



UNIVERSITÀ
DI PAVIA

UNIVERSITÀ DEGLI STUDI DI PAVIA

DIPARTIMENTO DI STUDI UMANISTICI
CORSO DI LAUREA MAGISTRALE IN LINGUISTICA
TEORICA, APPLICATA E DELLE LINGUE MODERNE

Completamento automatico dell'Oggetto Implicito in italiano:
un'esplorazione con BERT

RELATRICE

Prof.ssa Elisabetta Jezek

CORRELATRICE

Prof.ssa Claudia Roberta Combei

Tesi di Laurea Magistrale di

Agnese Daffara

Matricola n. 502318

Anno Accademico 2022/2023

“Il postino suona [il pianoforte] sempre due volte”

– UmBERTo

Sommario

Introduzione.....	2
Definizione del task.....	2
Il corpus: una nuova risorsa per l'italiano.....	4
Domande di ricerca e struttura della tesi.....	6
Capitolo 1. Studi precedenti.....	7
1.1. Il Complemento Oggetto Implicito.....	7
1.1.1. Classificazione degli Argomenti Impliciti.....	7
1.1.2. Parametri dell'omissione dell'Oggetto.....	9
1.1.3. Omissione dell'Oggetto e classi verbali.....	11
1.2. Trattamento automatico dell'Oggetto Implicito.....	15
1.2.1. I modelli <i>Transformer</i>	17
1.2.2. Capacità e limiti linguistici di BERT.....	19
Capitolo 2. Il corpus.....	21
2.1. Ontologia e classi verbali.....	21
2.2. Raccolta dei dati.....	27
2.3. Annotazione e risultati.....	29
2.3.1. EXPLICIT dataset.....	30
2.3.2. IMPLICIT dataset.....	30
Capitolo 3. Metodologia e design dell'esperimento.....	39
3.1. Il problema del doppio [MASK].....	39
3.2. Script.....	41
3.2.1. Interrogare i modelli.....	42
3.2.2. Calcolare la <i>similarity</i>	45
3.2.3. Analizzare e visualizzare i risultati.....	49
Capitolo 4. Risultati.....	51
4.1. EXPLICIT dataset.....	51
4.2. IMPLICIT dataset.....	55
Capitolo 5. Discussione.....	61
5.1. Panoramica.....	61
5.2. Esplorazione delle classi verbali.....	66
5.2.1. <i>Verbs of Creation and Transformation</i>	66
5.2.2. <i>Verbs of Ingesting</i>	73
5.2.3. <i>Verbs Involving the Body</i>	76
5.2.4. <i>Verbs of Motion</i>	78
5.2.5. <i>Psych-Verbs</i>	80
5.2.6. <i>Verbs of Change of Possession</i>	82
5.2.7. <i>Verbs of Existence</i>	86
5.2.8. Altri verbi.....	88
Conclusioni.....	97
Riferimenti bibliografici.....	99

Sitografia.....	104
APPENDICE A.....	110

Elenco delle sigle e abbreviazioni

cfr.	confronta
DNI	<i>Definite Null Instantiation</i>
es.	esempio
EXP	EXPLICIT dataset
GS	<i>Gold Standard</i>
IAA	<i>Inter-Annotator Agreement</i>
id	numero identificativo della frase nel corpus
IMP	IMPLICIT dataset
INI	<i>Indefinite Null Instantiation</i>
LD	<i>Lexical Defaulting</i>
MWE	<i>Multi-Word Expression</i>
out	output del modello
pag.	pagina
PD	<i>Pragmatic Defaulting</i>
POS	<i>Part of Speech</i>
pp.	pagine
sez.	sezione
sim	<i>cosine similarity</i>
SN	Sintagma Nominale
trad.	traduzione
v.	vedi
vs.	<i>versus</i>

Introduzione

Definizione del task

Quando incontriamo una struttura formata da un verbo e i suoi Argomenti, abbiamo la capacità cognitiva di comprenderne il significato, formando una rappresentazione della situazione nella nostra mente. Anche nei casi in cui uno o più Argomenti siano impliciti, siamo comunque in grado di comprendere il senso complessivo, basandoci sul significato lessicale intrinseco del verbo e sugli indizi contestuali dateci dalle parole vicine. Questa osservazione è in linea con l'ipotesi distribuzionale, proposta in parallelo da Harris (1954) e Firth (1957), secondo cui il significato di una parola viene dato dal contesto in cui si trova. L'ipotesi sta anche alla base di alcune recenti applicazioni computazionali, tra cui l'idea di rappresentare le parole come vettori matematici orientati nello spazio (*Word Embeddings*).

Nel campo della Linguistica Computazionale, è fondamentale che un modello linguistico sia in grado di replicare l'abilità di individuare i partecipanti all'evento quando non sono esplicitati, per affinare task come quelli di *Information Retrieval*, *Question Answering* e *Summarization*, dove è richiesta la comprensione e rielaborazione di informazioni semantiche. Il presente lavoro di tesi si pone su questa linea di studi, focalizzandosi sul completamento automatico del Complemento Oggetto Implicito in italiano. Il progetto consiste nell'organizzare e valutare un esperimento cognitivo su alcuni modelli derivati da BERT (*Bidirectional Encoder Representations from Transformers*) (Devlin et al., 2019).

Per introdurre il task, si osservi l'esempio (1). Sugeriamo che in questa frase il verbo *bere* implichi un Complemento Oggetto, un partecipante semantico all'evento non espresso a livello sintattico. Siccome si tratta di un elemento la cui *Part of Speech* (POS) è Nome (*Noun*), segnaliamo questo Oggetto con [SN], che indica il suo Sintagma Nominale. Assumiamo che la frase possa essere parafrasata, esplicitando un tale Oggetto nella posizione evidenziata senza stravolgerne il senso.

(1) Da quel 26 dicembre non vuole più *bere* [SN] né lavarsi.

Il compito dei modelli è completare automaticamente la testa del SN mancante, restituendo come output una singola parola che possa adattarsi al contesto semantico della frase e riempire quella posizione. Per applicare e valutare il task, sono state annotate manualmente delle frasi con Complemento Oggetto Implicito. Nel caso della frase (1) il candidato ideale è un tipo di liquido, probabilmente *acqua*, se assumiamo che il Soggetto non voglia più

svolgere delle funzioni vitali di base come bere acqua e lavarsi. Completando la frase (1) con *acqua*, si ottiene (2), una frase di senso compiuto.

(2) Da quel 26 dicembre non vuole più *bere* [acqua] né lavarsi.

In questo caso avremmo potuto completare lo spazio vuoto anche con *liquidi*, che indica generalmente la classe di sostanze che si possono bere, assumendo che la persona di cui si parla non voglia ingerire alcun tipo di sostanza liquida. L'annotazione del candidato ideale dipende dai giudizi degli annotatori ed è collegata al tipo di implicazione in atto (v. sez. [1.1.2.](#) e [2.3.2.](#)). Inoltre, la scelta del completamento non è univoca; nella frase (1), per esempio, altri tipi di liquidi o bevande potrebbero essere implicati. Per questo motivo, come metodo di valutazione dei modelli, si è deciso di usare la *cosine similarity*, che misura la similarità semantica tra l'annotazione manuale e l'output, assegnando punteggi da 0 a 1 alle coppie di parole, dove 0 indica la maggiore distanza e 1 significa identità (v. [sez. 3.2.](#)).

A livello metodologico, si propone di trattare l'esperimento come un task di *fill-mask*, chiamato anche *cloze-task*, *fill-in-the-gap* o *masked-word completion*. Questo task consiste nella ricostruzione automatica di una parola precedentemente sostituita con il token [MASK] ed è nato insieme ad una famiglia di modelli linguistici basata sulla tecnologia dei *Transformer*, il cui capostipite è BERT (Devlin et al., 2019). I modelli della famiglia BERT hanno un funzionamento bidirezionale, ovvero, nell'assegnare i pesi alle parole, tengono conto del contesto precedente e successivo. Questi modelli sono allenati proprio sul task di *fill-mask* poiché il *training* prevede il mascheramento di parole randomiche (v. [sez. 1.2.1.](#)).

I cinque modelli scelti per l'esperimento sono allenati sull'italiano e rappresentano le principali famiglie di modelli BERT: BERT base, RoBERTa, DistilBERT e ELECTRA. La scelta ricalca la selezione di Puccinelli et al. (2021); dove erano disponibili la versione *cased* e *uncased*, è stata preferita la prima versione, che distingue tra lettere maiuscole e minuscole. Qui di seguito vengono elencati i modelli scelti:

- BERT-multi, la versione base di BERT allenata su 104 lingue, adottata come *baseline*¹.
- BERT-it, un modello XXL allenato sull'italiano e rilasciato dalla *Bayerische Staatsbibliothek*².
- ELECTRA-it, la versione italiana di ELECTRA, sempre rilasciata dalla *Bayerische Staatsbibliothek*³.
- UmBERTo (Parisi et al., 2020), la versione italiana di RoBERTa⁴.
- BERTino (Muffo e Bertino, 2020), la versione italiana di DistilBERT⁵.

L'esperimento di *masked-word completion* è una variante del *cloze test*. Questo test, nato negli anni '50, inizialmente era utilizzato come indice di leggibilità dei testi e venne poi adottato dalla teoria dell'informazione di Shannon (1948) come criterio per misurare l'entropia e la ridondanza, per essere poi riutilizzato nei test scolastici e, più recentemente, rivisitato in chiave computazionale con la nascita dei modelli *Transformer*. L'idea alla base è che, dove il completamento risulta più facile, il testo sia più ricco di informazioni (Chiari, 2002). Questa idea viene ripresa in questa tesi nel confronto tra i diversi tipi di omissione, in cui il contesto è più o meno ricco (v. [sez. 1.1.1](#)).

A differenza dei *cloze test* tradizionali, nei task computazionali spesso viene oscurato un solo elemento della frase con una precisa funzione sintattica, in modo da osservare il comportamento dei modelli su quella classe di parole. Nel nostro caso, siccome il Complemento Oggetto è implicito, non viene mascherata una parola, ma viene aggiunto il token [MASK] in una posizione originariamente vuota, che assumiamo possa essere riempita con un SN. Il completamento atteso per quella posizione svolge la funzione di Complemento Oggetto e rientra nella classe dei sostantivi⁶.

Il corpus: una nuova risorsa per l'italiano

Per mettere in pratica l'esperimento sui modelli, si è deciso di creare un nuovo corpus italiano di frasi contenenti Oggetti Impliciti. Da un lato, una risorsa di questo tipo può essere utile alla comunità per gli studi sull'omissione del Complemento Oggetto; dall'altro, la costruzione di un corpus consente di avere un dataset fatto su misura per il task e abbastanza esteso da

¹ training: Italian Wikipedia.

² training: OSCAR, OPUS e Italian Wikipedia.

³ training: OSCAR, OPUS e Italian Wikipedia.

⁴ training: UD Italian-ISDT, UD Italian-ParTUT, I-CAB e WIKINER.

⁵ training: Paisà e ItWac.

⁶ La categoria dei sostantivi è considerata "hard to fill" da Nuccorini (2001) rispetto ad altre parti della frase più facilmente prevedibili, come ad esempio le preposizioni. Questa difficoltà, dovuta al range più ampio di possibili candidati, emerge sia nell'annotazione sia nei risultati del modello.

permettere di sviluppare delle osservazioni linguistiche supportate dai dati. Inoltre, l'assenza di risorse linguistiche in italiano annotate con gli Oggetti Impliciti costituisce una ragione per la quale la creazione di un corpus di tale natura rappresenta un contributo necessario alla ricerca su questo argomento.

Come primo passaggio, si è stabilito un elenco di 30 verbi su cui formare il corpus (v. [sez. 2.1](#)). I verbi sono stati selezionati dalla risorsa T-PAS (*Typed Predicate-Argument Structures*) (Jezek et al., 2014). In base alla notazione di T-PAS, questi verbi reggono un Oggetto Opzionale, ovvero ne ammettono sia l'esplicitazione sia l'omissione. Tale presupposto ci permette di teorizzare che, per ciascuna delle frasi, esista effettivamente un Oggetto che può essere esplicitato. Senza questa assunzione, il rischio sarebbe di chiedere ai modelli di riempire uno spazio che non può essere riempito, ottenendo una frase grammaticale.

Sono state raccolte 600 frasi con Oggetto Implicito dal corpus *ItWac reduced* (Baroni e Kilgarriff, 2006), che contiene istanze di vario tipo tratte dal web. In ognuna delle frasi è stata segnalata la presenza di un Oggetto Implicito attraverso il token [MASK], corrispondente alla testa del SN^{7,8}. Così facendo, il task di *fill-mask* è stato riadattato: anziché mascherare una parola esplicita, abbiamo mascherato uno spazio vuoto. In seguito, a lato di ognuna di queste frasi è stata annotata la parola candidata per completare la testa del SN, a cui ci riferiamo con il termine *Gold Standard* (d'ora in poi GS)⁹. Nella frase (1), come è già stato accennato, suggeriamo che la parola GS sia *acqua*. Oltre all'annotazione del Gold Standard, è stato anche annotato il tipo di omissione in atto: *Lexical Defaulting*, in virtù della radice lessicale o *Pragmatic Defaulting*, in virtù del contesto (v. [sez. 1.1.1](#)).

Le 600 frasi annotate con Oggetti Impliciti sono andate a formare la prima parte del nostro corpus, che chiamiamo IMPLICIT dataset. Oltre a questo, è stata costruita una risorsa chiamata EXPLICIT dataset, la quale contiene gli stessi verbi, questa volta accompagnati da Oggetti Espliciti, i quali sono stati semplicemente mascherati e annotati a lato delle frasi. In totale, il corpus è composto da 1.200 frasi divise equamente in due dataset.

La nostra decisione di creare due diversi dataset è motivata dall'idea di osservare le differenze nelle performance dei modelli: funzionano meglio quando l'Oggetto originale è implicito o quando è esplicito? Questa domanda si fonda su un precedente esperimento

⁷ Solo nel caso del modello UmBERTo sono state usate le parentesi uncinate per indicare il token mascherato: <mask>.

⁸ Spesso la frase richiede anche un determinante perché il completamento risulti grammaticale; un problema affrontato aggiungendo un secondo token [MASK]. Per maggiori dettagli si rimanda alla [sezione 3.1](#).

⁹ Il termine viene usato per indicare il completamento ottimale per la posizione mascherata, cioè una parola che rappresenti la categoria semantica richiesta dal contesto e dalla semantica verbale. Usiamo questo termine non per indicare l'unico completamento possibile, ma perché la parola proposta è un'annotazione manuale su cui sarà basata la valutazione, quindi un *Gold Standard*.

condotto da Ye et al. (2020), che assumiamo come esperimento guida; secondo i loro risultati, le prestazioni di un modello migliorerebbero notevolmente se *fine-tuned* su un set in cui l'elemento è implicito, poiché ci sono più informazioni contestuali disponibili. Lo studio di Ye et al. (2020) si concentra sull'omissione di un verbo all'infinito in contesti di metonimia logica in inglese, come per esempio il verbo *drinking* nella frase “she enjoyed [MASK] her coffee” (v. [sez. 1.2.2.](#)). Vorremmo indagare se tali osservazioni possono essere generalizzate al nostro esperimento, pur consapevoli delle differenze tra l'uno e l'altro task, prima fra tutte, la possibilità di inserire un numero maggiore di scelte quando si tratta di un Complemento Oggetto rispetto al caso della metonimia logica, dove la selezione dei possibili verbi è limitata, e in secondo luogo l'influenza maggiore della sintassi nel nostro esperimento. L'ipotesi di partenza è che i modelli funzionino meglio sul dataset implicito.

Nella discussione finale, si intende osservare e discutere i risultati ottenuti dall'esperimento, per fare luce sul funzionamento di BERT in italiano ed elaborare una serie di osservazioni linguistiche, anche quantitative, sull'influenza del tipo di omissione, della classe semantica dei verbi e del tipo/range di Oggetti disponibili. Come già accennato, la metrica che viene usata nella valutazione è la *cosine similarity*, che misura il coseno tra i vettori del GS e dell'output del modello per quantificare la distanza semantica tra i due.

Domande di ricerca e struttura della tesi

Le domande di ricerca della tesi sono le seguenti:

- Ci sono differenze tra il completamento di un Oggetto Implicito e di uno Esplicito?
- Quali modelli producono output più simili al Gold Standard?
- Qual è il ruolo del contesto?
- Su quali classi verbali si ottengono risultati migliori?
- I modelli riconoscono il tipo semantico dell'Oggetto?

La tesi è strutturata in questo modo: nell'introduzione sono stati introdotti il progetto e le domande di ricerca. Nel capitolo 1 viene presentata una panoramica sugli studi precedenti, che riguardano, da un lato, il fenomeno linguistico dell'omissione del Complemento Oggetto ([sez. 1.1.](#)), dall'altro, il trattamento computazionale dell'Oggetto Implicito ([sez. 1.2.](#)). Nella [sezione 1.2.](#) viene descritto il funzionamento dei modelli *Transformer* come BERT e vengono citati alcuni lavori sperimentali che ne discutono le abilità linguistiche. Il capitolo 2 descrive la creazione del corpus, che viene divisa in tre parti: l'Ontologia ([sez. 2.1.](#)), la selezione delle frasi ([sez. 2.2.](#)) e l'annotazione dei due dataset EXPLICIT e IMPLICIT ([sez. 2.3.](#)).

