

LA TRADUZIONE AUTOMATICA

Seminario di Linguistica
Computazionale

Prof. A. Cappelli

Studente: Marco Maiani

Corso di laurea in Scienze della Comunicazione

1

INTRODUZIONE

La traduzione automatica (**MT = Machine Translation**) prevede la possibilità di affidare ad un computer la traduzione, parziale o completa, di un testo, riducendo al minimo la successiva revisione da parte dell'utente.

In questo settore di ricerca sono coinvolte varie discipline, tra cui Informatica, Linguistica Computazionale, Intelligenza Artificiale e Scienze Cognitive.

2

L'interesse in ricerche di questo tipo è giustificato dai numerosi campi di applicazione della MT, che si concentrano in vari ambiti:

Socio-politico: comunità dove si parla più di una lingua

Commerciale: fornire informazioni sui prodotti in lingue diverse

Scientifico: terreno di test per molte idee in informatica, intelligenza artificiale e linguistica

Filosofico: l'efficienza con la quale si può automatizzare la traduzione è un'indicazione dell'efficienza con la quale si può automatizzare il pensiero.

3

SISTEMI DI TRADUZIONE ON-LINE

Numerosi servizi di MT sono al momento disponibili on-line, permettendo all'utente di ottenere in pochi secondi la traduzione di una parola, una frase o un intero testo da una lingua sorgente a una lingua obiettivo (es.: dall'inglese all'italiano). E' sicuramente allettante la possibilità offerta da alcuni sistemi di tradurre pagine Web. Mettere in rete versioni in altre lingue di un proprio sito Internet è essenziale per assicurare una visibilità del sito a livello internazionale.

4

Purtroppo i sistemi di MT attualmente in uso commettono ancora molti errori e rendono indispensabile una revisione da parte di un traduttore umano per giungere a risultati di buona qualità.

Questo lavoro parte proprio dall'analisi di varie frasi che il sistema dà come risultato di una traduzione automatica. In alcuni casi è stata selezionata l'opzione "**Traduci questa pagina**" disponibile per certi siti Internet. Altre volte sono state richieste apposite traduzioni, impostando parole e frasi nelle caselle di testo di diversi traduttori automatici.

5

Se si prendono, ad esempio, tre sistemi di MT disponibili gratuitamente on-line (**Freetranslation**, **Google** e **Synthema**) si ottengono risultati differenti per la stessa frase inserita. Nei seguenti casi è stata richiesta una traduzione dall'inglese all'italiano.

INPUT:

The plane took off an hour late

OUTPUT:

- 1) L'aereo ha tolto un'ora tardi (*Freetranslation*)
- 2) L'aereo ha decollato un'ora ritardata (*Google*)
- 3) L'aereo ha richiesto un'ora tardi (*Synthema*)

6

INPUT:

We had run out of money

OUTPUT:

- 1) Eravamo rimasti senza denaro (*Freetranslation*)
- 2) Avevamo funzionato da soldi (*Google*)
- 3) Avevamo esaurito denaro (*Synthema*)

INPUT:

Self guided cycling and walking routes

OUTPUT:

- 1) Stesso ha guidato il cycling e che le camminando strade (*Freetranslation*)
- 2) Itinerari ciclisti ed ambulanti guidati auto (*Google*)
- 3) Lo sé ha guidato vie che e iterando camminando (*Synth.*)

7

Da questi esempi si evidenzia come si ottengano risultati notevolmente diversi a seconda del modo in cui viene programmata ogni macchina per la traduzione. In generale, comunque, resta sempre un margine di errore che non è possibile annullare neppure ricorrendo a complessi programmi di MT. Bisogna poi considerare il fatto che ogni lingua presenta caratteristiche proprie in quanto a strutture sintattiche e morfologiche, oltre che alla suddivisione dello spettro semantico, che rendono certe traduzioni più difficili da automatizzare rispetto ad altre. In questo lavoro si fa riferimento soprattutto alla traduzione dall'inglese all'italiano e viceversa.

8

COME FUNZIONA LA MT

Per capire da cosa derivino gli errori nella traduzione automatica bisogna innanzi tutto individuare i meccanismi che sono alla base del funzionamento di queste macchine.

Fra i motori per la MT si distinguono quelli ad architettura **Transformer** (sistema tradizionale) e quelli ad architettura **Linguistic Knowledge** (di recente sperimentazione).

