

## FRAMENET

Differentemente da WordNet che si concentrava sulla rappresentazione dei significati dei singoli sensi, Framenet ci permette di cogliere il senso di significati complessi. FrameNET è una risorsa lessicale che dispone di un'ampia rappresentazione dei significati dei termini.

I sistemi a frame sono un formalismo di rappresentazione della conoscenza che presenta vari punti di somiglianza con le reti semantiche.

L'intuizione alla base dei frame è che la cognizione umana richiede, per funzionare correttamente, un bagaglio culturale che è già disponibile, ovvero un insieme strutturato di conoscenze per risolvere le ambiguità e per interpretare le diverse situazioni che si trovano a dover affrontare. Questi elementi di conoscenza che sono reperibili a seconda dei vari casi sono dei pacchetti di informazione denominati **frame**. Di fronte alle situazioni in cui ci si imbatte non si parte da zero, al contrario si recupera dalla memoria una rappresentazione a carattere generale (frame) che viene poi raffinata e modificata per rendere conto dei dettagli della situazione corrente. Il frame è proprio questa struttura che viene dalla conoscenza pregressa. Gli esseri umani integrano la conoscenza catturata dagli organi di senso con la conoscenza passata.

FrameNet è un progetto che mira alla costruzione di un lessico per l'inglese e che tende di mettere insieme da una parte i termini e dall'altra i significati. Questa operazione avviene da una parte registrando i modi in cui le varie locuzioni e le varie frasi sono costruite attorno ai singoli termini, dall'altra sulla base della raccolta di corpora che collettivamente forniscono delle evidenze su come i vari termini nei vari sensi possono essere usati nella lingua.

Oltre ad un insieme di strutture che definiscono situazioni, c'è associato una componente che è un corpus annotato.

### Come funziona il lavoro della costruzione di un Frame?

Tipicamente, chi usa frameNet si avvale di porting dei frame in varie lingue, e si trova nella situazione di estenderlo per catturare delle rappresentazioni di conoscenza specifiche nel contesto di uso. Le operazioni che riguardano l'estensione della risorsa sono:

- Quali frame servono e come caratterizzarli? Scriverli in maniera di cogliere le necessità di rappresentazione del contesto d'uso. Caratterizzare un frame significa individuare i vari pezzi di informazione utili, dato un contesto, ovvero i frameElement
- Si trova il vocabolario, ovvero l'insieme di termini che sono richiesti ad esprimere lessicalmente i frame individuati. Quindi si cercano, i sinonimi, iponimi, ed altri elementi lessicali utili
- Si individuano delle frasi tipiche che possano funzionare per studiare ed analizzare i verbi e i loro dipendenti in modo da capire come si legano gli elementi rilevanti del frame

- Annotare gli esempi selezionati. Si tratta di un'operazione condotta largamente da esperti. Consiste nell' individuare frasi esemplificative dei frame per poi annotarle con i nomi dei frame element.
- Variazione della valenza. La valenza esprime nome e numero degli argomenti dei predicati. La valenza ci dice, dato un verbo quali sono le sue possibilità di combinarsi con gli altri elementi, dunque spiega la sua realizzazione sintattica. In questa fase si analizzano numero e tipo di argomenti che possono essere retti da un predicato in modo da cogliere il funzionamento del predicato.

### Uso tipico di frameNet?

Se nel caso di WordNet, l'uso tipico era: dato un termine polisemico e un contesto in cui occorre, cercare di capire il senso di quel termine nel contesto, nel caso di **FrameNet** il task è diverso. Dato un frame, che descrive uno scenario stereotipato (*frame del ristorante*) si tratta di capire come i vari elementi della frase che stiamo analizzando possono essere inseriti nei **frame element** che costituiscono il frame in questione.

Tutti gli elementi che rientrano nel frame sono concetti specifici relativi a quel dato frame. I termini possono evocare diversi frame, essendo polisemici, dopodichè, il lavoro dell'applicazione consiste nell'individuare il frame che meglio si adatta al contesto del termine in questione.

### Polisemia e fenomeni linguistici di FrameNET

In WordNet un termine se ha più di un senso viene associato a diversi synset, nel caso di FrameNet esiste questa nozione di lexical unit.