Nel capitolo 3 si descrive la seconda fase del progetto, l'esperimento sui modelli. Viene discusso il problema del doppio [MASK], ovvero la necessità di prevedere un determinante nelle frasi implicite ([sez. 3.1.](#)). Vengono poi presentati gli script usati per applicare il task di *masked word completion* ([sez. 3.2.1.](#)), calcolare la *similarity* ([sez. 3.2.2.](#)) e analizzare e visualizzare i risultati ([sez. 3.2.3.](#)). Nel capitolo 4, vengono presentati i risultati dell'esperimento e le relative analisi statistiche. Infine, nel capitolo 5 si discutono i risultati; nella [sezione 5.1.](#) si riprendono e discutono le domande di ricerca, nella [sezione 5.2.](#) si propone una discussione sui singoli verbi, divisi per classi semantiche, e vengono mostrati esempi tratti dall'IMPLICIT dataset.

Le due fasi del progetto, la costruzione del corpus e l'esperimento di completamento, sono strettamente collegate tra loro, in quanto il corpus viene realizzato su misura per essere usato nell'esperimento. Ci auguriamo tuttavia che possa essere riutilizzato e riadattato a studi futuri sull'omissione dell'Oggetto, sfruttando il ricco materiale linguistico che mette a disposizione. Confidiamo infine che l'esperimento sui modelli BERT possa fare luce sulle loro caratteristiche e sui loro limiti, delineandone meglio il funzionamento linguistico.

Per replicabilità dell'esperimento e rispondendo alla necessità di rendere disponibili in *open access* le nuove risorse alla comunità accademica, tutto il materiale utilizzato, inclusi gli script e il dataset annotato, viene pubblicato in un *repository* su *GitHub* all'indirizzo: https://github.com/agnesedaff/Implicit_obj_completion.

Capitolo 1. Studi precedenti

1.1. Il Complemento Oggetto Implicito

1.1.1. Classificazione degli Argomenti Impliciti

L'omissione degli Argomenti verbali è regolata da molti parametri e può avvenire in modi diversi, a seconda dell'influenza del contesto pragmatico, delle coordinate deittiche e della semantica inerente al verbo. L'individuazione di diversi tipi di Argomenti Impliciti viene tradizionalmente fatta risalire ai lavori di Charles J. Fillmore. In un articolo del 1986, chiamato *Pragmatically Controlled Zero Anaphora*, lo studioso distingue gli Argomenti Impliciti, da lui chiamati *Null Complements* (NC), in due categorie: gli *Indefinite Null Complements* (INC) e i *Definite Null Complements* (DNC).

La distinzione è spiegata in questo modo:

nei DNC l'elemento mancante si può ricavare da qualcosa di dato nel contesto; negli INC l'identità del referente è sconosciuta oppure indifferente

(Fillmore, 1986: 96; trad. mia)

I *Definite Null Complements* possono essere ricostruiti dal contesto perché «l'omissione è autorizzata da un particolare elemento lessicale o una costruzione grammaticale della lingua»; (Fillmore, 1986: 97; trad. mia). Gli *Indefinite Null Complements*, al contrario, non richiamano nessun elemento contestuale, e per questo hanno un'identità «sconosciuta o indifferente» (Fillmore, 1986: 97; trad. mia). Gli INC, inoltre, hanno una doppia lettura, per cui possono essere generici o specifici. Ciò equivale a quanto è stato già discusso per la frase (1) in merito all'alternanza tra *liquidi* e *acqua*.

Generalmente, anziché riferirsi agli Argomenti, si preferisce usare la nozione di *Instantiation*, cioè il meccanismo pragmatico sottostante all'omissione dell'Argomento. Nella teoria di Fillmore, i due tipi di *Instantiation* sono chiamati *Definite Null Instantiation* (DNI) e *Indefinite Null Instantiation* (INI). A questi se ne aggiunge un terzo, chiamato *Free Null Instantiation* (FNI), che si verifica quando c'è un riferimento pragmatico nella situazione comunicativa (Cennamo, 2017: 258).

Un parametro importante per l'omissione è la polisemia dei verbi. Alcuni verbi infatti ammettono l'omissione dell'Argomento solo sotto determinati sensi. Fillmore cita il caso di *lose* 'perdere', che non può implicare l'Oggetto nel suo significato primario di *smarrire* (1b),

ma può farlo quando invece viene usato in un registro sportivo o politico, riferendosi ad una partita, alle elezioni e così via (2b).

- (1) a. He *lost* his wallet / the keys to the office
 Ha perso il suo portafoglio / le chiavi in ufficio
 b. *He *lost*
 **Ha perso*
- (2) a. He *lost* the race / the election / the contest
 Ha perso la gara / le elezioni / il concorso
 b. He *lost*
 Ha perso

(Fillmore, 1986: 100; trad. mia)

Il verbo *perdere* in italiano si comporta allo stesso modo (v. [sez. 5.2.8.](#)) ed è compreso nell'Ontologia del corpus, presentata nella [sez. 2.1.](#) L'osservazione sulla polisemia verbale come parametro dell'omissione di un Argomento dimostra che è fondamentale selezionare i pattern semantici specifici da usare come base ontologica per la raccolta del dataset, perché uno stesso verbo può ammettere l'omissione se usato in un senso e non ammetterla in un altro.

Il registro della frase regola anche l'alternanza tra DNI e INI. Il verbo *mangiare*, per esempio, sia in inglese che in italiano ammette una DNI quando è usato nelle istruzioni con l'imperativo, come nelle ricette (3), mentre solitamente regge una INI. Notiamo che, in italiano, l'Oggetto omesso deve essere inanimato (Cennamo, 2017: 262).

- (3) Phil's Cherry Pie: cool Ø briefly, then *eat* Ø warm
 La torta alle fragole di Phil: raffreddare Ø brevemente, poi *mangiare* Ø calda

(Ruppenhofer e Michaelis, 2010: 164, in Cennamo, 2017; trad. mia)

Un'altra fonte importante per inquadrare l'omissione degli Argomenti verbali è la Teoria del Lessico Generativo di James Pustejovsky (1995). Qui è proposta una classificazione degli Argomenti verbali tripartita: ci sono i *True Arguments*, gli Argomenti espliciti veri e propri, i *Defaulted Arguments*, che sono gli argomenti omessi ma implicati e deducibili, e gli *Shadow Arguments*, che non possono essere esplicitati tranne che in casi particolari. Ai fini del nostro esperimento ci interessano i secondi, gli Argomenti opzionalmente omessi.

Jezek (2018) seguendo la classificazione di Pustejovsky propone di riconoscere solo due tipi di Argomenti: i *True Arguments* e gli *Shadow Arguments*. In aggiunta ci sarebbe poi una strategia pragmatica chiamata *Defaulting* che interessa l'omissione dei *True Arguments*; è proprio la possibilità di implicare semanticamente i *True Arguments* che li differenzia dai circostanziali e li pone in una relazione più stretta con il verbo. Il *Defaulting* può avvenire a livello pragmatico, quando è basato su segnali contestuali, e in questo caso viene chiamato *Pragmatic Defaulting* o *Defaulting Pragmatico*, d'ora in poi indicato con la sigla PD (4), oppure a livello lessicale, quando è autorizzato dal significato intrinseco del verbo, e in questo caso è detto *Lexical Defaulting* o *Defaulting Lessicale*, d'ora in poi LD (5).

(4) Luca iniziò Ø alle 8.30 e finì alle 15. PD

(5) Maria ha mangiato Ø alle 6. LD

(Jezek, 2018: 60-61)

Nel nostro studio e per l'annotazione del dataset abbiamo deciso di adottare questa terminologia. Notiamo che c'è una sovrapposizione, seppur non perfetta, tra DNI + FNI e PD e tra INI e LD. La frase (5) ad esempio mostra un caso di INI o *Lexical Defaulting* con il verbo *mangiare*.

Non sempre la distinzione tra *Lexical* e *Pragmatic Defaulting* è netta, e si ipotizza che esistano dei casi di sovrapposizione. Per esempio, nella frase (6), l'Oggetto Implicito potrebbe essere l'istanza specifica *il piatto preferito di Luca* oppure un'istanza che indica la classe più generica contenuta semanticamente nel verbo, come *cibo* (Jezek, 2018: 62). Nella [sezione 2.3.2](#), si spiega meglio come viene affrontato questo punto nell'annotazione del dataset.

(6) Maria preparò il piatto preferito di Luca e Luca mangiò troppo Ø.

(Jezek, 2018: 62)

1.1.2. Parametri dell'omissione dell'Oggetto

Cennamo (2017) presenta una panoramica comparativa dei principali fattori che influenzano la possibilità di omettere un Complemento Oggetto in varie lingue. L'omissione dell'Oggetto dipende dall'incrocio e dall'alternanza di diversi fattori, in particolare l'azionalità del verbo e la sua struttura aspettuale interna, le caratteristiche tematiche del Soggetto, il tipo di Oggetto, il grado di implicazione che sussiste tra il verbo e l'Oggetto (chiamato *lexical solidarity*), il contesto e il registro della frase (Cennamo, 2017: 251).

La struttura aspettuale dei verbi è il primo parametro discusso nell'articolo. Introduciamo in questa sede le nozioni di *aspetto* e *azionalità* (o *Aktionsart*), due concetti strettamente connessi tra loro e imprescindibili nello studio del comportamento verbale. L'azionalità indica alcune caratteristiche lessicali dei verbi, che permettono di dividerli tradizionalmente in quattro classi: *States*, *Achievements*, *Activities* e *Accomplishments*. Ognuna di esse varia per i tratti [\pm durativo], [\pm telico] e [\pm dinamico] (Vendler, 1957) (v. anche [sez. 2.1.](#))¹⁰. L'aspetto, invece, è una caratteristica grammaticale che indica il punto di vista adottato sulla situazione e descrive la struttura dell'evento, la quale può essere semplice oppure complessa.

Sia in italiano che in inglese, i verbi che indicano eventi complessi, ovvero scomponibili in più sub-eventi (per esempio il risultativo *uccidere*), tendono a non ammettere l'omissione dell'Oggetto, in quanto esso è un partecipante *strutturale* che è necessario esplicitare nella struttura eventiva, piuttosto che un partecipante *radicale* che dipende dalla semantica del verbo e si può quindi omettere. Anche gli *Achievements*, verbi puntuali e non durativi (come ad esempio *rompere*) non ammettono l'omissione dell'Oggetto, tranne per alcuni casi: in pattern con significato metaforico (es. "Marco rompe molto"), in alcuni contesti imperfettivi (es. "Giovanna prima rompe (sc. oggetti) e poi aggiusta") o in pattern idiomatici (es. "Rompiamo se continua così") (Cennamo, 2017: 266).

Può capitare che un verbo cambi comportamento quando cambia lettura aspettuale; per esempio, il verbo *kill* 'uccidere', non può omettere Oggetto se ha una lettura perfettiva (7), perché esso è un partecipante strutturale, codificato al di fuori della radice verbale, mentre lo può omettere quando viene usato con una lettura imperfettiva (8).

- (7) a. John *killed* ten people
 John *ha ucciso* dieci persone
 b. *John *killed*
 *John *ha ucciso* (lettura perfettiva)
- (8) a. John *has killed*, this is why he is in jail (lettura imperfettiva)
 John *ha ucciso*, è per questo che è in prigione.
 b. A tiger only *kills* at night
 Una tigre *uccide* solo di notte

(Goldberg, 2005, in Cennamo, 2017: 30; trad. mia)

¹⁰ A queste se ne aggiungono talvolta altre più particolari, come quella degli *Active Accomplishments*, verbi durativi, dinamici e atelici che possono diventare telici in alcuni contesti, per esempio quando si accompagnano a Oggetti definiti (es. *"Luca ha scritto in due ore" vs. "Luca ha scritto una lettera in due ore").

La spiegazione di questo comportamento, ovvero la possibilità di omettere l'Oggetto solo sotto certe letture aspettuali, potrebbe essere collegata anche a fattori pragmatici e discorsivi, come la salienza dell'Oggetto. Si potrebbe ipotizzare che un Oggetto meno saliente a livello discorsivo, che occupa una posizione meno focale e topicale, sia più candidato ad essere messo in background e ad essere lasciato inespresso. Questo fenomeno infatti è grammaticalizzato e si osserva sistematicamente in alcune lingue, per esempio in Hindi (Cennamo, 2017: 260). Oltre al livello di salienza, i fattori pragmatici e discorsivi includono il registro utilizzato; come è già stato discusso, alcuni linguaggi settoriali (per esempio, le ricette o il linguaggio sportivo) ammettono l'omissione anche per verbi che normalmente non l'ammetterebbero.

I fattori pragmatici si possono incrociare poi con le caratteristiche interne dell'Oggetto: in portoghese brasiliano, per esempio, è possibile omettere un Oggetto in posizione topicale che sia stato precedentemente esplicitato (quindi applicare un meccanismo di DNI) solo quando quest'ultimo è inanimato, oppure, in caso di Oggetti animati, quando l'antecedente è definito (Cennamo, 2017: 261). In italiano, come già menzionato a proposito dell'esempio (3), nelle costruzioni all'imperativo si possono ottenere DNI anche con verbi che solitamente non le ammettono, ma in questo caso l'Oggetto deve essere definito, referenziale e inanimato (Cennamo, 2017: 259).

Infine, anche il ruolo tematico del Soggetto ha un'importanza fondamentale. Sempre in portoghese brasiliano, se il Soggetto agentivo di una frase principale si trova come Oggetto della subordinata, non può essere omesso, mentre può essere omesso quando ha il ruolo di paziente o esperiente e quindi non è agentivo (Cennamo, 2017: 26). In italiano, *uccidere* e *assassinare* sono entrambi verbi risultativi che ammettono anche una lettura imperfettiva, come esemplificato dalle frasi (8a) e (8b), ma, mentre il primo ammette l'omissione dell'Oggetto in contesti imperfettivi (esempio 9), il secondo non lo può fare (esempio 10). Anche qui la differenza potrebbe stare nel grado di agentività del Soggetto: più il Soggetto è agentivo, più obbligatoria è l'esplicitazione dell'Oggetto.

(9) John *ha ucciso*, è per questo che è in prigione (lettura imperfettiva)

(10) *John *ha assassinato*, è per questo che è in prigione (lettura imperfettiva)

1.1.3. Omissione dell'Oggetto e classi verbali

Un'altra prospettiva fondamentale sull'omissione dell'Oggetto si trova nel libro *English Verb Classes and Alternations* di Beth Levin (1993). Il lavoro è diviso in due parti: nella prima, la

studiosa classifica i verbi dal punto di vista delle alternanze sintattiche che esibiscono, nella seconda, propone una classificazione in base alle loro caratteristiche semantiche. Questo libro si concentra sull'inglese ma è utile anche per gli studi in italiano, poiché le classificazioni sintattiche e semantiche sono in buona parte trasferibili da una lingua all'altra.

Nella sezione dedicata alle alternanze sintattiche transitive, vengono passati in rassegna i verbi che alternano tra una forma transitiva e una intransitiva. Un sotto-tipo di alternanza transitiva, chiamata *Unexpressed Object Alternation*, interessa proprio i verbi che possono omettere l'Oggetto e quindi alternano tra una forma transitiva del tipo "NP V NP" e una forma intransitiva del tipo "NP V". Per Levin (1993), la classe ha le seguenti proprietà:

[...] the subject of the transitive use of the verb bears the same semantic relation to the verb as the subject of the intransitive use does. The intransitive variant in each of these alternations involves an unexpressed but understood object.

[...] il soggetto dell'uso transitivo del verbo ha la stessa relazione semantica con il verbo del soggetto nell'uso intransitivo. La variante intransitiva in ciascuna di queste alternanze implica un oggetto non espresso ma sottinteso.

(Levin, 1993: 33; trad. mia)

Questo tipo di alternanza transitiva, a sua volta, comprende delle sotto-classi (figura 1).

- 1.2 Unexpressed Object Alternations
 - 1.2.1 Unspecified Object Alternation
 - 1.2.2 Understood Body-Part Object Alternation
 - 1.2.3 Understood Reflexive Object Alternation
 - 1.2.4 Understood Reciprocal Object Alternation
 - 1.2.5 PRO-arb Object Alternation
 - 1.2.6 Characteristic Property Alternations
 - 1.2.6.1 Characteristic Property of Agent Alternation
 - 1.2.6.2 Characteristic Property of Instrument Alternation
 - 1.2.7 Way Object Alternation
 - 1.2.8 Instructional Imperative

Figura 1. Alternanze transitive che coinvolgono un Oggetto Implicito in Levin (1993).

Tra queste, riportiamo una breve descrizione solo di quelle che sono trasferibili all'italiano e sono utili per spiegare la composizione dell'Ontologia di verbi scelta per il corpus. La prima classe comprende i verbi che esibiscono una *Unspecified Object Alternation*, cioè tutti i verbi

con INI riferita ad un Oggetto generico non specificato. Comprende soprattutto degli *Activities*, come *disegnare* ‘draw’, *bere* ‘drink’, *mangiare* ‘eat’, *cucinare* ‘cook’, *cantare* ‘sing’ e *leggere* ‘read’. L’Oggetto Implicito è «something that qualifies as a typical object of the verb; however, there is some discussion concerning how best to characterize this interpretation» (Levin, 1993: 33). Come specificato anche da Levin, in questa classificazione non sono tenute in conto le DNI con Oggetto specifico, per cui la studiosa rimanda all’articolo di Fillmore (1986). Questi verbi possono alternare tipicamente tra un Oggetto generico, rappresentante la classe semantica, e uno più specifico, come è già stato osservato per *bere* (cfr. introduzione, esempio 1) e *mangiare* (cfr. [sez. 1.1.2.](#), esempio 6).

La *Understood Reciprocal Object Alternation* coinvolge verbi di interazione sociale come *abbracciarsi* ‘hug’, *combattere* ‘fight’/‘battle’/‘war’/‘wrestle’/‘box’ e *giocare* ‘play’. Il tipico Oggetto di questi verbi è *each other* ‘l’un l’altro’ e il Soggetto nella forma intransitiva deve essere un nome collettivo. Nelle lingue romanze e slave il verbo si può accompagnare con la particella riflessiva, come in italiano nella frase “i ragazzi si abbracciano”.

