació de paràmetres. Aquesta combinació de paràmetres es farà servir en el sistema de traducció automàtica estadística resultant.

Traducció a partir de sintaxi (*syntax-based translation*)

La traducció a partir de sintaxi es basa en la traducció d'unitats sintàctiques en comptes de paraules aïllades o de cadenes de paraules. Per a poder dur a terme aquesta estratègia és indispensable disposar de *parsers* potents capaços de dur a terme l'anàlisi sintàctica, tant en el moment de l'entrenament del sistema com en el moment de dur a terme la traducció.

Models de traducció factoritzats (*factored translation models*)

En aquests models de traducció es treballa amb informació morfològica, sintàctica o semàntica integrant-les en diferents nivells. La motivació per treballar d'aquesta manera és doble:

- Els models de traducció construïts a partir de representacions més generals, com per exemple a partir de lemes en lloc de formes poden basar-se en estadístiques més riques i evitar els problemes derivats de l'escassetat de dades (*data sparseness*) deguts a les limitacions de mida de les dades d'entrenament.
- Molts aspectes de la traducció es poden explicar millor en el nivell morfològic, sintàctic o semàntic. Per exemple, la concordança es pot modelar millor a partir d'informació morfològica i la reordenació de paraules a partir de principis sintàctics.

Els models factoritzats es basen en els models a partir de frases però on els tokens ja no són simplement la forma de la paraula, sinó un vector que conté informació sobre la forma, el lema, la categoria gramatical, morfologia, semàntica, etc.

Per exemple, un sistema que treballés amb formes, lemes, categoria gramatical i informació morfològica funcionaria de la següent manera:

- Traduiria els lemes de la llengua de partida en lemes de la llengua d'arribada
- Traduiria els factors morfològics i de categoria gramatical
- Generaria les formes corresponents a partir dels lemes i la informació morfològica i de categoria gramatical

4.3.5. Sistemes de traducció automàtica basada en exemples

Els sistemes de traducció automàtica basada en exemples (EBMT – *Example Based Machine Translation*) es basen a explotar traduccions anteriors similars per a poder fer la traducció d'una nova oració. En certa manera es pot veure com un cas extrem d'ús de memòries de traducció. En el cas de les eines de traducció assistida el sistema busca traduccions d'oracions similars a la que estem traduint i deixa la tasca de generar la nova traducció al traductor humà. En el cas dels sistemes de traducció automàtica basada en exemples també se cerquen traduccions d'oracions similars a la que es vol traduir a unes memòries de traducció, però la composició de la nova traducció la fa totalment l'ordinador.

Veiem el següent exemple entre l'anglès i el japonès[2]

Anglès	Japonès
How much is that **red umbrella**?	Ano **akai kasa** wa ikura desu ka.
How much is that **small camera**?	Ano **chiisai kamera** wa ikura desu ka.

A partir d'aquests exemples el sistema podria aprendre que:

How much is that **X** ?	Ano **X** wa ikura desu ka.

Aquesta metodologia de traducció va ser proposada en primer lloc per Nagao (1984) on presentava un sistema de traducció entre el japonès i l'anglès. La traducció automàtica basada en exemples és especialment atractiva per a parells de llengües força allunyats, on la resta de metodologies de traducció automàtica presenten problemes difícils de resoldre.

El funcionament general d'aquests sistemes es pot explicar en els següents passos:

- Segmentació de l'oració d'entrada en fragments. Aquests fragments per regla general estan motivats lingüísticament, a diferència de la traducció automàtica estadística basada en frases, on els fragments són arbitraris.
- Se cerquen els exemples adequats per traduir aquests fragments.
- Es compon l'oració completa traduïda a partir dels fragments.

4.3.6. Sistemes híbrids

En els apartats anteriors hem vist diverses metodologies per a la traducció automàtica. Cada una d'aquestes metodologies té els seus avantatges i inconvenients i són capaces de solucionar alguns fenòmens lingüístics de manera satisfactòria i en canvi tenen problemes per solucionar altres aspectes del llenguatge. Aquests fets fan pensar que combinar d'alguna manera dos o més sistemes d'estratègies diferents poden fer millorar la qualitat final de la traducció. La manera de combinar els sistemes poden ser diferents i es poden distingir tres estratègies principals[3]:

- Combinació de diversos sistemes de traducció automàtica (*multi-engine approach*)
- Generació estadística de regles

2 Exemple extret de la Vikipèdia: http://en.wikipedia.org/wiki/Example-based_machine_translation
3 Font: http://en.wikipedia.org/wiki/Hybrid_machine_translation