9

ARCHITETTURE TRANSFORMER

Le frasi di ingresso vengono trasformate in frasi di uscita sostituendo le parole del linguaggio sorgente con il loro equivalente nel linguaggio obiettivo (sulla base di un dizionario bilingue). Le parole così ottenute vengono poi riordinate per soddisfare le regole grammaticali del linguaggio obiettivo.

Il vantaggio principale di questo sistema è di non bloccarsi in condizioni di errore, quando incontra input che contengono parole o strutture grammaticali sconosciute (conoscenza limitata della grammatica del linguaggio sorgente).

Del resto, in certi casi può produrre risultati del tutto inaccettabili nel linguaggio obiettivo (conoscenza limitata della grammatica del linguaggio obiettivo).

10

ARCHITETTURE LINGUISTIC KNOWLEDGE

Le architetture LK richiedono una grammatica dettagliata sia del linguaggio sorgente che del linguaggio obiettivo. Inoltre prevedono una grammatica comparativa addizionale usata per relazionare ogni rappresentazione della frase sorgente a qualche rappresentazione corrispondente nel linguaggio obiettivo, riconoscendone le differenze.

Con questo sistema si ottengono uscite più corrette grammaticalmente che con l'architettura Transformer, in quanto si producono sempre frasi "well-formed" anche nel caso di traduzioni non accurate. Il sistema fallisce in presenza di frasi in ingresso grammaticalmente complicate, che non vengono riconosciute, anche se corrette.

11

LA CONOSCENZA SINTATTICA

Entrambe le architetture utilizzate per la MT utilizzano particolari tecniche per rappresentare la conoscenza sintattica necessaria per la traduzione.

Prendiamo come esempio le seguenti frasi:

- 1) They **work** six days a week
- 2) Your **work** is gaining importance

Il sistema deve scomporre ciascuna frase in una struttura sintattica in modo da individuare la funzione di ciascuna parola e procedere alla successiva traduzione. Uno stesso termine può avere, infatti, valenza grammaticale diversa a seconda della frase in cui è collocato. Nel primo esempio la parola *work* è usata come verbo, mentre nel secondo caso come nome.

12

Lo studio sintattico riguarda l'analisi della struttura costituente e delle relazioni grammaticali.

Analizzare la **struttura costituente** significa associare ad ogni parola la categoria a cui appartiene, per es. nome (N), verbo (V), aggettivo (A), avverbio (ADV), preposizione (P)...

Per ogni linguaggio esistono delle regole che prescrivono come queste categorie possono essere combinate per creare frasi corrette. Ad esempio nella lingua inglese ogni frase grammaticalmente corretta è formata da un sintagma nominale (NP) seguito da un verbo modale o ausiliario (AUX), che può anche non esserci, e da un sintagma verbale (VP). Esempio:

The keeper should open the door

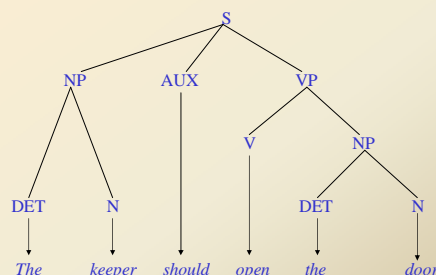
Nel sintagma nominale il nome può essere preceduto da un determinante (DET) come *the, a, many...*

13

Se cambiamo l'ordine degli elementi nella frase, la regola precedente non viene più rispettata e la frase risulta priva di senso:

Open keeper the should door the

Ecco la struttura ad albero della nostra frase (S):



14

Oltre a ricavare l'albero della struttura costituente il sistema deve procedere ad un'analisi delle **relazioni grammaticali** tra gli elementi della frase, individuando **SOGGETTO, OGGETTO, COMPLEMENTO...**

In inglese i **soggetti** sono normalmente sintagmi nominali che stanno prima del verbo, mentre gli **oggetti** tendono a trovarsi dopo il verbo. Non in tutte le lingue, però, funziona così. Per esempio, in giapponese l'ordinamento normale delle parole è soggetto-oggetto-verbo, in irlandese è verbo-soggetto-oggetto. In molte lingue, come il russo, il verbo, il soggetto e l'oggetto possono apparire in qualsiasi ordine.