La classe della *PRO-arb Object Alternation* comprende i verbi psicologici come *disgustare* ‘disgust’, *incantare* ‘enchant’ e *preoccupare* ‘preoccupy’ e i verbi di consiglio come *consigliare* ‘advise’ e *avvisare* ‘warn’. Il nome della classe, *PRO-arbitrary Object Alternations*, indica che l’Oggetto Implicito, tipicamente un paziente o esperiente, ha un’interpretazione arbitraria. Se esplicitato nella costruzione transitiva del verbo, tale Oggetto potrebbe corrispondere a *one* ‘uno’, *us* ‘noi’ o *people* ‘le persone’ (Levin, 1993: 38).

Un’altra alternanza transitiva è la *Characteristic Property Alternation*, che può essere di due tipi: nella *Characteristic Property of Agent Alternation*, il verbo indica una propensione del Soggetto agentivo verso un’azione (ad esempio “il cane morde”); nella *Characteristic Property of Instrument Alternation*, il verbo indica una funzione caratteristica del Soggetto strumentale (ad esempio “il coltello taglia”).

Infine, la classe di alternanza chiamata *Instructional Imperative* indica la possibilità di una costruzione intransitiva in cui si omette l’Oggetto quando si usa l’imperativo. L’Oggetto qui viene interpretato in base al contesto, quindi c’è una DNI. Abbiamo già visto che in italiano questa alternanza si può applicare con alcune restrizioni sulle caratteristiche dell’Oggetto, che deve essere definito, referenziale e inanimato (cfr. [sez. 1.1.2.](#), esempio 3).

La classificazione di Levin è stata in parte ripresa da Cennamo (2017), che, in base alle osservazioni comparative avanzate, propone di riconoscere tre tipi di omissione dell’Oggetto in italiano. Per ogni NI elenca i verbi che la esibiscono tipicamente, raggruppandoli per tipi azionali. Nella divisione, la studiosa tiene conto del parametro aspettuale, indicando la lettura

del verbo con cui può avvenire quel tipo di omissione (perfettiva o imperfettiva). I tre tipi di omissione dell'Oggetto individuati sono i seguenti:

1. INI in contesti perfettivi e imperfettivi. Questa omissione si trova con:

- *States* (*vedere, amare*).
- *Activities* (*cucire, studiare, cucinare*).
- *Active accomplishments* (*scrivere, disegnare, dipingere*). Si tratta dei verbi con *Unspecified Object Alternation* in Levin. Quando gli *Activities* sono usati in modo intransitivo, il verbo si può riferire in generale a «general attitudes, abilities, dispositions» (Lo Duca, 2000: 227; Jezek, 2003: 97 in Cennamo, 2017) (es. “Marco dipinge”). Questa classe comprende:
 - a. Verbi a tema incrementale, come i verbi di distruzione e creazione (*mangiare, bere, costruire*), che ammettono tipicamente una doppia lettura (generica o specifica) dell'Oggetto implicito.
 - b. Altri verbi con Oggetto generico, come per esempio i verbi dichiarativi (*borbottare, gridare, sussurrare*)¹¹.

2. INI in contesti imperfettivi. Questa omissione si trova con:

- *Activities* con Oggetto generico (*visitare*).
- *Accomplishments* con Oggetto generico, spesso plurale, umano e con il ruolo tematico di esperiente o paziente. Questa classe comprende:
 - a. Verbi psicologici (*affascinare*) che in Levin appartengono alla classe di verbi con *PRO-arb Object Alternation*, ovvero verbi con Oggetto arbitrario.
 - b. Verbi di cambio di stato (*corrodere*).

¹¹ Inoltre, in Cennamo (2017) si dice che questa classe «shows a clear similarity with cognate object verbs, which take a narrow and fixed range of objects, specified by the semantics of the verb, e.g., *vivere* ‘live’, *dormire* ‘sleep’, *combattere* ‘fight’, *cantare* ‘sing’, etc. (Lo Duca 2000: 228)» (Cennamo, 2017: 264).

3. DNI. Questa omissioni si trova con:

- *Activities* in cui l'Oggetto veicola al suo interno delle informazioni specifiche sull'evento anziché indicare una classe generica. Questo fenomeno è chiamato *lexical solidarity* (Jezek, 2003 in Cennamo, 2007: 264). L'omissione con questi verbi può avvenire solo in contesti letterali.
- *Accomplishments* con un Oggetto animato (es. "la mafia uccide"), la cui omissione dipende dal grado di agentività del Soggetto, oppure casi speciali con un Soggetto strumentale (es. "il coltello taglia"). Questa classe in Levin è indicata come un'alternanza di tipo *Characteristic Property Alternation*, dove il verbo indica una certa caratteristica del Soggetto (che è agentivo o strumentale).

Come abbiamo visto, la classificazione di Cennamo (2017) aggiunge il parametro dell'aspetto. Questo parametro differenzia in particolare il primo tipo di omissione dal secondo tipo, poiché nel secondo caso bisogna per forza avere un contesto imperfettivo per permettere l'omissione. Nel primo tipo di omissione troviamo i verbi che Levin (1993) indicava sotto *Unspecified Object Alternation*, che esibiscono una DNI con Oggetto più o meno specifico. Il secondo tipo di omissione interessa in particolare i verbi a Oggetto arbitrario di Levin (*PRO-arb Object Alternation*), che hanno un Oggetto generico. Il terzo tipo di omissione è una classe più variabile, che può comprendere verbi di tutte le classi azionali e si trova in contesti sia perfettivi che imperfettivi; qui troviamo i verbi con alternanza in base alle caratteristiche del Soggetto (*Characteristic Property Alternation*) di Levin (1993).

1.2. Trattamento automatico dell'Oggetto Implicito

Dal punto di vista della linguistica computazionale, il riconoscimento e completamento degli Argomenti Impliciti è un task sfaccettato che può essere affrontato in molteplici modi ed è collegato alle nozioni di *Semantic Role Labeling*, *Anaphora Resolution*, *Co-Predication*, coerenza e salienza discorsiva. La letteratura sull'argomento si concentra sull'inglese e sembra esserci un gap di ricerca nel trattamento del task in italiano, probabilmente per la mancanza di dati annotati.

La prima fase di ricerca, sviluppatasi prima della nascita dei *Large Language Models*, include una varietà di approcci basati su lessici computazionali annotati semanticamente, come *FrameNet* (Baker, 2014). Quest'ultimo è una risorsa elettronica sviluppata a partire dalle teorie di Fillmore (1982) che raccoglie delle strutture, chiamate *frames*, dove gli eventi

verbalì vengono collegati a ruoli semantici e unità lessicali. I *frames* sono utili per ricostruire gli Argomenti omessi grazie alla ricchezza di informazione semantica che mettono a disposizione. La linea di ricerca più recente, invece, si affida maggiormente a metodi statistici.

Un'importante tappa per il trattamento computazionale degli Argomenti Impliciti è il task 10 proposto a *SemEval 2010* (Ruppenhofer et al., 2009). Questo *shared task* prevedeva, da un lato, di individuare automaticamente i partecipanti semantici all'evento (*Semantic Role Labelling*), dall'altro, di riconoscere le NI, distinguendo tra INI e DNI. Le DNI dovevano poi essere ricondotte ad un referente nel testo (*NI linking*). Per esempio, nella frase (11), il referente è *I* perché il verbo *waiting* regge come Oggetto il sintagma *for me*.

(11) I saw him rejoin his guest, and I crept quietly back to where my companions *were waiting* \varnothing to tell them what I had seen.

L'ho visto riunirsi al suo ospite, e sono tornato furtivamente indietro dove i miei compagni *stavano aspettando* \varnothing per dire loro quello che avevo visto.

(SemEval 2010, task 10; trad. mia)

I partecipanti potevano scegliere di realizzare uno o l'altro task oppure entrambi, e avevano a disposizione dei dati (due racconti di Arthur Conan Doyle) annotati con *FrameNet* e *PropBank*. Per la valutazione, gli annotatori hanno annotato i dati con i ruoli semantici e le anafore. Tutte le proposte hanno adottato come punto di partenza lo schema di annotazione di *FrameNet*, ma solo un gruppo ha scelto di risolvere entrambi i task. Dei due gruppi che hanno risolto il task, uno ha riconosciuto il 63,4% delle NI e l'altro solo l'8%. In conclusione, il task si è rivelato complesso e ulteriori elaborazioni si sono rivelate necessarie (Ruppenhofer et al., 2009).

Anche lo studioso americano Michael Roth si è occupato estensivamente del tema dell'identificazione degli Argomenti Impliciti. Alla conferenza *Starsem 2013* (Roth e Frank, 2013) ha sottolineato che l'omissione di un Argomento è legata alla coerenza testuale e alla salienza (parametro discusso anche in Cennamo (2017), cfr. [sez. 1.1.2.](#)) e ha proposto un metodo di training con dati comparabili in cui strutture esplicite ed implicite sono affiancate parallelamente.

Tra gli altri approcci su questa linea, troviamo Tonelli e Del Monte (2011) in ambito italiano, Gerber e Chai (2010), che si concentrano sui predicati nominali utilizzando la risorsa *NomBank*, e Cheng e Erk (2018), che allenano i modelli fornendo eventi negativi e positivi e

informazioni sulla salienza delle entità all'interno della frase. Più recentemente, il gruppo di James Pustejovsky ha affrontato l'esplicitazione degli Argomenti Impliciti nell'ambito del *Dense Paraphrasing* (Tu et al., 2022), un approccio usato anche per la metonimia logica (v. [sez. 1.2.2.](#)), che prevede di produrre automaticamente una parafrasi dei testi per renderli più comprensibili ad un modello automatico; la parafrasi è *densa* perché vengono anche esplicitati gli Argomenti Impliciti. Il task è motivato dal fatto che, rendendo le frasi più chiare, si possono facilitare altri task più complessi come l'inferenza logica e il *Q&A Answering*.

Sebbene l'aumento dei modelli basati sui *Transformer* abbia apportato miglioramenti significativi a questo compito, come mostrato, ad esempio, in Zhang et al. (2020), il task rimane ancora una questione complicata e stimolante. Come accennato, per quanto riguarda la lingua italiana, abbiamo identificato una potenziale lacuna nella letteratura sull'individuazione e l'elaborazione computazionale degli Argomenti Impliciti, probabilmente a causa della mancanza di corpora annotati progettati per questo compito. È quindi della massima importanza indagare questo argomento e creare nuove risorse computazionali adeguate.

1.2.1. I modelli *Transformer*

I modelli computazionali *Transformer* sono stati introdotti nel giugno 2017 con l'articolo *Attention Is All You Need* (Vaswani et al., 2017) che sottolinea l'importanza dei cosiddetti *attention layers* interni all'architettura delle reti neurali, cioè i pesi nascosti assegnati alle parole ai vari livelli di computazione. In particolare, viene introdotto il meccanismo della *self-attention*, grazie al quale il significato di ogni parola viene ricostruito tenendo in considerazione le parole vicine, che possono avere più o meno influenza su di esso a seconda del loro valore di attenzione all'interno della frase. Di volta in volta, i token vengono trasformati in valori numerici vettoriali più o meno lunghi che rappresentano il contesto in cui si trovano (*Embeddings*) attraverso l'aggiustamento dei pesi a loro assegnati. Il calcolo dei pesi nella frase avviene in parallelo, cioè in modo *feedforward*, a differenza di quanto succede nelle *Recurrent Neural Networks* (RNN), dove si effettua in modo sequenziale con diversi cicli di informazione (Jezek e Sprugnoli, 2023: 103).

Inizialmente questi modelli erano usati per la traduzione automatica, successivamente sono stati adattati ad altri task, tra cui quello di *fill-mask*, anche chiamato *fill-in-the-blank* o *cloze-task*. I modelli linguistici basati sulla tecnologia dei *Transformer* si possono dividere in tre famiglie:

1. BERT (*auto-encoding*)
2. GPT (*auto-regressive*)
3. BART/T5 (*sequence-to-sequence*)

I modelli *Transformer* sono composti da due parti: l'*encoder*, che codifica le informazioni provenienti dal testo, e il *decoder*, che le traduce nell'output. I modelli come BERT (*Bidirectional Encoder Representations from Transformers*) utilizzano solo l'*encoder*, mentre quelli come GPT (*Generative Pre-Trained*) utilizzano il *decoder*. Per questo motivo, i primi sono in grado di concentrarsi meglio sul contesto dell'intera frase, hanno una comprensione globale della sequenza di input e si prestano a task come il *Named Entity Recognition* (NER) o la *Sentiment Analysis*, dove la comprensione dell'input è più importante della generazione di linguaggio, mentre i secondi si concentrano solo sul contesto precedente al token che stanno analizzando e sono più adatti a task di generazione del testo. I modelli di tipo BART/T5 tengono in considerazione entrambe le parti e funzionano bene per task che richiedono sia la comprensione dell'input che la generazione di testo, come la traduzione automatica.¹²

Concentriamoci ora sui modelli BERT usati per questo progetto. Come dice la prima parola dell'acronimo, *Bidirectional*, essi sono modelli bidirezionali, perché tengono in considerazione il contesto precedente e successivo nella frase quando devono codificare i singoli token. La prima uscita di un modello con questa architettura è il checkpoint rilasciato dal team Google nell'ottobre 2018, *bert-base-cased*, allenato con i due obiettivi di *Masked Language Modeling* e *Next Sentence Prediction*, cioè mascherando una percentuale di parole randomiche nei dati di training e successivamente allenando il modello da un lato a ricostruire la frase originale (*Masked Language Modelling*), dall'altro a predire se due frasi con una probabilità del 50% di trovarsi una dopo l'altra sono consecutive (*Next Sentence Prediction*).¹³ Questo primo modello è stato poi seguito da varie versioni, tra cui distilBERT, un modello molto più leggero, con *performance* complessiva del 95%, RoBERTa, in cui si elimina lo step di *Next Sentence Prediction*, ed ELECTRA, che viene allenato non più sostituendo dei token casuali con [MASK], ma sostituendoli con un altro token¹⁴.

Per l'italiano sono stati sviluppati, a partire dalla *baseline* di BERT, vari modelli, tra cui ricordiamo ALBERTo (Polignano et al., 2019), GILBERTo (Ravasio e Di Perna, 2020), UmBERTo (Parisi et al., 2020) e BERTino (Muffo e Bertino, 2020). In questa tesi si è deciso

¹² fonte: introduzione di Hugging Face al NLP.

¹³ fonte: introduzione di Hugging Face a BERT.

¹⁴ fonte: introduzione di Hugging Face a ELECTRA.

di usare UmBERTo, derivato da RoBERTa, che raggiunge le *performance* migliori tra quelli citati (Tamburini, 2020) e BERTino, che rappresenta un'alternativa leggera e meno dispendiosa poiché utilizza come corpora di training solo *ItWac* (lo stesso corpus da cui deriva il nostro dataset) e *Paisà*. Vengono usati poi una *baseline* multilingue e una italiana e infine la versione italiana di ELECTRA, che sono allenati su corpora più grandi (*OPUS*, *OSCAR* e *Wikipedia*) (v. introduzione). Seguendo l'approccio adottato da Puccinelli et al. (2020), siccome questi modelli BERT sono stati addestrati attraverso un task simile al *cloze test*, riteniamo di poterli usare senza un ulteriore *fine-tuning*.

Per l'esperimento è stata utilizzata la libreria *Transformers* di *Hugging Face*, che mette a disposizione la funzione *pipeline* per il task di *fill-mask*, in grado di convertire le parole in input ID e poi decodificarle automaticamente, restituendo i candidati più probabili per la posizione mascherata. Questa lista di candidati è stata sottoposta ad un *POS tagging* attraverso la libreria *SpaCy* e da qui è stato estratto il sostantivo con *prediction score* più alto per ogni frase (per dettagli metodologici, v. [sez. 3.2.](#)).

1.2.2. Capacità e limiti linguistici di BERT

L'articolo *A Primer in BERTology: What We Know About How BERT Works* (Rogers et al., 2020) è una revisione critica aggiornata al 2020 dello stato dell'arte di BERT e include una rassegna delle sue capacità linguistiche e cognitive. Stando a quanto riportato, sembra che BERT abbia sviluppato durante il training, in modo indiretto, delle capacità sintattiche di tipo gerarchico sulla base dei pesi assegnati alle parole, e che tenga in considerazione l'accordo Soggetto-verbo quando deve completare un *cloze test*. Tuttavia, alcuni esperimenti dimostrano che il modello fatica a collegare nodi sintattici distanti tra loro e a riconoscere la negazione.

Dal punto di vista semantico, BERT sembra riconoscere i ruoli tematici ed è molto bravo a trattare gli iperonimi (Ettinger, 2019). Per via della tokenizzazione dei *float*, numeri decimali, il modello ha dimostrato alcuni problemi a riconoscere il segmento da cui ricavare la rappresentazione vettoriale di un numero. Inoltre, il modello ha mostrato una fragilità nella sostituzione di *Named Entities* come nomi di persona e stati.

Per quanto riguarda la conoscenza del mondo, BERT ha difficoltà a trarre delle inferenze pragmatiche dalla situazione e a ricostruire l'evento sulla base dei ruoli tematici, nonché a ricostruire proprietà degli oggetti che vanno dedotte e non sono menzionate, anche nel caso di attributi inerenti ai nomi, come per esempio l'aggettivo *red* nell'espressione *red strawberry* (Apidianaki e Soler, 2021). In generale, la sua conoscenza del mondo sembra più affidata a unità lessicali e *n-grams* ricorrenti che ad un'effettiva capacità deduttiva. Un limite

importante, quindi, è che il modello ricostruisce il significato partendo dal contesto della singola frase, ma facendo affidamento soprattutto sulla frequenza delle collocazioni, con il rischio di cadere in bias cognitivi (Podkorytov et al., 2021). Questo limite è discusso varie volte durante l’osservazione degli output dei modelli nella [sezione 5.2.](#)

Grazie alla disponibilità recente di modelli BERT *open source*, il task di *fill-in-the-gap* o *fill-mask* è stato applicato a una varietà di argomenti linguistici e cognitivi diversi anche in italiano. Tra questi, ricordiamo uno studio sulla capacità di trattare Agentività e Telicità (Lombardi e Lenci, 2021) ed uno riguardante i connettivi (Albertin et al., 2021). Puccinelli et al. (2021) comparano i risultati di BERT in vari tipi di *cloze test* con quelli di studenti universitari, trovando che il modello risolve i test con successo.