- Multi-etapes (*multi-pass approach*)

Combinació de diversos sistemes de traducció automàtica (*multi-engine approach*)
En aquesta estratègia la traducció es du a terme en paral·lel amb diversos sistemes de traducció automàtica. La sortida final es genera a partir de les sortides dels diferents sistemes. En alguns casos s'apliquen tècniques estadístiques per determinar quina de las sortides és la millor i prenent la millor com a sortida del sistema híbrid. En sistemes més avançats la sortida final es genera mitjançant la combinació de les diferents sortides. Aquesta estratègia es fa servir habitualment per combinar sistemes basats en regles i sistemes estadístics, tot i que també s'han explorat altres combinacions.

Generació estadística de regles
Aquesta estratègia implica l'ús de dades estadístiques per a generar regles tant lèxiques com sintàctiques. Un cop generades aquestes regles el sistema funciona com a un sistema basat en regles. L'objectiu d'aquesta estratègia és evitar haver de generar regles lingüístiques de manera manual, ja que és una tasca que requereix un gran esforç humà. El sistema intenta generar les regles de manera automàtica a partir del corpus d'entrenament. Aquesta estratègia ha resultat especialment útil per a dominis restringits.

Multi-etapes (*multi-pass approach*)
Aquesta estratègia suposa el processament del text d'entrada diverses vegades en sèrie, és a dir, l'entrada es processa per un sistema i la sortida d'aquest primer sistema es processa per un altre sistema, i així successivament per tots els sistemes que conformen el sistema, fins a produir la sortida final. El cas més habitual de sistema multi-etapa es el processament del text d'entrada amb un sistema de traducció automàtica basat en regles i el processament posterior d'aquesta sortida amb un sistema de traducció automàtica estadístic que produeix la sortida final. L'avantatge d'aquesta aproximació és que es redueix la quantitat d'informació que necessita processar el sistema estadístic i el fet que el sistema basat en regles no necessiti ser un sistema de traducció complet, reduint d'aquesta manera l'esforç humà necessari para construir el sistema.

4.4. Usos de la traducció automàtica

4.4.a. Usos tradicionals de la traducció automàtica

Tradicionalment s'han distingit dos usos principals de la traducció automàtica (Hutchins, 2009): la *disseminació* i l'*assimilació*:

- **Assimilació**: l'ús de sistemes de traducció automàtica per a produir traduccions amb l'objectiu de tenir una idea de sobre què parla un text.
- **Disseminació**: l'ús de sistemes de traducció automàtica per a produir traduccions de prou qualitat per a ser publicades. En la majoria dels casos, el resultat de la traducció automàtica ha de ser revisat o posteditat.

Pel que fa a la disseminació, per a produir un text de qualitat suficient per a ser publicat pràcticament sempre es necessari una revisió o postedició per part de traductors humans. En molts casos, aquesta revisió és tant important que la traducció automàtica es pot considerar simplement com a un esborrany de la traducció. Per evitar o almenys reduir l'esforç de revisió i postedició es fan servir *llenguatges controlats*, és a dir, es parteix de textos d'entrada que s'han produït a partir d'unes normes que restringeixen la gramàtica i el vocabulari amb l'objectiu de disminuir la complexitat i eliminar l'ambigüitat. Partint d'aquests llenguatges controlats es poden desenvolupar sistemes de traducció automàtica que proporcionin una sortida de prou qualitat. Aquesta pràctica es fa servir en empreses que produeixen productes que requereixen molta documentació i que tinguin la necessitat de produir la documentació en diverses llengües.

Pel que fa a l'assimilació el sistema de traducció automàtica ha de ser capaç de produir una sortida que permeti fer-se una idea del contingut del text i entendre els aspectes fonamentals. Aquest ús és habitual en Internet, quan un usuari visita una pàgina en un idioma que no coneix i vols saber les idees principals del seu contingut. Aquest ús també ha estat propiciat per agències governamentals d'intel·ligència per poder filtrar certs textos sospitosos de contenir informació important. Davant un text en una llengua totalment desconeguda, el sistema de traducció automàtica ens permetrà saber si parla de terrorisme o altres temes sensibles, o si simplement és una crònica d'un partit de futbol. En cas de confirmar-se que es tracta d'un document sensible, aquest podrà ser traduït posteriorment per un traductor humà.

A aquests usos tradicionals actualment s'afegeixen molts altres, que resumirem a continuació seguint l'article de Hutchins (2009).