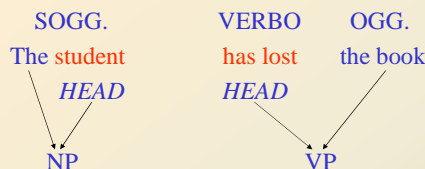
All'interno di ogni sintagma troviamo un elemento (**HEAD**) che è il più importante dal punto di vista grammaticale e che guida il significato. In un sintagma nominale l' **HEAD** è dato dal nome, in un sintagma verbale dal verbo.

15

Se prendiamo ad esempio la frase

The student has lost the book

possiamo associarvi la seguente rappresentazione:



Se riprendiamo ora le due frasi iniziali contenenti la parola *work* possiamo capire come il sistema riesca a distinguere i modi di tradurre questo termine in base alla funzione che esso svolge.

16



I DIZIONARI

Se si richiede la traduzione automatica del sito Web della città di Brighton ci si può imbattere in una versione italiana quanto mai bizzarra. In rosso le parole erroneamente tradotte dal sistema, in corsivo quelle che sono state lasciate nell'originale:

There's *lots* of *stuff* to be had in the City of Brighton and *Hove*.

This is one of Britain's most *happening* places with theatres, cinemas and restaurants all over the place, *loads* of music and *clubs*, *cafés* and *pubs* and...*well*, all the things that make life worth living.

Ci sono *lotti* di *stuff* da avere nella città di Brighton e *hanno sollevato*.

Ciò è uno dei posti *accidenti* della Gran Bretagna con i teatri, i cinematografi ed i ristoranti dappertutto il posto, i *carichi* di musica e dei *randelli*, i *cafés* e le *pubblicazioni* e...*well*, tutte le cose che fanno *la vita degno vivere*.

18

Si tratta di due frasi con un'altissima concentrazione di errori che hanno a che fare con la ricerca del termine da tradurre all'interno dei dizionari del sistema.

Nel caso delle parole non tradotte l'errore si deve al fatto che esse non sono state inserite nel dizionario bilingue, mancanza piuttosto grave per quanto riguarda termini ricorrenti come *stuff*, *well* o *café*. Questo problema può essere notevolmente ridotto se la macchina attinge ad un dizionario con un sufficiente numero di lemmi. In tal modo si garantisce che ad ogni parola (o quasi) corrisponda una traduzione, per quanto appropriata o meno essa sia.

Ma è sempre bene tradurre tutte le parole del testo? Così facendo si può incorrere nell'assurdità di considerare la città di Hove come il passato di "to heave" (= sollevare). Una cosa simile avverrebbe nel caso della frase *Reading is a city*, che il sistema traduce *La lettura è una città*.

Il traduttore PeTra, messo a punto da Synthema, propone all'utente l'opzione "non tradurre i termini in maiuscolo all'interno delle frasi", in modo da far valutare a chi richiede il servizio se questa alternativa sia migliore della traduzione integrale.

19

D'altra parte, lasciare invariati tutti i vocaboli in maiuscolo (pur eccettuando quelli che seguono un punto) comporterebbe errori abbastanza frequenti. E' il caso, per esempio, dei nomi delle nazioni (es. France), di tutti gli aggettivi di nazionalità, che in inglese si scrivono con la lettera maiuscola (es: British), di certe cariche importanti (es: the President, the Prince) o di nomi di organizzazioni o eventi (es: European Union, Olympic Games). Come regola generale sembra, quindi, preferibile tradurre tutto (anche il maiuscolo) ed eventualmente arginare certe eccezioni più semplici da trattare, come gli indirizzi o i nomi propri di persona. Ad esempio, per evitare che il sistema traduca *Lavender Street* come *via della lavanda* basterebbe introdurre la regola per cui i termini *Street*, *Road*, *Avenue*, *Square*, ecc. devono essere lasciati invariati, così come i termini in maiuscolo che li precedono (es: *Bond Street* non diventerebbe più *strada obbligazionaria*). Lo stesso principio è applicabile in presenza di appellativi come *Mr*, *Mrs*, *Miss*, *Prof*, *Dr*, ecc. ai quali segue un nome maiuscolo che non deve essere tradotto (es: *Mr Green* rimarrebbe invariato, o al limite *Signor Green*, ma non *Signor Verde*).

20

Riprendendo il testo iniziale possiamo vedere come nella traduzione siano stati commessi altri eclatanti errori di natura semantica. Quando si devono trattare **parole polisemiche**, che hanno cioè più di un significato, possono sorgere traduzioni errate, come è il caso di *lots* = lotti, *loads* = carichi, *clubs* = randelli, *pubs* = pubblicazioni.