Infine, la presente tesi è basata in parte su un esperimento sull’inglese del team di Pustejovsky (Ye et al., 2020) già menzionato nell’introduzione, in cui il task di *fill-mask* è applicato al tema semantico della metonimia logica. Le frasi usate per l’esperimento possono essere parafrasate tramite un task di *Dense Paraphrasing* esplicitando un verbo mancante. Per esempio, “Alex enjoyed [MASK] her coffee” diventa “Alex enjoyed [drinking] her coffee” poiché la frase contiene un verbo (*enjoyed*) che, attraverso una *semantic coercion* sull’Oggetto¹⁵, seleziona un evento (*drinking*). Il modello *BERT-base-uncased* viene *fine-tuned* sullo specifico task e si dimostra un significativo miglioramento della *performance*. In particolare, si osserva che il modello raggiunge risultati migliori quando è allenato sulle frasi implicite, confermando l’ipotesi avanzata dagli autori per cui la maggiore ricchezza contestuale nelle frasi implicite potrebbe avere un impatto positivo sull’addestramento.

¹⁵ Anche il meccanismo della *semantic coercion* (Pustejovsky e Jezek, 2008) viene osservato e discusso nella [sezione 5.2.](#)

Capitolo 2. Il corpus

In questo capitolo si descrive la creazione del corpus. Chiamiamo *Ontologia* l'elenco di verbi e relativi pattern verbali scelti per selezionare le frasi, il quale viene discusso nella [sezione 2.1](#). Nella [sezione 2.2](#), si descrive la raccolta delle frasi. Infine, nella [sezione 2.3](#), viene presentata la metodologia seguita per l'annotazione dei due dataset e vengono discussi i risultati delle annotazioni manuali.

2.1. Ontologia e classi verbali

Come già illustrato nella [sez. 1.1.1.](#), la polisemia dei verbi ha un ruolo fondamentale nella possibilità di implicare il Complemento Oggetto, poiché alcuni significati contestuali dei verbi ammettono l'omissione e altri no. Inoltre, è già stato detto che per l'esperimento sono stati scelti dei verbi che reggono dei Complementi Oggetti Opzionali, intesi come dei *True Arguments* che possono essere implicati oppure possono rimanere espliciti. Sono verbi con alternanza transitiva dell'Oggetto, chiamata *Unexpressed Object Alternation* in Levin (1993).

Il primo passaggio nella costruzione del corpus è stato la costruzione di un elenco di verbi e dei loro relativi pattern semantici con questa proprietà. La risorsa T-PAS (*Typed Predicate-Argument Structures*) (Jezek et al., 2014), un *repository* di strutture verbo-Argomenti sviluppato all'Università di Pavia, ha fornito la base di partenza. Questo *repository* contiene 1.000 verbi, per ciascuno dei quali vengono forniti i significati che può assumere in contesto, che in totale sono più di 5.000. Ogni significato è collegato alla relativa struttura argomentale e contiene anche l'annotazione del tipo semantico degli Argomenti (es. [Beverage], [Food], [Sound]...) e talvolta un *lexical set* di lemmi appartenenti al tipo semantico. Il sistema di tipi semantici si può confrontare sul sito della risorsa. T-PAS inoltre annota gli Oggetti Opzionali tra parentesi tonde. Per un esempio con i primi tre pattern di *bere*, di cui il secondo è metonimico, si veda la figura 1.

1	<p>[Animate] bere ([Beverage {birra caffè tè bibita bevanda aperitivo cocktail liquore vino acqua latte grappino birretta spritz mojito birrozza tisana cappuccino cioccolata whisky vodka rum rhum cognac pozione elisir sangue liquido acqua}])</p> <p>[Animate] ingerisce, assume [Beverage {birra caffè tè bibita bevanda aperitivo cocktail liquore vino acqua latte grappino birretta spritz mojito birrozza tisana cappuccino cioccolata whisky vodka rum rhum cognac pozione elisir sangue liquido acqua}]</p>	≡
1.m	<p>[Human] bere [Container {bicchiere bottiglia}] [Business Enterprise = Producer] [Quantity {sorso goccio}]</p> <p>[Human] ingerisce, assume [Container {bicchiere bottiglia}] (che contiene [Beverage]), [Business Enterprise = Producer] (che produce [Beverage]), [Quantity {sorso goccio}] (che è una porzione di [Beverage])</p>	≡
2	<p>[Human] bere</p> <p>[Human] ingerisce, assume una certa quantità di bevande alcoliche</p>	≡

Figura 1. Esempio di tre pattern del verbo *bere* su T-PAS.

Dai pattern verbali di T-PAS, sono stati isolati quelli indicati come fondamentali dal Nuovo Vocabolario di Base (NVdB) a cura di Tullio De Mauro, che sono 492. Le parole fondamentali coprono circa l'86% delle produzioni in lingua italiana¹⁶; sembra quindi un insieme rappresentativo da cui partire per la costruzione di un dataset. Dai pattern ottenuti sono state scartate le *Multi-Word Expressions* (d'ora in poi MWE), le espressioni fortemente idiomatiche e poco comuni, quelle passive e quelle causative. In seguito sono stati isolati i pattern con Oggetti Opzionali. Nella tabella 1 vengono riportate le quantità di verbi e pattern contenute in T-PAS e nel NVdB, la tabella 2 contiene invece le categorie escluse durante la selezione dei pattern con degli esempi.

¹⁶ La percentuale si riferisce all'italiano parlato del NVdB. Nel caso dei testi scritti prodotti attraverso l'uso del computer, questi vocaboli coprono circa l'83% delle occorrenze.

Risorsa	Categoria	Verbi	Pattern
NVdB	Totale	1.410	-
	Fondamentali	492	-
T-PAS	Totale	1.000	5.392
	Fondamentali	492	-
	Fondamentali con Argomento Opzionale	253	429
	Fondamentali con Oggetto Opzionale	56	63
	Selezione finale	30	37

Tabella 1. Verbi e pattern in T-PAS e NVdB.

Categoria	Esempio
Causativi	[Human] lasciare bollire ([Liquid])
Passivi	[Language] essere scritto (Manner)
Espressioni idiomatiche	[Human] accendere {un cero} (a [Deity])
Multi-Word Expressions	[Anything1] vuol dire ([Proposition]) [Anything2]
Significati meno comuni	[Event] bruciare (a [Human])

Tabella 2. Pattern di T-PAS esclusi durante la costruzione dell'Ontologia.

Il set risultante da questo processo di selezione è di 25 verbi fondamentali, a cui sono stati aggiunti altri 5 verbi citati nella letteratura sugli Oggetti Impliciti (Jezek 2003; Levin, 1993; Cennamo, 2017), che su T-PAS sono stati segnati senza l'annotazione in parentesi tonde dell'Oggetto Opzionale: *cantare*, *disegnare*, *mangiare*, *scrivere* e *suonare*. Questi verbi sono *Active Accomplishments* con una doppia lettura dell'Oggetto Implicito, generica e specifica, e sono descritti nella categoria 1 di DNI in contesti perfettivi e imperfettivi in Cennamo (2017) (v. [sez. 1.1.3](#)). In T-PAS compaiono nel loro significato intransitivo in cui indicano un'attitudine generale, abilità o predisposizione del Soggetto e l'eventuale Oggetto Implicito in questi casi è da interpretare come un rappresentante generico della categoria semantica dell'Argomento, es. *canzone*, *immagine*, *cibo*.

L'elenco finale, revisionato in modo da compattare alcuni pattern sovrapposti e aggiungere tutti i tipi semantici implicati, comprende 30 verbi e 37 pattern, elencati nella tabella 3.

Verbo	Pattern
ascoltare	1. [Human] ascoltare ([Sound] [Musical Performance] [Speech Act]) 2. [Human] ascoltare ([Human2])
attendere	[Human] [Institution] attendere ([Event])
bere	1. [Animate] bere ([Liquid] [Beverage]) 2. [Human] bere ([Alcoholic Drink])
cantare	[Human] cantare ([Musical Composition])
chiamare	[Human1] chiamare ([Human2] [Institution])
combattere	[Human] combattere ([War] [Competition])
condurre	[Human] condurre ([TV Program] [Activity])
consumare	[Human] [Human Group] [Machine] [Device] consumare ([Energy] [Gas] [Inanimate])
correre	[Human = Runner Pilot] correre ([Competition])
cucinare	[Human] cucinare ([Food] [Meal])
dirigere	1. [Human = Director] dirigere ([Musical Performance]) 2. [Human = Director] dirigere ([Movie])
disegnare	[Human] disegnare ([Image] [Physical Entity])
fumare	[Human] fumare ([Drug])
giocare	[Human] [Human Group = Team] giocare ([Competition])
guadagnare	[Human] guadagnare ([Money])
guidare	[Human] guidare ([Road Vehicle])
leggere	[Human] leggere ([Document])
mangiare	[Human] mangiare ([Food] [Meal])
ordinare	1. [Human] ordinare ([Artifact]) 2. [Human] ordinare ([Food] [Beverage] [Meal])
pagare	[Human] pagare ([Abstract Entity])
perdere	[Human] [Human Group] perdere ([Competition] [War])
pregare	1. [Human] pregare ([Deity]) 2. [Human1] pregare ([Human2]) di [Activity]
preoccupare	[Anything] preoccupare [Human]
provare	[Human = Artist] [Human Group = Artist] provare ([Artwork] [Performance])
respirare	[Animate] respirare ([Vapour])
scrivere	[Human] scrivere ([Document])
servire	[Human] servire ([Food] [Beverage] [Meal])
suonare	1. [Human = Artist] suonare ([Musical Instrument]) 2. [Human = Artist] suonare ([Musical Composition]) 3. [Human] suonare ({il campanello} {il citofono})

tirare	[Human = Football Player] tirare ([Ball])
vincere	[Human] [Human Group] vincere ([Competition] [War])

Tabella 3. Ontologia di verbi e relativi pattern semantici scelti per la costruzione del corpus.

Come si può vedere dalla tabella, alcuni verbi comprendono un solo pattern, altri due o più. Il set è eterogeneo dal punto di vista semantico; alcuni verbi sono usati nel loro significato principale, altri corrispondono a significati meno comuni. Per esempio, *condurre* è usato nel senso di *presentare una trasmissione* o *un incontro*; *correre* nel senso di *partecipare a gare di alta velocità*, *perdere* nel significato di *risultare perdente in una competizione*, che a differenza del significato principale di *smarrire qualcosa*, può ammettere l'omissione dell'Oggetto, come discusso nella [sez. 1.1.2.](#); *tirare* è un *achievement* usato nel senso calcistico di *calciare la palla*¹⁷.

Nei prossimi paragrafi vengono proposte tre diverse classificazioni dei verbi dell'Ontologia. La prima segue le classi di alternanza transitiva di Levin (1993); la seconda segue le classi semantiche dei verbi, proposte sempre da Levin (1993) per l'inglese e adattate all'italiano; la terza divide i verbi dell'Ontologia in tipi azionali seguendo la classificazione tradizionale di Vendler (1957) in *States*, *Activities*, *Accomplishments* e *Achievements*. Le classificazioni saranno riprese nel capitolo 5 (in particolare la [sezione 5.1.](#)) per discutere il comportamento dei modelli sulle varie classi e stabilire eventuali tendenze nei risultati.

Per quanto riguarda le classi di alternanza riconosciute da Levin (1993), la maggior parte dei verbi potrebbe essere ricondotta al primo tipo, cioè *Underspecified Object Alternation*. Essa comprende *Activities* che possono esibire una NI con Oggetto generico come rappresentante del tipo semantico (es. *mangiare cibo*) o con Oggetto più specifico (es. *mangiare la cena*) o referenziale (es. *mangiare la minestra di Luca*).

I verbi *giocare* e *combattere* hanno un'alternanza reciproca (*Understood Reciprocal Object Alternation*); tuttavia, il pattern riportato esclude questo tipo di interpretazione, perché il tipo semantico dell'Oggetto di entrambi i verbi è [War] o [Competition], quindi non dovrebbe includere un pronome reciproco come *l'un l'altro*. Il verbo *preoccupare* appartiene alla classe di *PRO-arb Object Alternation*, dove l'Oggetto ha spesso un'interpretazione arbitraria e può corrispondere a *uno*, *noi* o *le persone*. Alla stessa categoria appartiene *pregare* nel secondo pattern riportato, che solitamente si trova nell'uso impersonale *si prega di*. In Cennamo (2017), questi due verbi esibiscono una INI in contesti imperfettivi, con referente generico, spesso plurale e umano.

¹⁷ Le espressioni usate per parafrasare i significati dei pattern in tutta la tesi sono tratte dalla risorsa T-PAS.

Qui di seguito viene proposta una divisione dei verbi in classi semantiche (Levin, 1993):

- *Verbs of Creation and Transformation*: *cucinare* ‘cook’, *dirigere* ‘direct’, *giocare* ‘play’, *suonare* ‘play’, *scrivere* ‘write’, *cantare* ‘sing’, *disegnare* ‘draw’, *provare* ‘try’.
- *Verbs of Ingesting*: *bere* ‘drink’, *mangiare* ‘eat’.
- *Verbs Involving the Body*: *respirare* ‘breathe’, *fumare* ‘smoke’.
- *Verbs of Motion*: *correre* ‘race’, *guidare* ‘drive’.
- *Psych-Verbs (Verbs of Psychological State)*: *pagare* ‘suffer’, *preoccupare* ‘worry’.
- *Verbs of Change of Possession*: *guadagnare* ‘gain’, *ordinare* ‘order’, *vincere* ‘win’, *servire* ‘serve’, *perdere* ‘lose’.
- *Verbs of Existence*: *attendere* ‘wait’, *consumare* ‘consume’.
- *Verbs of Throwing*: *tirare* ‘hit’.
- *Learn Verbs*: *leggere* ‘read’.
- *Verbs of Social Interaction*: *combattere* ‘fight’/‘battle’/‘wrestle’/‘box’.
- *Verbs of Perception*: *ascoltare* ‘listen’.
- *Verbs of desire*: *pregare* ‘pray’.
- *Verbs of Communication*: *chiamare* ‘call’.
- *Verbs with Predicative Complements*: *condurre* ‘host’.

Tutti i verbi sono citati in Levin nel loro significato di riferimento, tranne *fumare*, *provare* (qui usato nel significato di *eseguire le prove di una performance*), *perdere* e *consumare*. Questi verbi sono stati aggiunti rispettivamente alle classi semantiche *Verbs Involving the Body* (*fumare*), *Verbs of Creation and Transformation* (*provare*), *Verbs of Change of Possession* (*perdere*) e *Verbs of Existence* (*consumare*).

Le categorie a volte non descrivono esattamente i pattern verbali scelti per l’Ontologia; nello specifico, il pattern 1 di *pregare*, che coinvolge un Oggetto di tipo [Deity], non è descritto dalla categoria *Verbs of Desire*. Lo stesso vale per il terzo pattern di *suonare*, dove il verbo viene utilizzato in un significato specifico e regge come Oggetto *il campanello* o *il citofono*, e in questo senso non è esattamente un *Performance Verb*. Tuttavia, la categoria serve a dare un’indicazione generale della classe del pattern principale di ciascun verbo. Alcuni verbi poi sono citati in più classi, tra cui è stata scelta la più rappresentativa del pattern rispetto all’utilizzo nel corpus.

Infine, per quanto riguarda la divisione dei verbi nei tipi azionali di Vendler (1957), la classificazione è stata realizzata applicando tre test sintattici: la compatibilità con l’espressione temporale “in X tempo”, la compatibilità con la forma progressiva “stare +

gerundio” e con la forma ingressiva “iniziare a + infinito” (Squartini, 2015). I test sono stati applicati alla forma intransitiva del verbo. Sono riassunti nella tabella 4.

*“in X tempo”		“in X tempo”	
*“stare + gerundio”	“stare + gerundio”	“iniziare a + infinito”	*“iniziare a + infinito”
<i>States</i>	<i>Activities</i>	<i>Accomplishments</i>	<i>Achievements</i>

Tabella 4. Test sintattici utilizzati per determinare il tipo azionale.

Applicando i test, è stata ottenuta la seguente classificazione:

- *States*: consumare, preoccupare.
- *Activities*: ascoltare, attendere, bere, cantare, combattere, condurre, correre, cucinare, dirigere, disegnare, fumare, giocare, guidare, leggere, mangiare, pregare, pagare, provare, respirare, scrivere, suonare.
- *Accomplishments*: guadagnare, ordinare, servire.
- *Achievements*: perdere, vincere, tirare, chiamare.

La maggioranza dei verbi appartiene al tipo *Activity*, ma alcuni di questi possono anche assumere una lettura telica nel contesto frasale e sono quindi meglio descritti come *Active Accomplishments*.

2.2. Raccolta dei dati

Sulla base dei pattern selezionati, sono state raccolte manualmente le frasi che costituiscono il dataset. La grandezza del corpus è stata fissata a 1.200 frasi. Per ogni verbo selezionato sono state selezionate 40 frasi, di cui 20 con Oggetto Implicito e 20 con Oggetto Esplicito. In questo modo, i due dataset IMPLICIT ed EXPLICIT contengono 600 frasi ciascuno (tabella 5).