4.4.b. La traducció automàtica com a ajuda al traductor humà

En un primer moment, la traducció automàtica es veia com a una veritable amenaça per a la professió de traductor i no es considerava que aquesta tecnologia pogués resultar d'ajut al traductor. Un cop s'ha pres consciència de les limitacions de la traducció automàtica i a la vegada els sistemes han millorat molt la qualitat de sortida, es considera ja que la traducció automàtica pot ser una ajuda a la professió. Una mica més endavant en aquest capítol veurem la possibilitat de combinar traducció automàtica i assistida dins d'una eina de traducció assistida per ordinador. Aquesta combinació pot ser molt productiva per a alguns parells de llengües i tipus de textos concrets. També veurem més endavant el tema de la postedició de traducció automàtica, que obre la porta a noves possibilitats professionals. De nou, per a certs parells de llengües i tipus de textos hi ha sistemes de traducció automàtica capaços de proporcionar una qualitat suficient perquè la revisió i correcció d'aquests textos (la seva postedició) fins a assolir la qualitat final necessària sigui més productiva que la traducció humana.

4.4.c. La traducció automàtica a Internet

A aquesta qüestió també li dedicarem un apartat sencer una mica més endavant. A Internet hi ha disponibles una sèrie de sistemes de traducció automàtica que ofereixen bona qualitat i que en la majoria de casos permeten un accés gratuït. Aquest fet ha popularitzat l'ús de sistemes de traducció automàtica entre usuaris generals. Els usos habituals passen per la navegació en planes escrites en llengües no prou conegudes per l'usuari i en la traducció de petits textos per a ús personal. Aquesta popularització dels sistemes de traducció automàtica ha fet que la societat en general sigui més conscient de la qualitat que es pot assolir i per a quins usos aquesta qualitat és suficient i també per a quins usos és imprescindible la participació d'un traductor humà.

4.4.c. La traducció automàtica i els dispositius mòbils

La disponibilitat actual de dispositius mòbils potents i amb connexió a Internet ha fet aparèixer tot un seguit d'aplicacions relacionades amb la traducció automàtica que ha popularitzat encara més l'ús d'aquests sistemes. Un bon exemple és la integració de tècniques d'OCR i traducció automàtica en aplicacions que permeten traduir petits textos fotografiats directament des de la càmera del dispositiu. La possibilitat de fotografiar textos i traduir-los és especialment interessant per a llengües que fan servir alfabets diferents al de l'usuari. Les aplicacions que fan servir aquestes tècniques el que fan és connectar-se a un servidor que du a terme les tasques d'OCR i traducció i que envia el resultat al dispositiu.

4.4.d. La traducció automàtica de veu

Els sistemes de traducció automàtica més coneguts són els que tradueixen un text d'una llengua a una altra. Hi ha sistemes, però, capaços de traduir missatges de veu en una llengua i produir el missatge de veu traduït a una altra llengua. En l'estratègia més senzilla, un sistema de traducció automàtica de veu es composa dels següents mòduls:

- Sistema de reconeixement de veu (*speech recognition*) capaç de transformar el missatge parlat en la llengua de partida a un text.
- Sistema de traducció automàtica que tradueix el text en la llengua de partida a un text en la llengua d'arribada
- Sistema de síntesi de veu (*speech synthesis*) que transforma el text traduït en veu en la llengua d'arribada

Hi ha sistemes de traducció automàtica de veu, però, que no són la simple concatenació d'aquests mòduls, sinó que combinen els models estadístics de les diferents etapes per poder millorar el resultat final.

La combinació d'aquests sistemes amb la disponibilitat de dispositius mòbils ofereix grans possibilitats en l'àmbit de la comunicació entre persones que parlen diferents llengües. També són importants les aplicacions militars d'aquests tipus de sistemes.

4.4.e Traducció automàtica i subtitulació

Actualment hi ha sistemes capaços de subtitular automàticament un vídeo. Aquestes eines es basen en el reconeixement de veu, que converteix el senyal de veu en text i aquest text s'insereix automàticament en el vídeo com a subtítol. Molts d'aquests serveis ofereixen també la traducció automàtica d'aquests subtítols. Aquestes funcionalitats poden ser interessants però cal tenir en compte que es poden produir errors tant en la generació automàtica dels subtítols originals com en la seva traducció.

4.4.f. Traducció automàtica i llenguatge de signes

També s'està experimentant en l'àrea del llenguatge de signes per a persones sordes. Aquests sistemes , com per exemple el Kinect de Microsoft[4] o el Sisi de IBM[5], transformen un missatge de veu en signes mitjançant l'ús d'avatars.