Una **lexical unit** è una **coppia: termine, senso**. Quello che in WordNET era: sensi diversi appartengono a synset diversi, qui in frameNet sensi diversi appartengono a frame diversi, quindi una lexical unit sarà un'associazione caratteristica di un senso all'interno di uno specifico frame. Quindi, sensi diversi, in questo contesto sono associati a **lexical unit** diverse. Anche in frameNet i sensi sono molto frammentati. Per ogni termine ci possono essere molti frame che individuano diversi significati, così come accade con WordNET. Si tratta di strumenti molto granulati e sottili e consentono di cogliere significati in maniera molto raffinata. Esempio: **verbo replace**

- *'put (sth) back where it belongs':* (rimettere a posto qualcosa nel posto giusto)
- *'occupy a position formerly occupied by (sth,sbd)'* (rimpiazzare qualcuno in un certo ruolo)
- *'put something in a position formerly occupied by (sth,sbd)'* (rimettere qualcosa in una posizione precedentemente occupata da qualcosa da qualcuno)

Il fatto di avere la possibilità di cogliere significati così sottili è stata una novità nell' NLP e in questo casi anzichè puntare a cogliere il significato di una singola parola in una frase,(WordNet) **il tentativo di frameNET è cogliere il significato di una espressione intera.**

Altro tipo di fenomeni linguistici rappresentati nella risorsa sono quelli in cui il soggetto non solo può essere l'agente o il parlante ma assume la posizione di medium, esempio: "l'autore spiega nel capitolo 2" ... "Ti riferivi nella tua lettera"

Altri fenomeni che gestisce frameNet sono i **verbi obliqui** ovvero quando *si dice qualcosa, si pensa che, si ritiene che*, qualcosa che ha a che vedere con delle credenze e pensieri anziché un atto comunicativo diretto, ad esempio: *Il vento spiega il numero di mulini a vento, questi fatti danno ragione alla tua ipotesi... ecc.* In tutti i questi casi è possibile a partire dal verbo ricostruire il significato dell'espressione.

Altro fenomeno linguistico che possiamo cogliere con frameNet è "**l'omissione degli argomenti**". Ad esempio, si può considerare che il senso comune di "give", "dare" e il senso di meno frequente di "contribuire" possono essere separati in quanto differiscono relativamente all'omissibilità dell'argomento. Quindi a domande come:

- "vuoi incontrare il rappresentante della croce rossa? Risposta: "I already gave (io ho già dato, contribuito)"
- "ti ricordi di prendere un regalo per il compleanno di tizio?" Risposta "I already gave (Si gliel'ho già dato)

Quest'ultimo è il senso più frequente, mentre nel primo caso vuol dire ho già dato il mio contributo. Questo meccanismo funziona anche quando l'oggetto è omissivo.

### Come funziona la rappresentazione dei significati?

Essenzialmente l'ipotesi alla base di frameNET è che esistono livelli diversi di analisi della frase con i quali possiamo associare a partire dall'analisi morfologica, per continuare con quella sintattica e semantica, dei ruoli semantici diversi ai vari pezzi della frase ai quali siamo interessati a fornire una rappresentazione semantica. Fare l'analisi sintattica di una frase, per quanto difficile, è un'operazione meno complessa che fare l'interpretazione semantica.

Prendiamo ad esempio il **frame della vendetta**. I tratti significativi del frame della vendetta sono:

- Un individuo A fa qualcosa a un individuo B
- B si vendica su A
- L'azione vendicativa di B è qualcosa di tribale, non istituzionale

**Come si agisce nel tentativo di costruire il frame della vendetta?** E per estensione, come si agisce nella costruzione di un qualsiasi frame, dato un contesto?

- Come prima cosa si fa un elenco dei termini che sono rilevanti per descrivere i contesti in cui può avvenire la vendetta. Quindi si elencano: **sostantivi, locuzioni, aggettivi, avverbi, verbi** e lessicalizzazioni di verbi che significano vendicare, in modo da avere un **dizionario minimo** per ogni componente dei frame.
- Dopo si individuano degli elementi semantici a partire dal **vocabolario** costruito. Tali elementi semantici si chiamano **frame element**. I frame element si riferiscono a dei

ruoli semantici in un dato contesto. Nel caso della vendetta esempi di frame element sono: *“il danno subito, la pena data al vendicatore, l’offeso, il vendicatore”*

- Il processo di interpretazione semantica della frase consiste nell’individuare i costituenti che rappresentano un certo frame element taggandoli in maniera appropriata. Quindi si prendono gli elementi della frase in input e si inseriscono negli opportuni frame element. Se il frame element è costituito da vendicatore, offeso e pena, si estraggono delle porzioni dal testo in input e si inseriscono tali porzioni nei giusti frame element, rispettando i ruoli semantici di un contesto.