	IMPLICIT dataset	EXPLICIT dataset	Corpus
Verbi	30	30	30
Frase per verbo	20	20	40
Tot. frasi	600	600	1.200

Tabella 5. Struttura del corpus.

Il corpus da cui sono state estratte le frasi è *ItWAC reduced*, una collezione di frasi dal web preparata al laboratorio di NLP all’università di Masaryk (Repubblica Ceca), il quale è stato a

sua volta ottenuto dallo snellimento del corpus italiano *ItWac* (Baroni e Kilgariff, 2006), un grande database di 1.5 milioni di token taggato con *Part of Speech* (POS) e dipendenze sintattiche. Entrambi i corpora sono interrogabili online tramite *Sketch Engine*. Inizialmente, la scelta è caduta su *ItWac reduced* perché è il corpus di riferimento per T-PAS, dal quale l'intero *repository* di pattern è stato sviluppato. In secondo luogo, essendo tratto dal web, si presta a scopi computazionali, trovando la sua origine proprio in un ambiente virtuale. Un altro vantaggio di questo corpus è che esso è stato parzialmente annotato con il numero T-PAS, che indica il pattern semantico specifico utilizzato nelle frasi. Inizialmente si pensava di usare la porzione annotata come database su cui operare direttamente. Tuttavia, isolando i sensi scelti nell'Ontologia e operando il processo di selezione descritto nella [sezione 2.1](#), si ottengono poche occorrenze, soprattutto per i pattern meno frequenti. Si è deciso quindi di tenerlo come base di partenza, ampliandolo con altri dati da *Sketch Engine*.

Nella raccolta sono state selezionate manualmente delle frasi con ordine sintattico lineare, scartando frasi passive, dislocazioni e frasi scisse. Nel caso del Complemento Oggetto Esplicito, sono state scartate anche le frasi con Oggetto pronominale, perché l'esperimento si concentra soltanto sui Sintagmi Nominali. Per neutralizzare al massimo l'influenza della sintassi, sono stati evitati i Sintagmi Nominali più pesanti, con modificatori come le relative oggettive e riferimenti anaforici precedenti o successivi, oltre alle MWE (con poche eccezioni citate nella [sez. 2.3.1](#), pag. 30). Il range di Oggetti è stato selezionato tenendo in considerazione le collocazioni verbo-Oggetto più frequenti, disponibili su *Sketch Engine* tramite il calcolo della *LogDice* (es. figura 2)¹⁸.

¹⁸ La *LogDice* è una metrica che misura la tipicità di una collocazione. Si considerano le collocazioni con punteggio più alto di nomi post-verbali (PostV_N) nel corpus *ItWac reduced*.

postV_N		
birra	11.1	...
bere birra		
caffè	10.5	...
bere un caffè		
vino	10.5	...
bere vino		
acqua	10.4	...
bere acqua		
bicchiere	10.3	...
bere un bicchiere		
sorso	9.7	...
bere un sorso		
sangue	9.5	...
bere il sangue		

Figura 2. Collocation scores dei primi sette PostV_N (nomi post-verbali) del verbo *bere* nel corpus *ItWac reduced*.

Dopo aver raccolto le 1.200 frasi per il corpus, quelle oltre i 400 caratteri sono state troncate, perché alcuni modelli hanno questo limite di lunghezza per il task di *masked-word*.

2.3. Annotazione e risultati

Secondo la prassi linguistica, un corpus annotato dovrebbe essere (a) riutilizzabile, (b) riproducibile e (c) stabile. Questo include la pubblicazione del materiale in *open source* e, nel caso di un'annotazione manuale, la redazione di linee guida e il calcolo dell'*Inter-Annotator Agreement* (IAA) (Hovy e Lavid, 2010).

Come già accennato, il materiale è stato pubblicato in un *repository* di *GitHub* in *open access* e mette a disposizione della comunità sia il corpus annotato sia gli script utilizzati per l'esperimento. Nelle prossime due sezioni vengono descritte le procedure seguite per annotare i due dataset EXPLICIT e IMPLICIT, con una panoramica del tipo e dell'ampiezza del range di Oggetti Impliciti per ciascun verbo e, per quanto riguarda l'annotazione dell'IMPLICIT dataset, delle osservazioni sulla distribuzione dei due tipi di *Defaulting* e sul range di Oggetti per verbo, utili in vista della discussione dei risultati nel capitolo 5.

2.3.1. EXPLICIT dataset

L'EXPLICIT dataset contiene frasi con un Complemento Oggetto originariamente espresso e successivamente mascherato. L'Oggetto è stato mascherato manualmente sostituendo la testa del Sintagma Nominale con il token [MASK], dove quindi un [MASK] corrisponde ad un singolo Nome. Fanno eccezione 9 occorrenze di Oggetti multiparola che sono stati mascherati interamente perché sono stati considerati unità semantiche indivisibili: *vigili del fuoco* (#95¹⁹), *Il ciclone* (#125), *Affari Tuoi* (#128), *Gran Premio* (#169), *Fiat 520* (#308), *Dirigenti scolastici* (433), *Direzioni Regionali* (#435), *Napoleone III* (#443), *Gruppo di lavoro* (#444).

Le (multi)parole mascherate sono poi state annotate a lato della frase, come esemplificato nella figura 3, che riporta l'annotazione di alcune frasi contenenti il verbo *bere* e gli Oggetti Espliciti *latte*, *alcool*, *acqua* e *vino*.

45	Da circa 3 settimane Federico ha cominciato a bere [MASK] parzialmente scremato alta qualità e sin dai primi giorni ha mostrato di gradire il nuovo alimento.	latte
46	La disposizione potrebbe essere utile anche nelle altre stagioni dell'anno, specie se si vieta ai minori di bere [MASK].	alcool
47	Bevi ogni giorno [MASK] in abbondanza è infatti la quinta regola della sana alimentazione, che invita a bere, ma rigorosamente acqua e non altre bevande, frequentemente e in piccole quantità.	acqua
48	Per il pubblico in generale e per i giovani studenti c'è come al solito il padiglione 9 aperto gratuitamente in cui si può fare shopping, bere del [MASK] e rilassarsi oppure informarsi su ciò che accade nella fiera.	vino

Figura 3. Esempio di annotazione dell'EXPLICIT dataset.

Il range di Oggetti mascherati per ogni verbo è molto vario e va da un minimo di 2 Oggetti per il verbo *tirare* ad un massimo di 20 Oggetti. L'ampiezza del range potrebbe essere un parametro nella riuscita del completamento: più la classe di Oggetti è ridotta, più potrebbe essere facile per un modello trovare un completamento. Questa ipotesi verrà discussa nel dettaglio per l'IMPLICIT dataset e nella discussione dei risultati ([sez. 5.1](#)).

2.3.2. IMPLICIT dataset

Il dataset IMPLICIT contiene 600 frasi con un Complemento Oggetto non esplicitato. Come già detto, assumiamo che per ciascuna di queste frasi ci sia un Oggetto Implicito che si può ricavare dal contesto della frase oppure dalla semantica verbale.

Inizialmente in ognuna di queste frasi sono stati inseriti due token [MASK] che indicano la posizione del Sintagma Nominale Implicito. I token sono stati inseriti in posizione immediatamente post-verbale, tranne in quei casi in cui c'era un avverbio collegato al verbo e

¹⁹ Il cancelletto indica il numero identificativo della frase nel dataset.

l'inserimento di un Complemento Oggetto tra i due elementi avrebbe prodotto una frase agrammaticale. Ciò si verifica soprattutto con il verbo *consumare* e il suo modificatore *meno* (esempio 1)²⁰.

- (1) Hai fatto sicuramente un ottimo acquisto, in particolare per la traina lenta è
l'ideale ed in più oltre a *consumare meno* [MASK] risparmi il motore principale.

L'annotazione dell'IMPLICIT dataset ha richiesto due diverse annotazioni manuali: da un lato il tipo di *Defaulting* (*Lexical* o *Pragmatic*) (v. [sez. 1.1.1.](#)), dall'altro il Complemento Oggetto Gold Standard (GS), che corrisponde alla testa di un Sintagma Nominale, da inserire nella posizione segnalata dal secondo [MASK].

Le due annotazioni sono state eseguite da due annotatori. Il primo annotatore è l'autrice di questa tesi, laureanda in Linguistica. Il secondo annotatore è un laureato in Informatica Umanistica a Pisa con conoscenze equivalenti derivate dal sostenimento di alcuni esami di Linguistica. Gli annotatori hanno svolto la loro annotazione indipendentemente, senza accedere all'altra nel corso dello svolgimento. Sono state impostate delle linee guida consultabili in appendice (Appendice A) che possono essere consultate anche sulla pagina *GitHub* del progetto. Nelle linee guida viene data innanzitutto una definizione di *Lexical* e *Pragmatic Defaulting* tratta da Jezek (2018: 62).

Attraverso quattro passaggi si spiega poi la procedura con cui identificare il tipo di *Defaulting*, indicando prima la presenza di un riferimento anaforico e di indizi contestuali ed extra-frasali come segnali di un *Pragmatic Defaulting*, poi la possibilità di ricavare un'entità specifica dalla semantica lessicale del verbo come indicatore di *Lexical Defaulting*, fornendo infine la possibilità di inserire l'annotazione *unknown* quando non si dia nessuno dei casi precedenti. Questa procedura permette di disambiguare i casi di sovrapposizione tra tipi di *Defaulting*, attraverso un filtro che dà priorità al primo tipo e lascia comunque una terza alternativa nei casi in cui non sia possibile scegliere una risposta univoca. Ricordiamo che si può ipotizzare che i due tipi di omissione possano sovrapporsi (Jezek, 2018: 62).

Per l'annotazione del Complemento Oggetto, è stato prima specificato che questo deve essere un sostantivo singolare o plurale, e che può essere preceduto da un eventuale articolo, che però non andrà annotato. Poi è stata fornita anche in questo caso una scaletta composta da quattro passaggi per identificare la parola da annotare, la quale ricalca la scaletta per l'identificazione del *Defaulting*. Anche qui si passa dall'identificazione di un Oggetto

²⁰ Frasi come (1) sono state inserite nel dataset a causa della difficoltà nel trovare esempi con un Oggetto Implicito.

istanziato dal contesto, poi da quella di un Oggetto più specifico istanziato dal verbo, per arrivare infine alla terza alternativa di annotare *unknown* in contesti ambigui o nel caso in cui non si riesca a trovare un completamento adatto. Questa ultima possibilità ricalca anche il task del modello computazionale, che annota proprio *unknown* quando tra le prime 20 parole più probabili non è presente un Nome (v. [sez. 3.2.](#)).

Per riconoscere un'entità istanziata dal verbo, viene indicato di consultare il pattern semantico, lasciato visibile durante l'annotazione. Ciò permette di fornire un'indicazione generica della categoria semantica dell'Oggetto, la quale viene usata come base per valutare le *performance* dei modelli e la loro capacità di riconoscere una parola appartenente a quella stessa categoria. La figura 5 riporta degli esempi di frasi con il verbo *bere*, annotate dal primo annotatore con l'Oggetto GS *acqua*.

42	Quand'ebbe finito di dargli da bere [MASK], disse: lo ne attingerò anche per i tuoi cammelli finchè abbiano bevuto a sufficienza.	Pragmatic Defaulting ▼	acqua
43	E'quindi della massima importanza adottare opportune contromisure: svolgere una regolare attività fisica, aumentare il consumo di alimenti ricchi di fibre, bere [MASK] regolarmente, evitare i grassi animali privilegiando quelli vegetali da consumare preferibilmente crudi.	Pragmatic Defaulting ▼	acqua
44	La giovinezza, l'ardore, il dolore di Giacinto si trasfondevano in lei: dimenticava i suoi anni, il suo aspetto, la sua assenza; le sembrava d'essere distesa sotto un'acqua limpida nel folto di un bosco e di vedere una figura curvarsi a bere, a bere [MASK], sopra la sua bocca: era Giacinto, ma era anche lei, Noemi viva, assetata d'amore: era uno spirito misterioso che sorbiva tutta l'acqua della sorgente, tutta la vita dalla bocca di lei, tanta sete insaziabile aveva; e si stendeva poi nel cavo	Pragmatic Defaulting ▼	acqua

Figura 5. Esempio di annotazione dell'IMPLICIT dataset (primo annotatore).

Nei prossimi paragrafi si propone una discussione dei risultati dell'annotazione. Per quanto riguarda il tipo di *Defaulting*, le tre possibili annotazioni sono state usate in questo modo: *Lexical Defaulting* è stato annotato 573/1.200 volte e *Pragmatic Defaulting* è stato annotato 627/1.200 volte. La distribuzione dei due tipi è molto bilanciata, con una percentuale di 47,5% vs. 52,25%. L'annotazione *unknown* invece non è mai stata usata per il *Defaulting* né dal primo né dal secondo annotatore. Questo significa che il filtro a tre passaggi proposto nelle linee guida era dettagliato e ha permesso agli annotatori di disambiguare tra un tipo e l'altro. Tuttavia, l'IAA mostra che l'interpretazione personale delle linee guida ha comportato qualche discordanza. Per calcolare l'IAA è stato usato il *Cohen's Kappa*, che misura 0,596. La cifra indica un accordo appena al di sotto della soglia tra moderato e sostanziale nella tabella proposta da Landis e Koch (1977, in Gagliardi, 2018: 2010).

Le annotazioni del tipo di *Defaulting* potrebbero essere indicative delle tendenze dei verbi verso l'uno e l'altro tipo di omissione. Il grafico nella figura 6 riporta l'occorrenza delle etichette nei vari verbi, prendendo in considerazione entrambe le annotazioni.

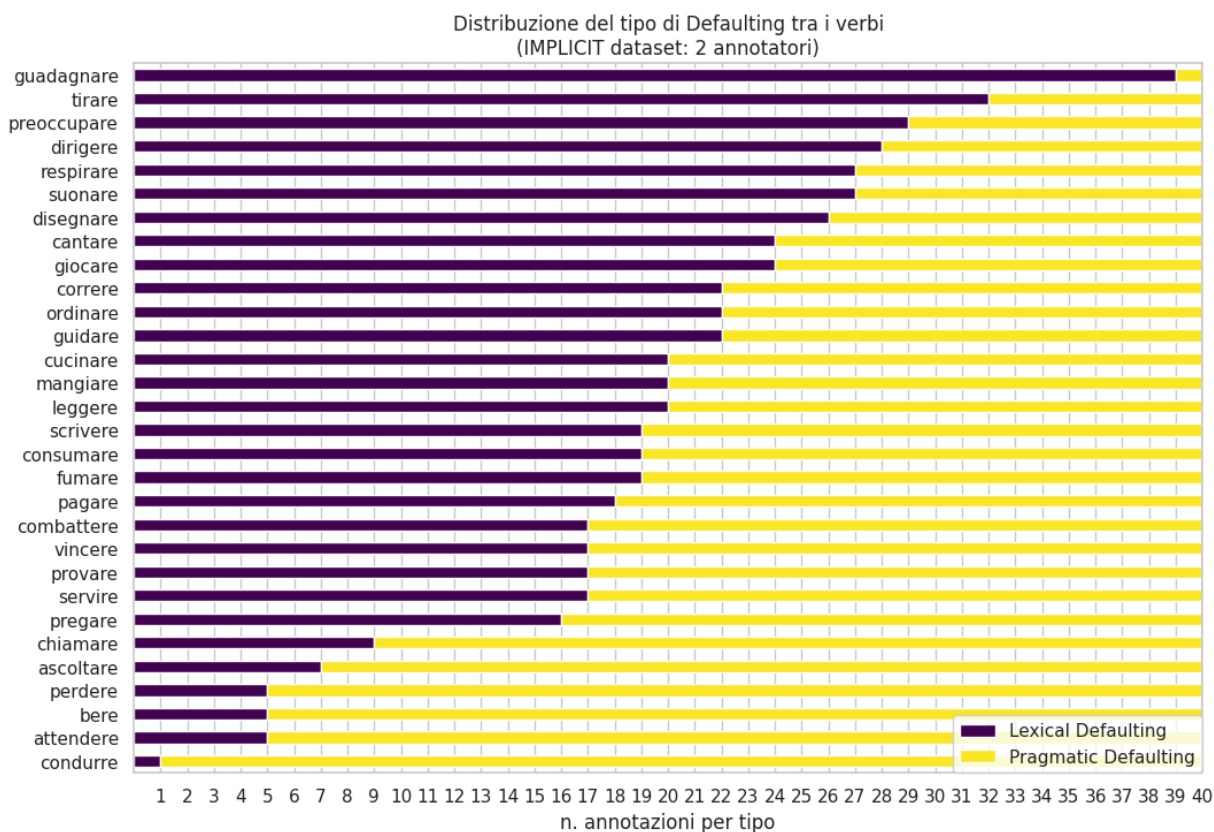


Figura 6. Distribuzione del tipo di *Defaulting* tra i verbi nell'annotazione manuale.

Gli estremi del grafico sono il verbo *guadagnare*, con 39/40 annotazioni di *Lexical Defaulting* e il verbo *condurre*, con 39/40 annotazioni di *Pragmatic Defaulting*. Il primo verbo viene tipicamente usato senza un riferimento contestuale, mentre il secondo si trova in contesti dove l'Oggetto è menzionato subito prima o dopo, come mostrato dai due esempi riportati nella tabella 6.

id	Frase	Defaulting		Oggetto GS	
		1 ²¹	2	1	2
281	Questa domanda presuppone che sia facile, o almeno più facile, <i>guadagnare</i> [MASK][MASK] vendendo software proprietario col solo codice binario (binary-only).	<i>Lexical Defaulting</i>	<i>Lexical Defaulting</i>	soldi	denaro
122	11.20 - FORUM, <i>condurre</i> [MASK][MASK] Claudia Terracina, giornalista	<i>Pragmatic Defaulting</i>	<i>Pragmatic Defaulting</i>	Forum	forum

Tabella 6. Esempi di frasi con *guadagnare* e *condurre*.