Perché il sistema fa corrispondere a certe parole delle uscite completamente sbagliate?

Consideriamo la frase seguente:

This is one of Britain's most happening places with loads of music and clubs, cafés and pubs.

Quando il sistema incontra il termine *clubs* ricerca nel dizionario il lemma *club*, da cui la parola *clubs* è derivata secondo un processo di "inflection". Si tratta di regole morfologiche che modificano la struttura interna delle parole. Nel caso dell'"inflection" una parola è derivata dalla forma di un'altra parola, acquisendo particolari caratteristiche grammaticali. Una volta isolate le eccezioni, questo processo si applica a tutti i membri di una data categoria.

21

Il termine *clubs* offre l'esempio di due regole di "inflection" della lingua inglese che possono essere state applicate:

1) plurale di un nome (*club*)

2) terza persona singolare di un verbo al "simple present" (*to club*)

In entrambi i casi la regola prevede di aggiungere una *s* al lemma nella sua forma base (quella presente nel dizionario). Le poche eccezioni vengono descritte esplicitamente. Il componente morfologico di cui è dotato il sistema deve associare alle parole "inflected" la corrispondente "head word" (forma base) ed estrapolare il significato che il processo di "inflection" ha aggiunto alla parola base. Il primo passo da compiere è quindi quello di stabilire se ci troviamo di fronte a un nome o a un verbo, visto che l'aggiunta della *s* è contemplata in entrambi i casi. Per risolvere questa ambiguità bisogna utilizzare le tecniche di riconoscimento sintattico viste in precedenza. Per capire qual è la giusta categoria da attribuire a *clubs* basta verificare se è grammaticalmente possibile avere un nome o un verbo nella posizione in cui la parola compare.

22

Dato che i sistemi di traduzione sono dotati della capacità di riconoscere frasi grammaticalmente corrette in base al rispetto di certe regole del linguaggio considerato, risulta facile arrivare ad un'unica interpretazione sintattica degli elementi della frase. Per esempio, non esistono in inglese sequenze grammaticali formate dall'articolo *the* seguito da un verbo. Nella frase sottostante, quindi, *clubs* non può essere analizzato come terza persona del presente del verbo "to club", ma piuttosto deve essere considerato come plurale del nome "club", unica alternativa coerente con le regole grammaticali conosciute dal sistema.

The clubs where I usually go are...

Se *clubs* fosse un verbo dovrebbe essere preceduto da un soggetto, ma *the* non può fungere da soggetto di *clubs*, quindi viene selezionata l'alternativa *clubs* = nome.

Dal momento che in inglese il soggetto è sempre espresso, davanti al verbo deve esserci obbligatoriamente un nome o un pronome personale. Nel nostro caso, scartata l'ipotesi *clubs*, il sistema

23

cercherà il verbo in un'altra parola della frase, ad esempio *are*, che quadra col fatto che il soggetto è al plurale.

Applicando l'analisi sintattica alla frase da cui siamo partiti

This is one of Britain's most happening places with loads of music and clubs, cafés and pubs.

concludiamo che *clubs* è il plurale del sostantivo "club", ma a questo punto sorge un nuovo problema: **disambiguare una parola polisemica**, cosa possibile solo se il sistema lavora con un dizionario con dettagliate informazioni sulle parole.

24

TIPICI DI INFORMAZIONE SULLE PAROLE

I dizionari sono le componenti più grandi di un sistema per la MT in quanto a informazione in essi contenuta. Per avere buone prestazioni è indispensabile che un dizionario non contenga solo una semplice lista di parole con la traduzione corrispondente, ma molte altre preziose informazioni per risolvere il problema dell'ambiguità semantica.

Si tratta di elencare innanzi tutto le **caratteristiche intrinseche della parola**, classificate secondo specifici parametri, che variano in base alla categoria grammaticale a cui la parola appartiene.

Nel caso, per esempio, di un sostantivo il dizionario potrebbe indicare, oltre alla sua forma base (lex) e alla sua categoria grammaticale (cat), se è un nome comune (ntype), concreto (concrete), se ha come attributo l'essere animato (human) o se si trova solo al plurale (number). Quella che segue è la scheda che un dizionario può associare al nostro nome *club*:

25

lex = club
cat = n
ntype = common
number =
human = no
concrete = yes

Oltre alle caratteristiche intrinseche della parola (quelle elencate nell'esempio precedente), in certi casi può essere utile annotare le restrizioni che un termine impone sugli altri vocaboli del suo ambiente grammaticale. Riguardo a ciò esiste un'informazione di "subcategorization", che indica i contesti sintattici all'interno dei quali una parola può apparire, e un'informazione relativa alle "selectional restrictions", che descrivono le proprietà semantiche del contesto stesso.