Un frame inoltre ha una definizione, che spiega brevemente il significato che cerca di catturare. Esempio con frame della vendetta: *“because of some injury to something-or-someone important to an avenger (maybe himself), the avenger inflicts a punishment on the offender. the offender is the person responsible for the injury.”*

### **Ruoli semantici (frame Element)**

I frame element costituiscono i differenti ruoli semantici che si possono catturare da un dato contesto. FrameNET usa come ruoli semantici dei specifici ruoli nel dato contesto. Ad esempio muovendoci nel contesto della vendetta è molto più informativo sapere di un ruolo di vendicatore o di offeso anziché definire ruoli generici che possono andar bene per più frame, come il ruolo di agente o il goal.

I ruoli semantici non hanno un numero fisso. Possono variare in base al contesto, da un minimo di 5-10 ad un massimo non specificato. Ogni frame è caratterizzato dai suoi frame element.

### **Raccogliere un set di esempi**

Dopo aver definito il frame e aver trovato i termini rilevanti che descrivono i ruoli semantici, la fase successiva è quella di raccogliere un set di esempi che possono essere presi da un corpus. Ciò significa che la descrizione e la risorsa è strettamente legata alle scelte del corpus, come ad esempio il british National Corpus, una collezione di centinaia di milioni di parole annotate. I corpora destinati a questa finalità raccolgono frasi da fonti differenti: giornali, tabloid, twit, articoli scientifici, trascrizioni di conversazioni... BNC è storicamente un esempio importante di corpus.

### **Annotare le frasi**

Definito il vocabolario e i ruoli semantici, raccolte un certo numero di frasi in cui sono ravvisati i frame element definiti, si parte con l’annotazione delle frasi con i nomi dei frame element a cui fanno capo. Si tratta di individuare e selezionare un certo numero di frasi e annotare singoli pezzetti con i ruoli semantici di appartenenza, così una frase che parla del Cid vendicatore avrà in corrispondenza del soggetto il ruolo di vendicatore. Nei corpora, le frasi vengono annotate su diversi livelli linguistici: morfologico, sintattico e semantico.

L'annotazione serve a 2 scopi fondamentali:

- **annotazione dei predicati:** i verbi che evocano i frame creano dei contesti per i filler dell'informazione sulle istanze dei frame. Questo è il momento dell'annotazione per estrarre informazioni e capire chi ha fatto cosa.
- **collocare correttamente elementi** delle frasi dal punto di vista semantico. (ad esempio: El Cid è il pezzo della frase che corrisponde al vendicatore)

### **FrameNet dal punto di vista dell'utente**

L'utente ha una risorsa che contiene frame. Ogni frame element ha associato un certo numero di frasi annotate, quindi a partire dai risultati della ricerca può capire gli usi dei termini ricercati.

E' possibile cercare un significato data una lessicalizzazione, ad esempio quale frame element espresso dall'oggetto avenger?

Altre operazioni sono:

- estrarre descrizioni
- costruire lessici
- utilizzare il corpus annotato all'interno della risorsa come corpus di training su task di apprendimento.

### **FrameNet per realizzare applicazioni**

FrameNet è una risorsa lessicale che mira a fornire, per una discreta porzione del vocabolario inglese, un corpus di frasi annotate sintatticamente e semanticamente. Da questo corpus annotato è possibile estrarre informazioni sui modi in cui i termini vengono usati. E' possibile usare frameNET per realizzare applicazioni nell'ambito della comprensione del testo e nell'ambito della estrazione di informazioni da documenti.

Il caso che consideriamo è tratto da un testo di cronaca nera, perchè le notizie di cronaca sono più semplici da analizzare, ci sono attori, agenti facili da individuare.

Immaginiamo di risolvere un certo numero di problemi collaterali relativi all'analisi sintattica e risoluzione dell'anafora, vorremmo:

- disambiguare i termini del testo
- risolvere alcuni problemi base di composizione semantica, cioè l'integrazione di informazioni associate con i dipendenti semantici in frames evocati dai loro reggenti semantici
- riuscire a scegliere tra possibili analisi sintattiche
- l'attivazione di un vocabolario topic-related per il riconoscimento e la selezione del senso nelle varie parti in cui si articola un testo.

Il testo che andiamo ad analizzare è formato da 5 frasi tratte da un caso di cronaca:

1. Washington (CNN)- Alleged White House gunman Robert Pickett was arraigned Wednesday at a federal court in Washington and ordered held without bond.
2. A federal magistrate informed Pickett of the charges against him — assaulting a federal officer with a deadly weapon, which carries a maximum of ten years in prison.
3. The magistrate set a preliminary hearing for next Tuesday and ordered Pickett held without bond.
4. Pickett, who was shot in the knee by the Secret Service after allegedly firing two shots outside the White House, used crutches to walk into the court.
5. He did not enter a plea.