²¹ I numeri 1 e 2 indicano il primo e il secondo annotatore nelle tabelle con gli esempi.

Per quanto riguarda l’annotazione del Complemento Oggetto Implicito, in alcuni casi gli annotatori hanno usato delle espressioni multiparola, soprattutto quando riportavano elementi anaforici espressi nel testo. Il primo annotatore 4 volte: *Isola dei famosi*, *figure geometriche*, *veicolo sportivo* e *junk food*; il secondo annotatore 17 volte: *programma TV* (5 annotazioni), *programma radiofonico*, *performance musicale* (4 annotazioni), *porta a porta*, *isola dei famosi*, *Clarinet Choir Amiata*, *pittura rupestre*, *junk food*, *comunità accademica* e *petto di pollo*. In varie frasi, il secondo annotatore ha aggiunto una specificazione dell’annotazione, per esempio, nella frase #345 con il verbo *mangiare* (tabella 7).

id	Frase	Defaulting	Oggetto GS
345	<i>Mangia</i> [MASK][MASK] con appetito, è regolarmente vaccinato e porta il collare scalibor dalla fine di maggio.	<i>Pragmatic Defaulting</i>	cibo (per cani)

Tabella 7. Esempio di annotazione dell’IMPLICIT dataset (secondo annotatore).

Unknown è stato annotato 13 volte, 4/600 dal primo annotatore (*chiamare* #89 #95 #98; *combattere* #109) e 9/600 dal secondo annotatore (*attendere* #28 #30; *chiamare* #89; *combattere* #109; *condurre* #133; *ordinare* #365; *pagare* #383; *perdere* #420; *tirare* #580). In tutti questi casi, l’omissione è stata annotata come *Lexical Defaulting*, ad eccezione delle frasi con il verbo *chiamare*, in cui il primo annotatore ha annotato *Pragmatic Defaulting*.

La difficoltà nell’annotare l’Oggetto di *chiamare* e la confusione tra i tipi di *Defaulting* potrebbe derivare dal fatto che nella maggior parte dei casi la parola omessa con questo verbo è un pronome personale, il quale però non poteva essere annotato secondo le linee guida che sono state fornite, dove si diceva di annotare un sostantivo. Si guardi per esempio la frase #95 (tabella 6), in cui l’Oggetto si riferisce ad una persona esterna che si può indicare con il pronome *lui*. La difficoltà può essere dovuta anche alla mancanza di un contesto precedente in cui probabilmente viene menzionato l’Oggetto (per una discussione sul verbo *chiamare*, si veda anche la [sezione 5.2.8.](#)). Inoltre, per alcune frasi contenenti altri verbi, come *combattere* (frase #109, tabella 8), si potrebbe discutere la presenza effettiva di un Oggetto nominale, poiché il verbo potrebbe implicare un Oggetto reciproco, come anche accennato quando sono state esposte le classi di alternanza di Levin (1993) ([sez. 1.1.3.](#)).

id	Frase	Defaulting		Oggetto GS	
		1	2	1	2
95	in piu <i>ho chiamato</i> [MASK][MASK] e mi ha detto che nn ci sono piu biglietti...	<i>Pragmatic Defaulting</i>	<i>Lexical Defaulting</i>	<i>unknown</i>	istituzione
109	Si tratta di un tale con cui ho dovuto <i>combattere</i> [MASK][MASK] in una piazza locale.	<i>Lexical Defaulting</i>	<i>Lexical Defaulting</i>	<i>unknown</i>	<i>unknown</i>

Tabella 8. Esempio di annotazioni di *unknown* in due frasi dell'IMPLICIT dataset.

Il range di Oggetti va da un minimo di 2 parole nel caso di *guadagnare* ad un massimo di 24 parole con il verbo *attendere* (figura 7). Siccome gli annotatori sono due, il numero massimo possibile di parole nel range è 40.

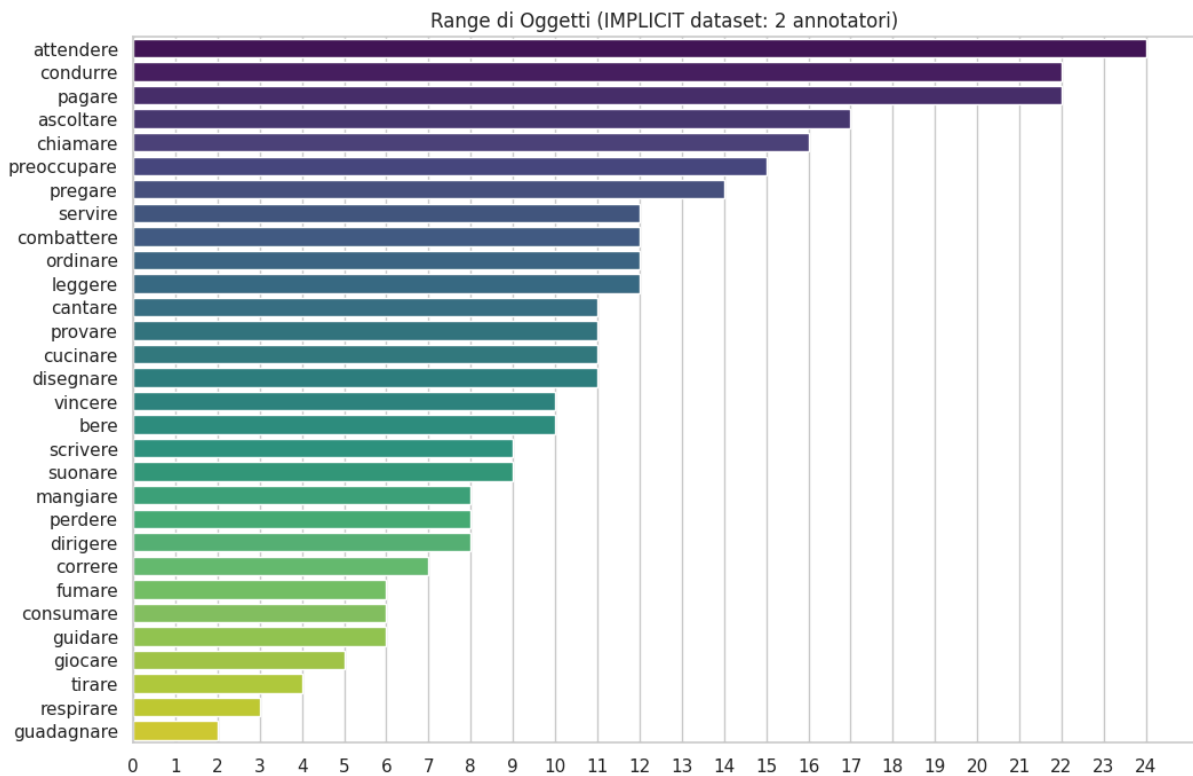


Figura 7. Range di Oggetti annotati per ogni verbo.

Notiamo che il verbo che presenta più volte un'annotazione di *Lexical Defaulting*, *guadagnare*, è anche quello con il range di Oggetti più piccolo. Ciò ha senso poiché l'Oggetto è codificato nella radice lessicale del verbo e quindi è soggetto a meno variazione. Viceversa, i verbi *attendere* e *condurre* sono i verbi con più annotazioni di *Pragmatic Defaulting* e sono tra i primi tre verbi che presentano la classe di Oggetti più ampia, perché l'Oggetto viene ricavato di volta in volta da indizi contestuali.

Gli *Active Accomplishments* con una doppia interpretazione dell'Oggetto come *cucinare*, *mangiare*, *scrivere* e *bere* si trovano a metà per range di Oggetti, perché nell'uso intransitivo alternano tra referenti specifici e generici e sono usati talvolta per indicare attitudini o abitudini generali²². Il verbo che mostra la più alta varietà di Oggetti in entrambi i dataset è *attendere*.

Nella tabella 9 si riassumono i tipi di *Defaulting* e il range di Oggetti annotati per ogni verbo. Come si può osservare, i verbi con un range sotto i 5 elementi sono concentrati nella parte alta della tabella e sono associati con l'annotazione del *Lexical Defaulting*. In particolare, *guadagnare*, *tirare* e *respirare* sono i verbi dove l'Oggetto è più fisso e codificato nella radice. Viceversa, i verbi con un range di Oggetti maggiore di 20 elementi si concentrano nella parte bassa della tabella e sono associati con il *Pragmatic Defaulting*. Tra questi, spiccano *ascoltare*, *attendere* e *condurre*, in cui l'Oggetto sembra più svincolato dalla semantica della radice e più codificato nel contesto pragmatico, di conseguenza più variabile.

²² Tra questi verbi ci sono anche i cinque per cui T-PAS non segna l'Oggetto Opzionale: *cantare*, *cucinare*, *disegnare*, *mangiare*, *suonare*.

	Tipo di Defaulting		
Verbo	<i>Lexical Defaulting</i>	<i>Pragmatic Defaulting</i>	Range di Oggetti GS
guadagnare	39	1	2
tirare	32	8	4
preoccupare	29	11	15
dirigere	28	12	8
respirare	27	13	3
suonare	27	13	9
disegnare	26	14	11
cantare	24	16	11
giocare	24	16	5
correre	22	18	7
guidare	22	18	6
ordinare	22	18	12
cucinare	20	20	11
leggere	20	20	12
mangiare	20	20	8
consumare	19	21	6
fumare	19	21	6
scrivere	19	21	9
pagare	18	22	22
combattere	17	23	12
provare	17	23	11
servire	17	23	12
vincere	17	23	10
pregare	16	24	14
chiamare	9	31	16
ascoltare	7	33	17
attendere	5	35	24
bere	5	35	10
perdere	5	35	8
condurre	1	39	22

Tabella 9. Annotazioni del tipo di *Defaulting* e range di Oggetti GS per ogni verbo.

È stato calcolato il coefficiente di Pearson per verificare la correlazione tra il tipo *Defaulting* e il range di Oggetti (per la metodologia, v. [sez. 3.2.3.](#)). Il risultato, calcolato sul *Pragmatic Defaulting*, è 0,629, *p-value*: 0,000198. Il risultato indica una correlazione positiva tra le due variabili e il *p-value* inferiore a 0,05 indica che questa correlazione è statisticamente significativa. Il risultato del test calcolato sul *Lexical Defaulting* indica gli stessi valori con una correlazione negativa (-0.629). Possiamo affermare allora che, nella nostra annotazione, il numero di annotazioni di *Pragmatic Defaulting* è correlato positivamente al range di Oggetti, e, viceversa, il numero di annotazioni di *Lexical Defaulting* è correlato negativamente al range di Oggetti.

Per il calcolo dell'IAA sull'annotazione del Complemento Oggetto Implicito GS non sono state usate le metriche tradizionali, perché la classe non comprende delle categorie fisse, è potenzialmente infinita e le annotazioni sono interscambiabili. L'unica metrica utilizzabile per questo calcolo è la *cosine similarity*, la stessa usata per valutare gli output dei modelli. Il valore medio di *similarity* tra gli annotatori è 0,72 sulle annotazioni non lemmatizzate e si alza leggermente su quelle lemmatizzate, raggiungendo un accordo sostanziale di 0,735. La possibilità di consultare il pattern verbale e completare l'annotazione con il rappresentante della categoria semantica ha contribuito ad un accordo più alto.

La tabella 10 riassume i punteggi di IAA per l'annotazione del tipo di *Defaulting* e dell'Oggetto GS nell'IMPLICIT dataset.

Annotazione	Metrica di valutazione		Valore
Tipo di <i>Defaulting</i>	<i>Cohen's Kappa</i>		0,596
Oggetto GS	<i>similarity</i>	originale	0,720
		lemmatizzata	0,735

Tabella 10. Valutazione dell'annotazione manuale.

Capitolo 3. Metodologia e design dell'esperimento

3.1. Il problema del doppio [MASK]

Quando si chiede ad un modello di completare una parola mascherata, il modello tiene conto del contesto e cerca di restituire una parola che vi si adatta anche grammaticalmente. Per questo motivo, quando chiediamo di completare uno spazio che potrebbe essere destinato ad un intero sintagma, il modello tende a completarlo con nomi massa singolari oppure nomi plurali, che non devono essere accompagnati da un articolo o un determinante per essere grammaticali. Per avere un sostegno a questa osservazione, prima di scegliere il design dell'esperimento di completamento dell'Oggetto Implicito, è stata fatta una prova su 8 frasi rappresentative create *ad hoc*, utilizzando il modello BERT-it.

La tabella 1 mostra la creazione in *Python* di una lista di frasi su cui operare, chiamata *sentences*, che contiene al suo interno le otto frasi scelte. Si tratta di frasi dove il verbo *bere* ammette l'omissione dell'Oggetto. In tutte le frasi c'è un Oggetto Implicito con un riferimento anaforico precedente, quindi sono casi di DNI. La presenza del riferimento rende il giudizio sull'output dei modelli meno ambiguo.

```
#creo una variabile con le frasi di prova
sentences = ["1) Mi piace bere il vino. Quando bevo [MASK] mi sento bene.",
             "2) Mi piace bere l'acqua. Quando bevo [MASK] mi sento bene.",
             "3) Mi piace bere succhi di frutta. Quando bevo [MASK] mi sento bene.",
             "4) Mi piace bere i succhi di frutta, l'acqua e il vino. Quando bevo [MASK] mi sento bene.",
             "5) Versò la minestra e gliela diede da mangiare. Lui mangiò [MASK] e ringraziò.",
             "6) Versò la minestra e gliela diede da bere. Lui bevve [MASK] e ringraziò.",
             "7) Versò la pozione e gliela diede da bere. Lui bevve [MASK] e ringraziò.",
             "8) Versò la pozione verde e gliela diede da bere. Lui bevve [MASK] verde e ringraziò."]
```

Tabella 1. Frasi selezionate per la prova su come trattare i due [MASK].

Le frasi (1), (2) e (3) contengono come riferimento anaforico dell'Oggetto Implicito rispettivamente un Complemento Oggetto maschile singolare (*vino*), un nome massa (*acqua*) e una MWE plurale (*succhi di frutta*). La frase (4) contiene una lista di elementi (tuttavia casi simili sono stati evitati nella selezione delle frasi per il dataset). Le frasi (5) e (6) contengono una il verbo *mangiare* e l'altra il verbo *bere* con riferimento a *minestra*. Le frasi (7) e (8)

infine hanno un riferimento a *pozione*, rispettivamente senza modificatori e con il modificatore *verde*.

Nella tabella 2 si possono vedere le frasi completate con gli output di BERT-it.

- 1) Mi piace bere il vino. Quando bevo *vino* mi sento bene.
- 2) Mi piace bere l'acqua. Quando bevo *acqua* mi sento bene.
- 3) Mi piace bere succhi di frutta. Quando bevo *succhi* mi sento bene.
- 4) Mi piace bere i succhi di frutta, l'acqua e il vino. Quando bevo *acqua* mi sento bene.
- 4) *unknown*
- 6) Versò la minestra e gliela diede da bere. Lui bevve *acqua* e ringraziò.
- 7) Versò la pozione e gliela diede da bere. Lui bevve *acqua* e ringraziò.
- 8) Versò la pozione verde e gliela diede da bere. Lui bevve *acqua* verde e ringraziò.

Tabella 2. Output di BERT-it sulle otto frasi di prova.

Per le frasi (1), (2) e (3) il risultato è ottimo. Nella frase (4) iniziamo a vedere come il modello sia fortemente influenzato dalle distribuzioni statistiche delle parole dei corpora di addestramento e tenda a scegliere la parola più frequente (*acqua*). Per la frase (5) non viene trovato un completamento²³, perché l'inserimento di *minestra* senza un articolo creerebbe una frase agrammaticale e contemporaneamente il verbo *mangiare* esclude l'inserimento di *acqua* per debolezza di associazione, parola che invece viene inserita nei casi (6), (7) e anche (8), in cui la specificazione attraverso un aggettivo non cambia i risultati.

La conclusione di questo esperimento è che i risultati del modello sono molto limitati dalla necessità di inserire una sola parola al posto di un Sintagma Nominale destinato al Complemento Oggetto Implicito. Inoltre non esiste la possibilità di generare più di una parola al posto di una sola posizione mascherata, perché c'è una corrispondenza 1:1 nel numero di token. Questa limitazione influenzerebbe la riuscita dell'esperimento, di conseguenza l'inserimento di un solo [MASK] non è ottimale.

La soluzione adottata è stata quella di inserire due [MASK], dove il primo va completato con un *Determiner*²⁴ e il secondo con un nome. Sono state proposte e sperimentate tre modalità per realizzare questa soluzione su una selezione di 80 frasi del dataset (corrispondenti ai primi 4 verbi):

²³ Il modello non trova un sostantivo neanche estendendo la ricerca alle 20 *predictions* più probabili (per la metodologia, si rimanda alla [sezione 3.2.1](#)).

²⁴ La libreria utilizzata, *SpaCy*, usa le labels di *Universal Dependencies*, che indicano come *Determiner* vari elementi, tra cui articoli e quantificatori (fonte: <https://universaldependencies.org/u/pos/DET.html>).

- PROVA 1 (“iterazione sul *Determiner*”): fornire una lista di Articoli su cui iterare per completare il primo [MASK], per ogni Articolo selezionare i Nomi restituiti come output, alla fine dell’iterazione prendere il Nome con il punteggio più alto tra tutti quelli ottenuti.
- PROVA 2 (“approccio a due step”): riempire il primo [MASK] con il *Determiner* più probabile (se possibile, altrimenti lasciare vuoto), creare una nuova frase con un solo [MASK] e poi utilizzare la frase per riempire il [MASK] rimanente con il Nome più probabile.
- PROVA 3 (“approccio ridotto”): riempire il secondo [MASK] con il Nome più probabile ignorando il primo [MASK].