4.4.g. Traducció automàtica i recuperació i extracció d'informació

La *recuperació d'informació* (*information retrieval*) pretén trobar els documents interessants d'una col·lecció de documents a partir d'una cadena de cerca de l'usuari. Sovint la col·lecció de documents està en diversos idiomes i llavors es parla de *cross-language information retrieval* (CLIR), que el que intenta és trobar documents rellevants a partir d'una consulta de l'usuari quan aquesta consulta pot estar en una llengua diferent de la dels documents. Aquesta tècnica implica estratègies relacionades amb la traducció automàtica.

L'*extracció d'informació* (IE-*information extraction*) és la tasca d'extreure automàticament informació estructurada a partir de documents no estructurats o semiestructurats. Els documents dels quals volem extreure la informació estructurada poden estar en diversos idiomes. En aquests casos també es combinen estratègies d'extracció d'informació amb estratègies de traducció automàtica.

Altres tasques que atreuen l'atracció dels investigadors és el resum automàtic multilingüe i els sistemes de pregunta-resposta també multilingües. En el primer cas es pretén crear un resum en una llengua d'un document escrit en una altra llengua. Els sistemes de pregunta-resposta (*question-answering*) són capaços de respondre a preguntes formulades en llenguatge natural a partir d'informació enmagatzemada en bases de dades. El fet de poder formular aquestes preguntes i rebre també la resposta en diversos idiomes és una tasca complexa.

4 http://research.microsoft.com/en-us/collaboration/stories/kinect-sign-language-translator.aspx
5 Say It, Sign It, http://mqtt.org/projects/sisi

4.5. La traducció automàtica a Internet

A Internet es poden trobar una gran quantitat de productes i serveis relacionats amb la traducció automàtica. Com hem comentat en apartats anteriors, el fet d'estar disponible a Internet ha popularitzat l'ús d'aquest tipus de sistemes, cosa que ha fet que la societat sigui més conscients de la utilitat i limitacions de la traducció automàtica. En aquest apartat presentarem alguns dels sistemes més populars i els classificarem atenent a 5 variables, Aquestes variables són:

- **Mode d'accés**: que pot ser el clàssic a través una plana web o automàticament mitjançant una API (A*pplication Programming Interface*). La connexió automàtica permet que un programa es connecti automàticament al sistema de traducció, li enviï una petició de traducció i rebi el resultat. Aquest tipus de connexió permetrà la connexió de sistemes de traducció assistida amb sistemes de traducció automàtica (com veurem a l'apartat següent).
- **Utilització: on-line o descàrrega**. La majoria de sistemes de traducció automàtica que s'ofereixen per Internet, almenys de forma gratuïta, permeten únicament la seva utilització on-line. Una excepció interessant, com veurem, és el cas d'Apertium.
- **Preu: gratuït/de pagament**. Molts sistemes comercials ofereixen serveis gratuïts amb una limitació del nombre de paraules a traduir. Per a traducció massiva de documents habitualment es requereix algun tipus de pagament. També hi ha empreses que ofereixen la traducció automàtica gratuïta i ofereixen paral·lelament un servei de postedició de pagament.
- **Servei principal o servei de valor afegit**. Les principals empreses que desenvolupen sistemes de traducció automàtica tenen presència a Internet i ofereixen l'accés al seus productes (habitualment amb alguna limitació pel que fa a la quantitat de text a traduir). Però les empreses desenvolupadores no són les úniques que ofereixen traducció automàtica. Molts dels sistemes de traducció automàtica disponibles a Internet s'ofereixen com a valor afegit d'un altre tipus de servei. Entre aquests cal destacar les opcions de traducció automàtica que ofereixen els principals cercadors d'Internet.
- **Estratègia de traducció automàtica**: intentarem determinar quina de les estratègies estudiades fa servir el sistema de traducció: directa, de transferència, interlingua, estadística o basada en exemples.

Google Translate (https://translate.google.com)

El motor de traducció de Google està disponible per a 80 llengües i en totes les combinacions. L'estratègia de traducció és estadística. A Google Translate es pot accedir de diverses maneres:

- Amb la clàssica interfície on es pot escriure o enganxar l'original i seleccionar la llengua de partida (que també pot ser detectada automàticament) i la d'arribada. En la mateixa interfície es poden escriure adreces d'Internet per poder traduir planes web i navegar en la versió traduïda automàticament. Permet també pujar documents en diversos formats i traduir-los. Aquests accessos són gratuïts.
- També es pot accedir al traductor mitjançant una API que permet que una aplicació (per exemple de traducció assistida) es connecti automàticament i recuperi la traducció. Aquest accés actualment és de pagament.
- Google també ofereix el *Translator Toolkit* que és una eina de traducció assistida on-line. L'usuari pot pujar memòries de traducció i glossaris terminològics que es faran servir durant la traducció. També es podrà fer servir Google Translate, ja que quan passem d'un segment a un altre es mostra també el resultat de la traducció automàtica. L'accés a aquesta eina és gratuït.
- Google Translate admet suggeriments dels usuaris per millorar la qualitat de les traduccions.