26

Una tipica informazione riguardo la "subcategorization" può essere l'indicazione che un verbo è transitivo (Tn). Ciò significa che compare come HEAD di frasi con un sintagma nominale SOGGETTO e un sintagma nominale OGGETTO. E' il caso, ad esempio, del verbo "to club" (= picchiare, bastonare):

They clubbed the demonstrators to the ground
SOGG. OGG.

Nelle frasi seguenti i verbi in rosso hanno altre caratteristiche, indicate dalla sigla fra parentesi:

The train arrived immediately (I = intrans.)

I bought Mary a book (Dn.n = ditrans. con un sogg. e due ogg.)

I sold a car to Jim (Dn.pr = ditrans. con un sogg. e due ogg., il secondo introdotto da *to*)

They told him to try again (Cn.t = trans. con un sogg., un ogg. e una clausola infinitivale introdotta da *to*)

27


In ambito semantico intervengono le "selectional restrictions" che ci danno informazioni importanti sul contesto nel quale un certo vocabolo, in genere, si inserisce. Prendiamo, ad esempio, il verbo "to eat" (= mangiare). Se esso è seguito da un OGGETTO (come previsto dalla sua categoria di verbo transitivo), questo OGGETTO o PATIENT del verbo (elemento su cui ricade l'effetto dell'azione espressa dal verbo) deve essere qualcosa di commestibile, concreto e solido (tranne poche eccezioni) perché la frase abbia senso. Inoltre ci si aspetta che il SOGGETTO o AGENT del verbo (l'ente che compie l'azione) sia un essere animato. Possiamo codificare questi vincoli nella nostra grammatica associando le caratteristiche HUMAN e EDIBLE con appropriati nomi nel nostro dizionario. Se il sistema sa che il verbo "to eat" può essere seguito solo da un nome con l'attributo "commestibile" selezionerà, nel caso di una parola polisemica, il significato che soddisfa questa caratteristica, scartando le altre possibili traduzioni.

In corrispondenza del lemma *eat* (verbo) potrebbe essere associata nel dizionario la seguente scheda:

28

lex = eat
cat = v
vtype = main
finite =
pers. =
numb. =
subcat. = **sogg.-ogg.**
sem. agent = **human**
sem. patient = **edible**


La grammatica ora accetterà solo oggetti commestibili per il verbo "to eat" e questo può essere fondamentale per disambiguare certe parole in base al contesto. Prendiamo, ad esempio, la frase



29

La parola inglese *game* è ambigua, in quanto può avere più significati, tra cui "gioco", "partita", "cacciagione". Se nei dizionari del sistema sono state inserite schede dettagliate per ogni accezione della parola "game", verrà selezionata la traduzione che presenta l'attributo "commestibile", scartando quindi le alternative "gioco" e "partita", inconciliabili con le caratteristiche semantiche del verbo.

Tutto questo ragionamento crolla nel caso in cui il testo sia scritto in **stile metaforico**, per cui si va al di là del significato letterale attribuendo a un elemento della frase certe caratteristiche semantiche proprie di un altro. Esempio:



Qui si ricorre ad una metafora per indicare che la macchina in questione richiede molti soldi per essere mantenuta. Non essendo i soldi qualcosa di commestibile, il sistema fallisce nel tentativo di voler soddisfare i vincoli semantici del verbo.

30

RISOLVERE L'AMBIGUITA'

Dotare le macchine per la traduzione di dizionari complessi è l'unico modo per ridurre (ma non eliminare) le possibilità di errore nel caso di parole polisemiche. Tutto sarebbe più facile se ogni vocabolo avesse un solo significato, ma l'ambiguità è un fenomeno molto diffuso nelle lingue umane. Quando una parola ha due possibili significati si dice che presenta **ambiguità 2**, ed è difficile trovare parole che non abbiano almeno questo grado di ambiguità. Nel caso peggiore, una frase contenente 2 parole, ognuna delle quali con ambiguità 2, può avere 4 possibili interpretazioni (2 X 2), una frase con tre parole può comportare fino a 8 traduzioni differenti (2 X 2 X 2), e così via. Secondo questa logica si possono ottenere numeri veramente elevati, ma al traduttore umano risulta, quasi sempre, semplice disambiguare anche una frase con alto grado di ambiguità, analizzando il contesto in cui la frase è inserita e basandosi sul senso comune e sulla conoscenza del mondo.