Le tre prove sono state sottoposte al modello per vedere quale restituiva un punteggio di *similarity* più alto con l’annotazione manuale realizzata dal primo annotatore. La *similarity* media del modello sulle tre prove è mostrata nella tabella 3. La prova che ha dato i risultati migliori è la prova 2, un approccio a due step in cui prima si riempie prima la posizione del *Determiner* e poi si trova un Nome per la seconda posizione. Pertanto, questo è l’approccio scelto per l’intero esperimento.

Prova	<i>Similarity</i> media
1. “Iterazione sul <i>Determiner</i> ”	0,44
2. “Approccio a due step”	0,45
3. “Approccio ridotto”	0,41

Tabella 3. Punteggi medi di *similarity* ottenuti nelle tre prove per come trattare il doppio [MASK].

3.2. Script

In questa sezione si presentano gli script utilizzati per l’esperimento, focalizzandosi sulle scelte di design che influenzano maggiormente i risultati dei modelli. Gli script sono stati creati in *Python*, sfruttando la piattaforma interattiva *Google Colab*. Il modello preso come esempio a carattere illustrativo è BERT-it. Nella prima sezione vengono presentati gli script per interrogare i modelli e ottenere il completamento delle posizioni mascherate ([sez. 3.2.1.](#)), nella seconda quelli per calcolare la *similarity* tra gli Oggetti GS e gli output dei modelli ([sez.](#)

[3.2.2.](#)), infine sono riportati alcuni script usati per analizzare e visualizzare i risultati ([sez. 3.2.3.](#)). Tutti i riferimenti alle risorse utilizzate sono riportati nella sitografia.

3.2.1. Interrogare i modelli

In questa sezione vengono presentati prima lo script per l'EXPLICIT dataset e poi quello per l'IMPLICIT dataset, che segue la procedura scelta per trattare il doppio [MASK] discussa nella [sezione 3.1.](#)

Come prima cosa, vengono importate la libreria *Transformers* di *Hugging Face* e la funzione *pipeline* di tale libreria. La funzione *pipeline* è un oggetto che contiene al suo interno tutte le *pipelines* del modello e ne facilita l'utilizzo.

Si definisce una variabile *unmasker* in cui si specificano i parametri della funzione, ovvero il task di *fill-mask*, il modello e il tokenizzatore da usare per codificare i dati; in tutti i casi è stato usato il tokenizzatore del modello, seguendo le indicazioni fornite nelle pagine dei singoli modelli su *Hugging Face*.

```
import transformers
from transformers import pipeline

unmasker = pipeline(task = "fill-mask", model =
"dbmdz/bert-base-italian-xxl-cased", tokenizer =
"dbmdz/bert-base-italian-xxl-cased")
```

In seguito si importano altre due librerie. Prima *Pandas*, una libreria per manipolare e analizzare dati, che qui viene usata per aprire e leggere il dataset. Poi *SpaCy*, che serve per effettuare il *POS tagging* degli output e isolare i sostantivi. Per poter usare *SpaCy*, viene caricata la versione *large* della *pipeline* per l'italiano *it_core_news_lg* e incorporata nella variabile *nlp*. Tale *pipeline*, allenata su testi di tipo *news* e *media*, è la più grande disponibile in italiano (541MB). Comprende gli step: *tok2vec*, *morphologizer*, *tagger*, *parser*, *lemmatizer*, *attribute_ruler* e *ner*. Nel suo vocabolario comprende 500.000 vettori a 300 dimensioni ricavati da *FastText*, che verranno usati per calcolare la *cosine similarity* tra le annotazioni GS e gli output dei modelli.

```
import spacy

!python -m spacy download it_core_news_lg
nlp = spacy.load("it_core_news_lg")
```

In seguito viene importato l'EXPLICIT dataset in formato .xlsx da *Google Drive* con *Pandas* e si assegnano alla variabile *sentences* le frasi mascherate.

```
file =  
pd.read_excel(r"/content/drive/MyDrive/IOcompletion_tesi_agnese_daffara/Dataset/  
EXPLICIT_results.xlsx")  
sentences = file["sentence_masked"]
```

A questo punto viene lanciato il modello. Innanzitutto si creano due liste in cui conservare gli output isolati e i *prediction scores* associati, che misurano la probabilità con cui il token restituito riempie la posizione mascherata²⁵. Per ogni frase, si crea una lista chiamata *noun_predictions* in cui verranno inserite tutte le *predictions* del modello con POS = *Noun*. Viene poi avviato l'*unmasker* sulla frase; impostando il parametro *top_k* = 10 si chiede al modello di restituire le prime 10 *predictions*.

```
output = []  
scores = []  
  
for sentence in sentences:  
    noun_predictions = []  
    results = unmasker(sentence, top_k=10)
```

Ogni *prediction* viene restituita nella forma di un dizionario, all'interno del quale sono riportati: 1) il *probability score* (*score*); 2) il numero identificativo del token all'interno del vocabolario del modello (*token*); 3) la stringa corrispondente (*token_str*); 4) la frase completata con il token restituito (*sequence*).

```
{'score': 0.8377352952957153, 'token': 26994, 'token_str': 'alcolici',  
'sequence': "La disposizione potrebbe essere utile anche nelle altre stagioni  
dell'anno, specie se si vieta ai minori di bere alcolici."}
```

Tra le 10 *predictions* ottenute, si isolano tutti i nomi attraverso il *POS tagging* del token (*token_str*) con la libreria *SpaCy*. Gli output con POS = *Noun* vengono aggiunti alla lista *noun_predictions*. Dopo aver ordinato i candidati in base al punteggio, si seleziona quello con il punteggio più alto e lo si aggiunge alla lista *output*; se non viene trovato un Nome, si aggiunge *unknown*. In questo modo si isola per ogni frase, se presente, il sostantivo con la probabilità più alta. Alla fine dell'iterazione sulle frasi, si stampano le due liste con gli output e gli *scores* del modello per l'EXPLICIT dataset.

²⁵ Il *prediction score* non è stato considerato nella discussione dei risultati.


```

for prediction in results:
    doc = nlp(prediction["token_str"])
    for token in doc:
        if token.pos_ == "NOUN":
            noun_predictions.append(prediction)
if noun_predictions:
    noun_predictions.sort(key=lambda x: x["score"], reverse=True)
    output.append(noun_predictions[0]["token_str"])
    scores.append(round(noun_predictions[0]["score"], 2))
else:
    output.append("unknown")
    scores.append("unknown")

print(output)
print(scores)

```

Questo procedimento è stato ripetuto su ogni modello e gli output con i relativi punteggi sono stati copiati in un file riassuntivo comprendente il dataset annotato e i risultati.

Come discusso nella [sezione 3.1](#), per l'IMPLICIT dataset è necessaria una procedura più lunga che comprende due step, ovvero: 1) il completamento del primo [MASK] con un *Determiner* e 2) il completamento del secondo [MASK] con un *Noun*. Si importa l'IMPLICIT dataset, dove tutte le frasi oltre i 400 caratteri sono troncate. Dopo aver importato anche le librerie e le *pipelines*, si interrogano i modelli. Così come per l'EXPLICIT dataset, si creano le due liste *output* e *scores*, che verranno riempite con gli output e i punteggi di probabilità, e si itera su ogni frase del dataset. Si lancia l'*unmasker* con il parametro *top_k* sempre impostato su 10 risultati. Una volta ottenute le prime 10 *predictions* per ogni [MASK], si considerano soltanto quelle riferite al primo [MASK] (*results[0]*). Si applica il *POS tagging* di *SpaCy* per isolare i *Determiners*, che vengono aggiunti alla lista *det_predictions*. Tra questi, si seleziona quello con il punteggio maggiore. A questo punto, viene creata una nuova frase dove il primo [MASK] è sostituito dal *Determiner* appena isolato.

```

output = []
scores = []

for sentence in sentences:
    results = unmasker(sentence, top_k=10)
    det_predictions = []
    for prediction in results[0]:
        doc = nlp(prediction["token_str"])
        for token in doc:
            if token.pos_ == "DET":
                det_predictions.append(prediction)
    if det_predictions:
        det_predictions.sort(key=lambda x: x["score"], reverse=True)
        determiner = det_predictions[0]["token_str"]
        new_sentence = sentence.replace("[MASK]", determiner, 1)
    else:
        new_sentence = sentence.replace("[MASK]", "", 1)

```

Notiamo che la posizione per il *Determiner* nella nuova frase può anche rimanere vuota, nel caso in cui tra le prime 10 *predictions* non ci sia un token con questa POS; in tal caso, il primo [MASK] viene semplicemente eliminato. Questo passaggio dà la possibilità al modello di completare il Sintagma Nominale anche con nomi massa o plurali senza articolo, se la probabilità che si trovi un *Determiner* nella prima posizione è troppo bassa.

A questo punto si procede nello stesso modo illustrato per l'EXPLICIT dataset, cioè si completa la nuova frase ottenuta, che adesso contiene un solo [MASK], con il *Noun* più probabile. L'unica differenza è che qui il parametro *top_k* è alzato a 20, per permettere di trovare eventuali nomi massa o plurali tra i primi 20 risultati.

Riassumendo, per l'EXPLICIT dataset si isola il *Noun* più probabile tra i primi 10 risultati; per l'IMPLICIT dataset si isola il *Determiner* più probabile tra i primi 10 risultati, lo si sostituisce al primo [MASK], ammettendo la possibilità di lasciarlo vuoto, e infine si isola il *Noun* più probabile tra i primi 20 risultati per il secondo [MASK].

3.2.2. Calcolare la *similarity*

In questa sezione si descrivono gli script per calcolare la *similarity* e lemmatizzare i risultati dell'IMPLICIT dataset. Dopo aver importato le librerie *Pandas* e *SpaCy*, si caricano i dati da *Google Drive* assegnandoli alla variabile *file*, come già visto nella [sezione 3.2.1.](#) A carattere illustrativo, viene mostrato qui di seguito il procedimento su un solo modello, mentre nello script caricato nel *repository* di *GitHub* si itera su tutti i modelli.

Iniziamo dal calcolo della *similarity* sui risultati dell'EXPLICIT dataset. La variabile *output* contiene gli output del modello BERT-it, la variabile *predicted* contiene gli Oggetti GS. Le due liste vengono unite con la funzione *zip(predicted, output)*.

```
output = [file["BERT-italian-xxl"]  
predicted = file["GS_Obj"]  
  
couples = zip(predicted, output)
```

Viene definita una funzione chiamata *calculate_similarity* per calcolare la *similarity* tra le coppie di parole con *SpaCy*. La funzione prima di tutto lancia la *pipeline* di *SpaCy* sui due token. Se l'output del modello è *unknown*, assegna automaticamente un punteggio di *similarity* uguale a 0; questo permette di eliminare eventuali punteggi negativi quando il risultato è *unknown*. Viene poi applicato un filtro per verificare se entrambi i token hanno un vettore nel vocabolario di *SpaCy*, con la funzione *has_vector*. Se la funzione *has_vector* dà un risultato positivo, viene restituito il punteggio di *similarity* per la coppia di parole arrotondato a due decimali; altrimenti, viene restituito 0, anche in questo caso per limitare il numero di punteggi negativi.

```
def calculate_similarity(word1, word2):  
    word1_token = nlp(word1)  
    word2_token = nlp(word2)  
    if word2 == "unknown":  
        return 0  
    if word1_token.has_vector and word2_token.has_vector:  
        similarity = word1_token.similarity(word2_token)  
        return similarity  
    else:  
        return 0
```

Dopo aver definito la funzione, la si applica ad ogni coppia e si stampano i risultati, arrotondati a due decimali, ottenendo così i punteggi di *similarity* del modello per ciascun completamento nell'EXPLICIT dataset.

```
for couple in couples:  
    similarity = calculate_similarity(couple[0], couple[1])  
    print(round(similarity,2))
```

Per quanto riguarda l'IMPLICIT dataset, prima di calcolare la *similarity* tra gli Oggetti GS e gli output del modello, sono state compiute delle operazioni per preparare i dati. Prima di tutto, dalle annotazioni del secondo annotatore sono state tagliate manualmente le

informazioni aggiuntive, per facilitare il calcolo della *similarity* e la lemmatizzazione (v. esempio nella tabella 7, [sez. 2.3.2.](#)).

In seguito, sia le annotazioni che gli output dei modelli sono stati lemmatizzati con la libreria *SpaCy*. Si è scelto di lemmatizzare i risultati sull'IMPLICIT dataset perché nei contesti impliciti non sono presenti indizi sintattici sul Numero del completamento e spesso entrambe le forme sono logicamente e semanticamente accettabili. Inoltre, la *similarity* tra il singolare e il plurale dello stesso lemma è minore di 1 e varia in base al lemma, come dimostrano gli esempi riportati nella tabella 4.

id	Frase	Oggetto GS		
		1	2	sim
64	Infine elogio i ragazzi della nord che <i>cantano</i> [MASK][MASK] incitano ma sono spesso poco seguiti (in nord ci sono molte persone in età avanzata che non hanno voglia di cantare ma da loro è capibile) e mi ritorna in mente una lettera di un ragazzo che si lamentava che i biglietti della Sud per il derby fossero dati anche a quelli della Nord, dicendo che la Sud è roba loro.	cori	coro	0,62
74	A cinque anni, <i>cantava</i> [MASK][MASK] a due voci con il nonno, direttore del coro di Monfalcone, e a undici scriveva la sua prima canzone.	canzoni	canzone	0,7

Tabella 4. Esempio di punteggio di *similarity* non lemmatizzato sull'IMPLICIT dataset.

Di conseguenza, si è deciso di trattare l'annotazione di un token al singolare e al plurale allo stesso modo, applicando una lemmatizzazione alle annotazioni manuali e ai risultati dei modelli. Per fare questo, è stata definita la funzione *lemmatize_text*. Tale funzione avvia la *pipeline* di *SpaCy* sulle parole, trova il lemma di ogni token nel testo e poi unisce i lemmi per restituire il testo lemmatizzato. Tale passaggio è necessario perché le annotazioni manuali dell'IMPLICIT dataset contengono talvolta delle MWE, che la *pipeline* di *SpaCy* non può lemmatizzare, in quanto può applicare la lemmatizzazione solo ai singoli token (per un elenco di tali MWE, v. [sez. 2.3.2.](#), pag. 34). Il problema è stato affrontato lemmatizzando i token e unendo i lemmi; per esempio, la MWE *isola dei famosi* è stata lemmatizzata in *isola di il famoso*. Non è una soluzione ideale ma comunque non risulta troppo intrusiva, visto il basso numero di MWE nelle annotazioni manuali e considerato che alcune sono state lemmatizzate correttamente, per esempio *programma radiofonico* > *programma radiofonico*.

```
def lemmatize_text(text):
    doc = nlp(text)
    lemmatized_text = " ".join([token.lemma_ for token in doc])
    return lemmatized_text
```

La funzione è stata applicata a tutte le annotazioni manuali e a tutti gli output dei modelli prima di calcolare la *similarity*. I risultati lemmatizzati sono stati poi assegnati alle variabili *predicted_lemmatized1* per il primo annotatore, *predicted_lemmatized2* per il secondo e *output_lemmatized* per il modello.

```
predicted_lemmatized1 = []
for i in predicted1:
    predicted_lemmatized1.append(lemmatize_text(i))
predicted_lemmatized2 = []
for i in predicted2:
    predicted_lemmatized2.append(lemmatize_text(i))
output_lemmatized = []
for i in output:
    output_lemmatized.append(lemmatize_text(i))
```

Il calcolo della *similarity* è stato effettuato in modo analogo a quanto visto per l'EXPLICIT dataset, con poche variazioni. La prima differenza è che nella definizione della funzione *calculate_similarity* è stato aggiunto un filtro che restituisce 0 anche nel caso in cui la parola della prima lista, cioè l'annotazione manuale, sia *unknown*; anche qui, questo passaggio è necessario per evitare i punteggi di *similarity* negativi. L'altra differenza è che in questo caso, per ogni completamento, prima viene calcolata la *similarity* con il primo annotatore, poi con il secondo annotatore, e infine viene calcolata una media tra questi due punteggi. Il metodo usato per trattare in parallelo le colonne del dataset è sempre la funzione *zip*. La media è calcolata usando la funzione *np.mean* della libreria *NumPy*.

```
similarity_scores = []
for pred1, pred2, out in zip(predicted_lemmatized1, predicted_lemmatized2,
                              lemmatized):
    similarity1 = calculate_similarity(pred1, out)
    similarity2 = calculate_similarity(pred2, out)
    mean_similarity = np.mean([similarity1, similarity2])
    similarity_scores.append(round(mean_similarity, 2))
print(similarity_scores)
```

Con questo procedimento di calcolo, se il modello restituisce un output simile ad una delle due annotazioni manuali ma diverso dall'altra, il punteggio ottenuto è intermedio. Notiamo che, se il completamento è uguale ad una sola delle annotazioni manuali, il punteggio è

inferiore a 1, come esemplifica la frase #70, riportata nella tabella 5. Quando invece la *similarity* è uguale a 1, indica una completa identità tra le due annotazioni manuali e l’output del modello (per il numero di risultati uguali a 1 per modello, cfr. tabella 8, [sez. 4.2.](#)).

id	Frase	Oggetto GS			BERT-it		UmBERTo		BERTino	
		1	2	sim	out	sim	out	sim	out	sim
70	O, se è stata, chissà quando, chissà dove: al tempo di Gilgamesh, magari, quando sulle mura di Ninive qualcuno <i>cantava</i> [MASK][MASK] alla luna convinto di aver liberato il mondo.	inno	canzone	0,5	canto	0,47	inno	0,75	inno	0,75

Tabella 5. Esempio di punteggi di *similarity* sull’IMPLICIT dataset.