Microsoft Bing Translator (http://www.bing.com/translator/)

El traductor automàtic de Microsoft fa servir també una estratègia estadística i està disponible per a 44 llengües i en totes les direccions. Ofereix diverses formes d'accés:

- La clàssica interfície per traduir frases, planes web i documents. L'accés és gratuït.
- L'accés per API, que és gratuït fins a un límit de 2.000.000 de caràcters mensuals[6] i a partir d'aquest límit és de pagament.
- Microsoft ofereix també la integració del traductor en moltes de les seves aplicacions, com Word, Excel, etc.
- Microsoft també ofereix molts serveis a empreses relacionats amb l'ús del seu traductor automàtic.

Systran (http://www.systransoft.com/)

Systran és un dels sistemes de traducció automàtica més antics que encara està en funcionament i ofereix programes de traducció per instal·lar a l'ordinador. Fa servir una estratègia de traducció directa i està disponible per les següents llengües (però només en algunes de les combinacions): anglès, castellà, alemany, xinès, coreà, francès, grec, italià, japonès, holandès, polonès, portuguès, rus, suec i àrab. Des de la seva web es poden fer traduccions gratuïtes de petits textos. Aquest sistema de traducció es feia servir en *Yahoo Babelfish* (servei que ja no està en funcionament) i en els serveis de traducció de Google fins a l'any 2007, moment en què va ser substituït pel seu propi sistema estadístic.

Apertium (http://www.apertium.org/)

Apertum és un sistema de transferència superficial basat en regles. És de programari lliure i es distribueix sota llicencia GNU General Public License. Apertium es va dissenyar inicialment per traduir entre llengües properes (com català i castellà) però es va expandir per tractar llengües més diferents (com per exemple català i anglès). El sistema Apertium es pot descarregar i instal·lar en un ordinador (preferentment amb sistema operatiu Linux, tot i que també hi ha disponibles instal·ladors per Windows) i proporciona els següents elements:

- Un motor de traducció automàtica independent de la llengua.
- Eines per manipular les dades lingüístiques necessàries per a construir un sistema de traducció automàtica per a qualsevol parell de llengües
- Dades lingüístiques per a una gran quantitat de parells de llengües.

Actualment el sistema està disponible per a més de 30 parells de llengües, tot i que no tots els parells es troben en el mateix estadi de desenvolupament. Des de la web es poden traduir petits textos i documents.

Open Logos (http://logos-os.dfki.de/)

Open Logos es un sistema de traducció automàtica de codi obert que es basa en el sistema Logos, que va ser, junt amb Systran, un dels primers sistemes comercials. Fa servir una estratègia de traducció directa i tradueix de l'anglès i l'alemany al francès, italià, castellà i portuguès.

El sistema es distribueix sota una doble llicència: per una banda la de programari lliure GNU-GPL, que permet fer servir i redistribuir el programari lliurement; per una altra banda es pot adquirir una llicència comercial que permet integrar el sistema de traducció en aplicacions propietàries.

6 Cal tenir en compte que les condicions d'accés al servei poden canviar en qualsevol moment.

4.6. Combinació de traducció automàtica i traducció assistida

La majoria de sistemes de traducció assistida actuals permeten una connexió amb sistemes de traducció automàtica, tant si aquests estan instal·lats al nostre ordinador, com si es tracta de sistemes remots. La connexió entre aquests sistemes acostuma a ser molt simple: a l'usuari se li mostren les coincidències obtingudes de la memòria de traducció i la provinent de la traducció automàtica. Depenent de l'índex de similitud de la recuperada de la memòria, s'insereix aquesta o la provinent de la traducció automàtica. Quan no hi ha cap segment en la memòria amb un índex de similitud suficient, s'insereix també el provinent de la traducció automàtica.