31

Riprendiamo ora la frase che è stata il punto di partenza della nostra analisi:

This is one of Britain's most happening places with loads of music and clubs, cafés and pubs.

Vediamo come un dizionario cartaceo classifica il lemma *club*, la nostra parola polisemica da disambiguare. La freccia verde indica la traduzione italiana che il sistema dovrebbe abbinare in ciascun caso.

club¹ n (a) *a group of people who meet together regularly, esp. for a particular activity such as a sport* → club; (b) *the building or rooms used by a club* → club

- club together v (-bb-) *to make separate contributions of money, etc. so that the total can be used for a specific purpose* → associarsi

club² n 1) *a heavy stick with one end thicker than the other, used as a weapon* → mazza; 2) *a stick with a specially shaped end for hitting the ball in golf* → mazza

32

- club v (-bb-) *to hit or beat sb/sth with a club or heavy object* ➡ bastonare
 club³ n (a) clubs *one of the four suits in a pack of cards* ➡ fiori; (b) *a card of this suit* ➡ fiori

Se si eccettua “club together” che, in quanto “phrasal verb”, deve essere classificato nel dizionario come lemma proprio, la ricerca del termine **club** dà origine a quattro possibili interpretazioni, precisamente 1 verbo e 3 sostantivi.

Il traduttore automatico disponibile on-line, una volta scartata l’ipotesi verbo, seleziona la traduzione “randello” (equivalente a “mazza” o “bastone”), cioè il secondo significato presente nel dizionario alla voce **club** come sostantivo. Evidentemente il contesto della frase ci suggerisce che si tratta in realtà del luogo di ritrovo che anche in italiano chiamiamo “club”. L’eclatante errore si produce perché il sistema è in grado di far corrispondere alla nostra parola un solo ed unico significato per la categoria nome, che a volte è quello giusto e altre volte no, come in questo caso

33

Se si lavora con un sistema che prevede per ogni lemma una sola traduzione possibile, indipendentemente dal contesto, è chiaro che si produrranno frequenti errori di natura semantica. In questi casi il criterio più logico per l’impostazione dell’unico significato (e della rispettiva traduzione) da associare a una data parola nel dizionario della macchina sembra essere quello della frequenza d’uso. Si cerca di ridurre le possibilità di errore scegliendo l’accezione che ricorre più spesso nell’uso della lingua (che è quella indicata per prima nel dizionario cartaceo). Spesso, però, non ci sono grosse differenze nella frequenza d’uso tra un significato e l’altro di uno stesso termine, quindi la probabilità di errore rimane molto elevata.

Il problema che la MT ha cercato di risolvere è proprio quello di far attivare il significato giusto di una data parola in base al contesto in cui essa compare, codificando ogni parola con certi attributi semantici. Per rappresentare il significato delle parole sono state messe a punto varie tecniche, tra cui quella delle “selectional restrictions” discussa in precedenza e per la quale è stata dimostrata una notevole utilità nel campo della MT. Questa modalità prevede che alle parole vengano

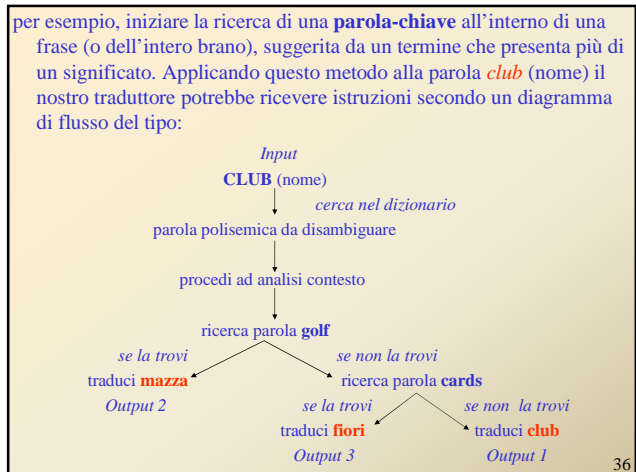
34

associate delle **caratteristiche semantiche** che corrispondono alle loro componenti di senso. Ad esempio, i sostantivi “man”, “woman”, “boy” e “girl” hanno in comune la natura di essere vivente, ma presentano tratti diversi riguardo alle contrapposizioni maschile ➡ femminile e giovane ➡ adulto.

