

Identificazione e disambiguazione della costruzione NPN con BERT

Caso studio di Scivetti e Schneider sulla costruzione inglese

Corso di Semantica a.a. 2025/2026

18 Novembre 2025

Università di Bologna

Indice

La costruzione NPN

Il Dataset

BERT, modello di encoder

Task 1: Identificazione

Task 2: Identificazione (perturbando l'ordine delle parole)

Task 3: Disambiguazione semantica

La costruzione NPN

Cos'è la costruzione NPN

La costruzione NPN è stata ampiamente studiata in inglese da una prospettiva costruzionista, in particolare da Jackendoff (2008) e Sommerer & Baumann (2021).

Construction Schema

Noun₁ Preposition Noun₂

I due nomi nella costruzione devono avere forma identica, anche per la declinazione del numero. La costruzione non ammette nomi accompagnati da determinati.

Cos'è la costruzione NPN

La costruzione può comparire in diverse posizioni sintattiche, ad esempio come modificatore avverbiale o come modificatore nominale.

Esempi

- *I need you to get this word for word.*
(modificatore avverbiale)
- *There is a rebellious quality to your day to day responses which have not gone unnoticed.* (modificatore nominale)

Significato e funzione

Significati delle costruzioni NPN istanziate da *to*

SUCCESSION

Testo del blocco esplicativo.

JUXTAPOSITION

Testo del blocco esplicativo.

II Dataset

Il Dataset

- estrazione da COCA
- Eliminazione dei casi PNPN
- Identificazione dei distrattori
- Annotazione per tutte le istanze delle etichette semantiche

Affidabilità dell'annotazione

Doppia annotazione

- Annotato **25%** del dataset
- Accordo grezzo: **84%**
- Cohen's kappa: **0.754** (accordo forte)

Dimensioni del dataset

- **6599** istanze totali (N-to-N)
- **1885** istanze con doppia annotazione

BERT, modello di encoder

Embedding e trasformazioni del testo

Embedding

Le parole e le frasi vengono trasformate in **vettori numerici** chiamati embedding. Questi vettori catturano somiglianze semantiche e relazioni tra parole, permettendo al modello di "comprendere" il testo.

Perché trasformare il testo

Il testo deve diventare numerico per essere elaborato dai modelli. Trasformare significa codificare ogni parola in uno spazio continuo dove vicinanza = somiglianza.

Transformer e Encoder

Transformer

Modello basato su **self-attention**: ogni parola osserva tutte le altre per capire il contesto. Non usa ricorrenza e permette di gestire sequenze lunghe in parallelo.

Encoder

L'encoder produce rappresentazioni contestuali di ogni parola, che riflettono sia il significato intrinseco sia le relazioni con le altre parole nel testo.

BERT: bidirezionale e contestuale

Cos'è BERT

BERT è un Transformer **solo encoder**, che legge il contesto a sinistra e a destra di ogni parola. Produce embedding contestuali che catturano significato, struttura sintattica e relazioni tra parole.

Addestramento

- **Masked Language Modeling:** predire token mascherati.
- **Next Sentence Prediction:** capire se due frasi sono in sequenza.

Perché usarlo

Le rappresentazioni di BERT possono essere adattate a molti task NLP: classificazione, NER, question answering. Per questo caso studio, mostrano come la semantica delle costruzioni è rappresentata.

Tadsk 1: Identificazione

Dataset

Esempio di blocco

Testo del blocco espositivo.

Task 2: Identificazione (perturbando l'ordine delle parole)

Esempio di blocco

Testo del blocco espositivo.

Task 3: Disambiguazione semantica

Esempio di blocco

Testo del blocco espositivo.