La integració podria anar més enllà i en un futur s'espera que els sistemes de traducció assistida puguin combinar d'una manera intel·ligent fragments de segments provinents de la memòria amb fragments provinents de traducció automàtica. En Simard (2009) es presenta un sistema d'integració de traducció assistida amb un sistema de traducció automàtica estadístic a partir de frases. Aquesta integració precisa de de sistemes de traducció automàtica que es comportin més com el mòdul de memòria de traducció dels sistemes de traducció assistida. Això implica per una banda produir traduccions automàtiques que siguin consistents amb la memòria de traducció quan es troben segments amb un alt índex de similitud; i per l'altra banda desenvolupar un mòdul del sistema de traducció automàtica que sigui capaç de filtrar les traduccions automàtiques que tingui poca probabilitat de ser útils.

En Forcada (2014) trobem una interessant proposta d'ampliació de l'estàndard TMX per a incloure informació provinent de diferents fonts (memòries de traducció, sistemes de traducció automàtica, bases de dades terminològiques, glossaris i alineadors estadístics a nivell de paraula). La idea és proporcionar tota aquesta informació a l'eina de traducció assistida de manera que sigui capaç d'identificar coincidències a nivell subsegmental.

4.7. Postedició de traducció automàtica

La postedició aplicada a la traducció automàtica és el procés de millorar una traducció generada per un sistema de traducció automàtica. El procés de postedició el du a terme una persona amb una formació específica per aquesta tasca de manera que l'esforç i el temps dedicat sigui el mínim possible. Per a parells de llengües properes per als que hi ha disponibles sistemes de traducció automàtica prou madurs la traducció automàtica amb postedició pot ser una estratègia molt productiva per a traduir documents assolint una molt bona qualitat final.

El procés de postedició està relacionat amb el procés de preedició, és a dir, el fet de modificar el text de partida amb l'objectiu d'assolir una millor qualitat de traducció automàtica, per exemple, aplicant algun tipus de llenguatge controlat. Amb la preedició es canvien les oracions amb estructures sintàctiques complexes i lèxic poc habitual per unes altres més simples que puguin ser traduïdes automàticament de manera més satisfactòria. El procés combinat de preedició + traducció automàtica + postedició és molt habitual per a la traducció de documentació tècnica.

El procés de postedició, però, està molt sovint present en fluxos de treball en els que no es du a terme preedició. En el cas que es vulgui una traducció d'un text original mantenint certa fidelitat a l'original, no convé preeditar-lo, ja que els canvis d'estructures es veurien reflectits a la traducció.

Així doncs, la postedició consisteix en l'edició de traduccions realitzades automàticament per assegurar que s'assoleixi el nivell de qualitat acordat entre el client i el proveïdor.

Es distingeixen dos tipus de postedició:

- **La postedició simple** (*light postediting*): l'objectiu és assolir una traducció que sigui comprensible i implica una intervenció mínima del posteditor. La traducció final tindrà una finalitat d'assimilació.
- **La postedició completa** (*full postediting*): l'objectiu és assolir una traducció correcta i estilísticament apropiada que pugui ser emprada tant per finalitats d'assimilació com de disseminació. Aquest tipus de postedició, en el seu grau màxim, té com a objectiu que el nivell de qualitat de la traducció la faci indistingible de la traducció realitzada per un traductor humà.

La postedició de textos traduïts automàticament pretén augmentar la productivitat i eficiència del procés de traducció. Tot i que hi ha estudis que intenten establir aquest grau de millora de la productivitat, encara no queda del tot clar si aquesta millora depèn del parell de llengües ni de quin percentatge d'estalvi de temps s'assoleix (que sembla que són d'entre un 15% i un 40%).

La professió de posteditor és relativament nova i requereix una formació específica. Aquesta formació específica encara no està totalment introduïda en els plans estudis de les titulacions relacionades amb la traducció i molt sovint el professional s'ha format a partir d'una pràctica continuada en aquest sector. Tampoc hi ha unanimitat pel que fa als processos que es duen a terme ni les eines emprades. En alguns casos, el posteditor s'enfronta a un text ja totalment traduït, i té accés també al text original. En altres casos aquesta postedició es fa dins d'una eina de traducció assistida i el posteditor té accés automàtic, a l'original, la traducció, memòries de traducció i bases de dades terminològiques.