MAN = + HUMAN, + MASCULINE, + ADULT
WOMAN = + HUMAN, - MASCULINE, + ADULT
BOY = + HUMAN, + MASCULINE, - ADULT
GIRL = + HUMAN, - MASCULINE, - ADULT

L’obiettivo è quello di confrontare un dato vocabolo (es. una parola polisemica) con un altro che compare nella stessa frase (es. un verbo) per verificare la compatibilità fra le caratteristiche semantiche di entrambi e attivare così il significato giusto nella traduzione in uscita. Si tratta di effettuare un’**analisi del contesto** volta a cercare elementi che ci indichino in quale senso è usata la frase. Il sistema potrebbe,

35



Una tale procedura, però, è solo ipotetica perché si possono incontrare tantissimi casi in cui gli abbinamenti che il sistema suggerisce portano a conclusioni sbagliate. L'esempio più evidente lo dimostra l'espressione *golf club*, che in base allo schema precedente verrebbe tradotta con "mazza da golf", mentre in certi casi la traduzione appropriata potrebbe essere "club di golf", per cui il sistema fallirebbe in pieno.

Meglio forse mettere in relazione le caratteristiche dell'oggetto con il verbo della frase. La mazza è uno strumento (quindi un oggetto che si usa) e generalmente si impiega per colpire, picchiare qualcuno o qualcosa. C'è buona probabilità che il termine *club* in questa accezione si trovi accompagnato da verbi come "use", "hit", "beat". Se la scheda del lemma *club* rimanda a questi verbi il sistema potrebbe avviare una ricerca in tale direzione e, in caso non trovi nessuna corrispondenza, tradurre con l'alternativa "club".

Insomma, quando si ha a che fare con la conoscenza semantica qualunque soluzione si proponga richiede un notevole grado di complessità e il margine di errore resta sempre alto.

37

LE PROBLEMATICHE DELLA TRADUZIONE

Oltre ai casi di ambiguità semantica possono sorgere problemi di traduzione dovuti alle **differenze strutturali e lessicali** tra i linguaggi. Le **incongruenze lessicali** derivano dal fatto che ogni lingua classifica il mondo in maniera diversa, per esempio esprimendo certi concetti con una singola parola, oppure non inserendone affatto altri nel proprio lessico.

Quando ad un concetto (al quale la lingua sorgente associa un solo termine) la lingua obiettivo fa corrispondere più lessemi (che descrivono aspetti diversi dello stesso concetto) si complica notevolmente il processo di traduzione.

Quelli che seguono sono alcuni esempi di suddivisione dello spettro semantico per certe parole in spagnolo, inglese e francese, a differenza dell'italiano.

38

SPAGNOLO	
portare	<i>llevar (da vicino a lontano; addosso)</i>
	<i>traer (da lontano a vicino)</i>
bello	<i>guapo (di persona)</i>
	<i>bueno (del tempo; di aspetto)</i>
	<i>bonito (di un posto)</i>
	<i>bello (concetto di bellezza più elevato)</i>
INGLESE	
ospite	<i>host (persona che ospita)</i>
	<i>guest (persona ospitata)</i>
FRANCESE	
gamba	<i>jambe (di un umano)</i>
	<i>pied (di un tavolo)</i>

39

Nell'associazione di termini del linguaggio sorgente con quelli del linguaggio obiettivo si possono incontrare i cosiddetti "**lexical holes**". Si tratta di quei casi in cui un linguaggio deve utilizzare un sintagma o un'espressione linguistica per esprimere ciò che in un altro linguaggio è esprimibile con una singola parola (es. "suicidarsi" si traduce in inglese con "to commit suicide", oppure "diventare" è reso in spagnolo con "llegar a ser").

Le **incongruenze strutturali** nascono quando due lingue utilizzano diverse costruzioni per lo stesso scopo e la stessa costruzione per scopi diversi. Nei seguenti esempi si utilizzano strutture differenti per ottenere lo stesso effetto.

- 1) a. *His name (= il suo nome) is Tom*
- b. *Er heißt (= ha nome) Tom*
- c. *Il s'appelle Tom*
 Se llama Tom
 Si chiama Tom

40

- 2) a. *Frank has just seen Paul*
 b. *Frank acaba de ver a Paul*

Il problema fondamentale in questi casi è che la rappresentazione astratta della frase nel linguaggio sorgente e quella della rispettiva traduzione nel linguaggio obiettivo sono notevolmente diverse. Il passare dall'una all'altra richiede regole di trasformazione particolarmente complesse.