Hi ha algunes pautes i consells donats per algunes institucions, entre les que podem destacar les que proporciona TAUS[7]. Els punts que considerem més importants per fer una tasca de postedició amb garanties són:

- La persona ha de tenir una bona formació en les llengües de partida i d'arribada, i preferentment formació i experiència en traducció. Com hem comentat molts posteditors s'han format mitjançant la pràctica. En aquests casos caldrà proporcionar unes pautes i suport, així com comentaris sobre les tasques realitzades. D'aquesta manera la formació serà més ràpida i efectiva.
- Cal que el posteditor conegui les característiques principals del sistema de traducció automàtica que es fa servir (quina estratègia de traducció fa servir, si és un sistema general o específic per l'especialitat que s'està traduint, etc). D'aquesta manera el posteditor es podrà fixar més en aquells aspectes que siguin més susceptibles de contenir errors.
- És imprescindible pactar amb el client quin és el resultat que espera: serà molt diferent si el que vol és tenir una traducció que li permeti comprendre el contingut del text, o bé si vol una traducció que pugui ser publicada. El client també haurà de tenir clar que el preu d'un servei i de l'altre són diferents.
- Cal assegurar-se que el text original sigui correcte. En alguns casos específics també es podrà optar per fer una preedició que asseguri que el resultat de la traducció automàtica sigui òptim.
- És també important que els posteditors tinguin accés ràpid a tota la informació necessària: text original, traducció, bases de dades terminològiques i memòries de traducció.
- És important que els posteditors proporcionin informació sobre els tipus d'errors més habituals. Aquesta informació serà interessant per a introduir millores al sistema de traducció automàtica. Ara bé, si es demana aquesta informació als posteditors, han de rebre també una compensació econòmica.

Tot i que la majoria de les eines de traducció assistida habituals poden fer-se servir amb èxit en la tasca de postedició, hi ha diversos estudis que pretenen dissenyar eines específiques per aquesta tasca. En Aziv (2012) i Roturier (2013) podem trobar algunes propostes. També és interessant l'estudi de Vieira (2011) que mostra una classificació dels sistemes de traducció automàtica disponibles des del punt de vista de la tasca de la postedició.

7 https://evaluation.taus.net/resources/guidelines/post-editing/machine-translation-post-editing-guidelines-spanish

4.8. Conclusions

En aquest capítol hem exposat els conceptes bàsics sobre la traducció automàtica que tot traductor hauria de conèixer. La traducció automàtica no ha esdevingut, com alguns vaticinaven en els primers anys de la història d'aquests sistemes, una amenaça per a la professió de traductor, sinó que s'ha convertit en un aliat a la seva tasca. En els darrers anys la qualitat dels sistemes de traducció automàtica ha millorat enormement i s'esperen més millores així com la disponibilitat de molts més parells de llengües en els propers anys. Tot i aquesta gran millora, ningú no dubta avui dia que el traductor humà serà imprescindible sempre per assolir els nivells de qualitat òptims.

4.9. Per ampliar coneixements

4.9.1. Història de la traducció automàtica

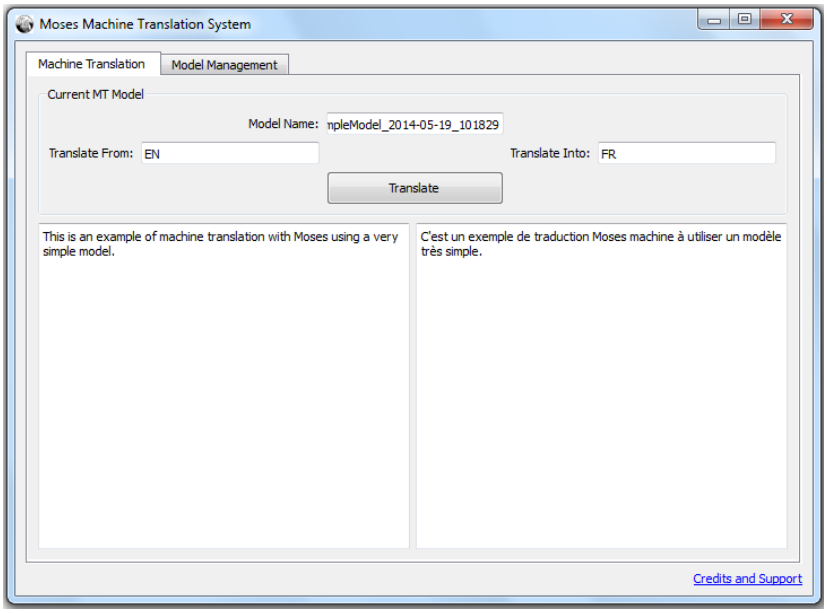
Per ampliar molt més sobre la història de la traducció automàtica, es pot llegir:

Hutchins, J.(1986) *Machine Translation: past, present, future* Ellis Horwood Series in Computers and their Applications 382 pp. Chichester (UK): Ellis Horwood, 1986. (ISBN: 0-85312-788-3) New York: Halsted Press, 1986. (ISBN: 0-470-20313-7)

Que està accessible lliurement a: http://www.hutchinsweb.me.uk/PPF-TOC.htm

4.9.2. Moses: un sistema de traducció automàtica estadística

Moses (http://www.statmt.org/moses/) és un sistema de traducció automàtica estadística que permet entrenar i fer servir sistemes de traducció automàtica per a qualsevol parell de llengües. Disposa d'un descodificador eficient que permet trobar una de les traduccions més probables a partir dels models entrenats. El sistema compta amb una documentació acurada i una sèrie de tutorials que permeten introduir-se en les tasques d'entrenament de sistemes de traducció automàtica estadística d'una manera relativament fàcil. Es requereix, però, una certa expertesa en l'ús de sistemes Linux/Unix. Per a usuaris de Windows hi ha una versió senzilla i amb interfície gràfica que permet provar sistemes ja entrenats. Podem observar aquesta interfície en la següent figura:



A la mateixa web també es poden descarregar sistemes ja entrenats per a diversos parells de llengües.

4.9.3. Altres sistemes de traducció automàtica

Al mercat hi ha una gran quantitat de traducció automàtica. Es pot trobar un recull de sistemes a: http://en.wikipedia.org/wiki/Comparison_of_machine_translation_applications

Per tenir una idea dels sistemes de traducció automàtica i altre programari relacionat amb llicencia lliure es pot consultar l'enllaç: http://www.computing.dcu.ie/~mforcada/fosmt.html

Bibliografia

Aziz, W.; Sousa, S. C. M.; Specia, L. (2012). *PET: a tool for post-editing and assessing machine translation*. In The Eighth International Conference on Language Resources and Evaluation, LREC '12, Istanbul, Turkey. May 2012.

Forcada, M. (2014) *On the Annotation of TMX Translation Memories for Advanced Leveraging in Computer-aided Translation. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14).* Reykjavik (Iceland). Eds. Nicoletta Calzolari (Conference Chair) and Khalid Choukri and Thierry Declerck and Hrafn Loftsson and Bente Maegaard and Joseph Mariani and Asuncion Moreno and Jan Odijk and Stelios Piperidis. European Language Resources Association (ELRA). ISBN 978-2-9517408-8-4.

Green S., Heer J. and Manning c. D. (2013) *The Efficacy of Human Post-Editing for Language Translation.* ACM Human Factors in Computing Systems (CHI). http://vis.stanford.edu/papers/post-editing

Hutchins, J. 2007. *Machine translation: A concise history. In Computer Aided Translation: Theory and Practice*, C. S. Wai, Ed. Chinese University of Hong Kong.

Hutchins, J. (2009). *Multiple Uses of Machine Translation and Computerised Translation Tools*. In Proceedings of the International Symposium on Data and Sense Mining, Machine Translation and Controlled Languages – ISMTCL 2009. http://www.hutchinsweb.me.uk/Besancon-2009.pdf;

Koehn, P. (2010) *Statistical Machine Translation* Cambridge University Press

Makoto Nagao (1984). *A framework of a mechanical translation between Japanese and English by analogy principle* In A. Elithorn and R. Banerji. Artificial and Human Intelligence. Elsevier Science Publishers.

Oliver A. (2006) *La traducció automàtica a Internet* - Revista Tradumàtica – Traducció i Tecnologies de la informació i la Comunicació 04 : Traducció Automàtica : http://www.fti.uab.cat/tradumatica/revista ISSN 1578-7559

Roturier J., Mitchell L. and D. Silva (2013) *The ACCEPT Post-Editing Environment: a Flexible and Customisable Online Tool to Perform and Analyse Machine Translation Post-Editing* in Proceedings of the MT Summit XIV Workshop on Post-editing Technology and Practice. Nice (France)

Schubert, K. (1988) *The architecture of DLT – interlingua or double direct?* In Maxwell, Dan, Klaus Schubert and Toon Witkam (eds): (1988). *New Directions in Machine Translation.* Dordrecht: Foris.

Simard M. And Isabelle P. (2009) *Phrase-based Machine Translation in a Computer-assisted Translation Environment.* In The Twelfth Machine Translation Summit (MT Summit XII), pages 120--127, Ottawa, Ontario, Canada.

Trujillo, A. (1999) *Translation Engines: Techniques for Machine Translation*. Springer. ISBN 978-1-85233-057-6

Vieira. L.; Specia, L. (2011). A Review of Machine Translation Tools from a Post-Editing Perspective. 3rd Joint EM+/CNGL Workshop Bringing MT to the User: Research Meets Translators (JEC 2011), Luxembourg.

Llicència d'aquest document