Per la frase (2) tali regole si possono articolare nei seguenti punti:

- 1- L'avverbio *just*, quando si trova tra l'ausiliare e il participio passato di un verbo al "present perfect", deve essere tradotto nel verbo *acabar de* (= terminare di) al presente, cosa che appare del tutto innaturale.
- 2- *Frank*, il soggetto di *see*, deve diventare il soggetto di *acabar de*.
- 3- Alcune informazioni riguardo la coniugazione del verbo devono essere prese dal sintagma del quale *see* è la HEAD, e portate sul sintagma la cui HEAD è *acabar de*, mentre ci si aspetterebbe che tale informazione andasse a finire sul sintagma la cui HEAD è la traduzione di *see*, cioè *ver*.

41

Un'altra problematica che un sistema di MT si trova a dover affrontare è quella della traduzione degli **idiomi**. Gli idiomi (o **unità multiparola**) sono espressioni con un significato proprio che è diverso dalla somma dei significati delle parole componenti. Se prendiamo, ad esempio, la frase:

I dropped a brick when I mentioned her ex-husband

e la traduciamo parola per parola arriviamo a questa traduzione:

Ho lasciato cadere un mattone quando ho menzionato il suo ex-marito

E' evidente che l'espressione inglese "to drop a brick" contiene un particolare significato ("fare una gaffe") che va al di là di quello letterale ("lasciar cadere un mattone"). Dal contesto della nostra frase il traduttore umano capirebbe subito che si tratta di un idiomma, ma un sistema di MT non può operare volta per volta questa distinzione. Il metodo migliore per la gestione degli idiomi sembra quello di rappresentarli come unità singole nel dizionario della macchina, dove verrebbe in questo caso inserita un'entrata lessicale *drop_a_brick*.

Il vero problema con gli idiomi è che questi non hanno generalmente

42

una forma fissa. Nell'esempio precedente il verbo "to drop" è usato al passato (*dropped*), forma comunque riconoscibile dal sistema trattandosi di parola derivata tramite processo di "inflection".

Il discorso si fa più complesso quando ci sono variazioni all'interno della sintassi della frase come avviene nelle traduzioni delle espressioni "to find one's feet" o "to fill sb's shoes":

I'm still finding my feet (= sto ancora orientandomi)

He can't fill John's shoes (= non può sostituire John)

Gli esempi riportati evidenziano la complessità del trattamento degli idiomi nel campo della MT. Molte di queste espressioni richiedono un'analisi sintattica così dettagliata per essere riconosciute che allo stato attuale i sistemi di MT non riescono a garantire risultati accettabili.

43

CONCLUSIONI

Dall'analisi svolta fin qui emerge come ancora ci siano da risolvere molti problemi nell'ambito della MT. Quello che allo stato attuale si può offrire è un validissimo aiuto nel lavoro di traduzione, che resta comunque imprescindibile dalla supervisione dell'uomo.

La ricerca si sta dirigendo verso nuovi argomenti che con buona probabilità diventeranno parte integrante dei futuri sistemi di MT.

Tra i nuovi approcci alla traduzione automatica emergono soprattutto i cosiddetti **approcci empirici**, che utilizzano tecniche di "pattern matching" e basate su statistiche. Con il termine empirico si vuole evidenziare come qualsiasi conoscenza linguistica che il sistema utilizza viene derivata empiricamente, esaminando testi reali, piuttosto che esplicitata da qualche linguista.

In qualsiasi direzione si sviluppi, la ricerca in MT sarà sempre complessa in quanto rappresenta un tentativo di automatizzare una attività che può richiedere l'utilizzo dell'intero campo della conoscenza umana.

44

BIBLIOGRAFIA

-W.J.Hutchins and H.L.Somers – “ An Introduction to Machine Translation”. *Academic Press, London, 1992.*

-A.Spencer – “ Morphological Theory”. *Basil Blackwell, Oxford, 1991.*

-Ronnie Cann – “Formal Semantics”. *Cambridge University Press, Cambridge, 1993.*

Siti Web consultati:

www.essex.ac.uk/linguistics/clmt/MTbook

www.freetranslation.com

www.babelfish.altavista.com/translate.dyn

www.google.com/language_tools

www.systranbox.com

www.synthema.it