3.2.3. Analizzare e visualizzare i risultati

In questa sezione verranno esaminate le metodologie utilizzate per analizzare i risultati, con riferimenti alle librerie impiegate durante l’analisi. I risultati ottenuti con queste metodologie sono presentati e discussi in parte nel capitolo 2 ([sez. 2.3.2.](#)) e in parte nei capitoli 5 e 6.

L’IAA tra annotatori sull’IMPLICIT dataset è stato calcolato nello stesso modo già visto per calcolare la *similarity* tra le annotazioni manuali e i risultati dei modelli ([sez. 3.2.2.](#)). Anche in questo caso, se almeno uno dei due token è *unknown*, la *similarity* restituita è automaticamente uguale a 0. Ciò significa che, se almeno uno dei due annotatori ha annotato *unknown*, la *similarity* tra annotatori è 0; tuttavia, la *similarity* del modello con gli annotatori potrebbe essere diversa da 0, in quanto consiste in una media tra le due *similarity*. L’IAA sul tipo di *Defaulting* invece è stato calcolato usando la funzione *cohen_kappa_score* del modulo *sklearn.metrics* della libreria *Scikit-learn*.

```
from sklearn.metrics import cohen_kappa_score

def1 = file["defaulting_1"]
def2 = file["defaulting_2"]

mapping = {"Lexical Defaulting": 0, "Pragmatic Defaulting": 1}
def1n = [mapping[a] for a in def1]
def2n = [mapping[a] for a in def2]

print(cohen_kappa_score(def1n, def2n))
```

Tutti i grafici sono stati creati con la libreria *Seaborn* e il modulo *matplotlib.pyplot* di *matplotlib*, dopo averli organizzati all’interno di un dataframe di *Pandas*. In particolare, la

distribuzione dei tipi di *Defaulting* è stata visualizzata attraverso un grafico a barre impilato (v. figura 6, [sez. 2.3.2.](#)).

Passando all'analisi dei risultati dei modelli, le metriche di valutazione classiche (*precision*, *recall* e *F-measure*) sui risultati per l'EXPLICIT e l'IMPLICIT dataset sono state calcolate con la funzione *precision_recall_fscore_support* del modulo *sklearn.metrics* di *Scikit-learn*, mentre le statistiche descrittive complete dei punteggi di *similarity* sono state calcolate utilizzando la libreria *NumPy* e il modulo *scipy.stats* di *SciPy*. Le medie di *similarity* di ogni modello sono state calcolate facendo una media dei punteggi ottenuti su ogni frase del dataset (già arrotondati a due decimali) con la funzione *np.mean* di *NumPy*.

Anche il coefficiente di correlazione di Pearson è stato calcolato utilizzando il modulo *scipy.stats*, con la funzione *pearsonr*. Tale metrica è stata usata per calcolare la correlazione tra il tipo di *Defaulting* e il range di Oggetti nell'annotazione manuale dell'IMPLICIT dataset (v. [sez. 2.3.2.](#)) e per verificare l'effetto del tipo di *Defaulting* e del range di Oggetti sul punteggio di *similarity* (v. [sez. 5.1.](#)).

Invece, per calcolare l'effetto delle classi verbali sul punteggio di *similarity*, è stato calcolato il test *oneway ANOVA*, utilizzando la funzione *f_oneway*, sempre appartenente al modulo *scipy.stats* di *SciPy*. Successivamente, nel caso di correlazione statisticamente significativa, per verificare la differenza di punteggio tra le classi è stato usato il test di Tukey HSD (*Honestly Significant Difference*), che compara le medie dei gruppi. Il test è stato calcolato con la funzione *pairwise_tukeyhsd* del modulo *statsmodels.stats.multicomp* della libreria *statsmodel*.

Capitolo 4. Risultati

In questo capitolo vengono riportati i risultati dei modelli, calcolati sia con le metriche di valutazione *precision*, *recall* e *F-measure*, sia con la *cosine similarity*. Per ogni modello si propone anche una statistica descrittiva dei 600 punteggi di *similarity* ottenuti nel dataset in analisi. Nel caso dell'IMPLICIT dataset, tutti i risultati sono stati calcolati dopo la lemmatizzazione del GS e degli output dei modelli (v. [sez. 3.2.2.](#)). Inoltre, vengono presentati i risultati medi totali dei modelli su ogni verbo del dataset in vista della discussione nel capitolo 5.

4.1. EXPLICIT dataset

La tabella 1 riporta punteggi di *precision*, *recall* e *F-measure* sull'EXPLICIT dataset che, come previsto, sono abbastanza bassi, con una *F-measure* inferiore a 0,45 in tutti i casi.

Modello	<i>Precision (weighted)</i>	<i>Recall (weighted)</i>	<i>F-measure (weighted)</i>
BERT-multi	0,17	0,12	0,12
BERT-it	0,42	0,31	0,33
UmBERTo	0,52	0,41	0,44
BERTino	0,44	0,37	0,38
ELECTRA	0,38	0,3	0,31

Tabella 1. Risultati delle metriche di valutazione classiche sui punteggi dei modelli.

Tuttavia, le normali misure di *evaluation* non sono adatte in questo caso a valutare adeguatamente la *performance* dei modelli, dato che la classe di completamenti è aperta e il task è di natura semantica. I risultati, come spiegato nella [sezione 3.2.](#), sono stati quindi calcolati attraverso la *cosine similarity*, facendo una media dei punteggi ottenuti su ogni frase. Essi sono riportati nella figura 1.

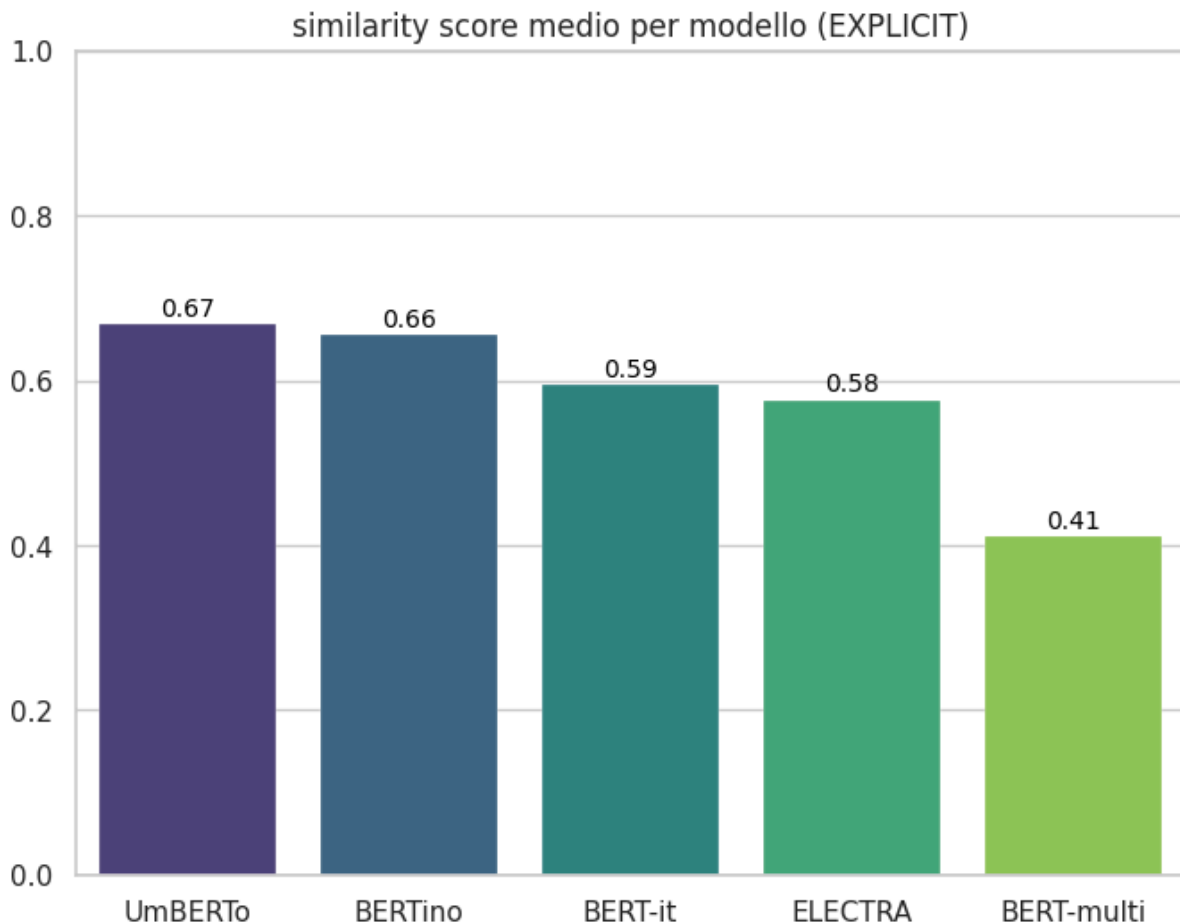


Figura 1. Risultati dei singoli modelli.

Vediamo che UmBERTo ha raggiunto il punteggio più alto (0,67), seguito da BERTino (0,66). Il risultato è positivo, considerando il rapporto tra risorse impiegate e *performance*: entrambi i modelli sono stati allenati con un dataset ridotto rispetto ai più grandi BERT-it e ELECTRA. Questi ultimi due modelli hanno ottenuto un punteggio più basso, rispettivamente di 0,59 e 0,58. Infine, BERT-multi ha raggiunto un punteggio di 0,41, che è del 17% più basso degli altri modelli; questo risultato è in linea con quanto atteso e osservato in altri studi, poiché la *baseline* multilingue performa peggio dei modelli *fine-tuned* in tutti i task .

Per un'idea più completa della distribuzione dei valori di *similarity*, nella tabella 2 viene riportata una statistica descrittiva dei risultati dei vari modelli sull'EXPLICIT dataset, comprendente i valori di Moda, Mediana, Media, Deviazione Standard, Minimo, Massimo e Campo di Variazione.

Modello	Moda	Mediana	Media	Deviazione Standard	Minimo	Massimo	Campo di Variazione
BERT-multi	1,0	0,36	0,410	0,288	-0,18	1,0	1,18
BERT-it	1,0	0,54	0,594	0,320	-0,16	1,0	1,16
UmBERTo	1,0	0,69	0,668	0,323	-0,16	1,0	1,16
BERTino	1,0	0,63	0,655	0,310	0,0	1,0	1,0
ELECTRA	1,0	0,52	0,576	0,329	-0,16	1,0	1,16

Tabella 2. Statistica descrittiva dei *similarity scores* dei modelli.

La Moda è 1 per ogni modello e significa che lo *score* di *similarity* ottenuto più frequentemente è 1. Il numero di risultati coincidenti con l'originale per ogni modello, che corrisponde al valore di *accuracy*, è riportato nella tabella 3. Nonostante l'*accuracy* sia bassa, le percentuali di risposte corrette sono da considerare ottime, considerato che il range di completamenti per uno spot nominale è ampio e che questi risultati non sono stati lemmatizzati. In particolare, UmBERTo si avvicina alla metà di risposte corrette sul totale (41,5%).

Modello	Risposte corrette	Percentuale	Accuracy
BERT-multi	70/600	11,67%	0,12
BERT-it	186/600	31,0%	0,31
UmBERTo	249/600	41,5%	0,42
BERTino	224/600	37,33%	0,37
ELECTRA	181/600	30,17%	0,3

Tabella 3. Risposte esatte per ogni modello.

Il valore Minimo e il Campo di Variazione nella tabella 2 invece ci mostrano che sono stati ottenuti dei valori di *similarity* negativi, che corrispondono ad un errore di computazione della libreria *SpaCy*. Questi valori negativi si trovano principalmente con parole straniere, che tuttavia sono riconosciute come parole del vocabolario italiano (la funzione *has_vector* impostata come filtro nello script ha valore positivo, v. [sez. 3.2.2.](#)). L'unico modello senza valori negativi è BERTino, ma questo risultato sembra essere casuale.

Come esempio di *similarity* negativa, si osservi la tabella 4, dove vengono presentati i risultati dei modelli (*out*) e i relativi punteggi (*sim*) per le frasi #54 e #284. Nella frase #54, il completamento GS è *the* con grafia inglese. Gli output dei modelli, *cielo* e *vino*, hanno ottenuto un punteggio negativo, con l'eccezione dell'output di BERTino, *caffè*, che ha

ottenuto una *similarity* positiva. Il risultato non ha una spiegazione linguistica, poiché *vino*, che appartiene allo stesso tipo semantico di *the*, cioè [Beverage], avrebbe dovuto ricevere un punteggio positivo. Questo dimostra che il funzionamento dei vettori di tipo *FasText* utilizzati dalla libreria *SpaCy* per il calcolo della *similarity* può essere oscuro dal punto di vista linguistico, in particolare nel caso di grafie non standard o prestiti.

Inoltre, alcune volte i modelli hanno generato un output troncato che ha ottenuto una *similarity* negativa, come nella frase #284 (tabella 4).

id	Frase	GS	BERT-multi		BERT-it		UmBERTo		BERTino		ELECTRA	
			out	sim	out	sim	out	sim	out	sim	out	sim
54	Continuai a bere in silenzio il [MASK], mentre il sole che tramontava tingeva di rosso il cielo.	the	cielo	-0,04	vino	-0,16	vino	-0,16	caffè	0,01	vino	-0,16
284	Guadagna [MASK] a palate e mangia solo tonno in scatola.	soldi	###re	-0,04	soldi	1	pesce	0,06	soldi	1	soldi	1

Tabella 4. Esempio di *similarity score* negativo sull'output dei modelli.

Nella tabella 5 si riassumono gli errori di computazione per ogni modello, cioè quei casi in cui l'output è troncato o sconosciuto (*unknown*) e in cui il punteggio di *similarity* è negativo. ELECTRA e BERT-multi contengono il numero più alto di errori, UmBERTo e BERTino quello più basso.

Modello	Output troncato	<i>unknown</i>	<i>Similarity</i> negativa
BERT-multi	6	38	11
BERT-it	1	14	4
UmBERTo	0	10	3
BERTino	0	8	0
ELECTRA	2	22	6

Tabella 5. Errori di computazione per ogni modello.

Infine, è stato calcolato il *similarity score* medio raggiunto dai modelli sui singoli verbi. Il punteggio è stato ottenuto calcolando la media totale dei risultati di ogni modello sulle 20 frasi di ogni verbo. I risultati sono riportati nella figura 2.

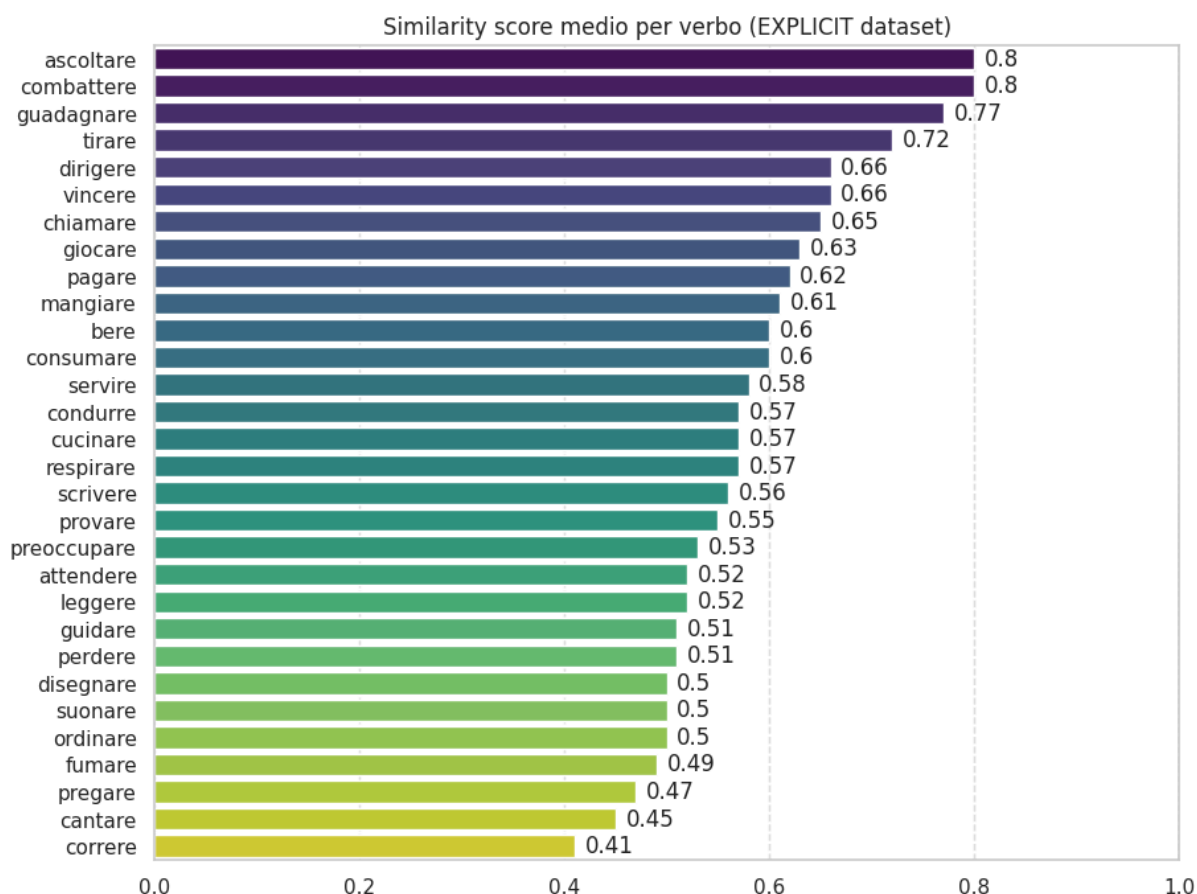


Figura 2. Risultati dei modelli sui singoli verbi.

4.2. IMPLICIT dataset

Per quanto riguarda l'IMPLICIT dataset, tutti i risultati riportati in questa sezione sono calcolati dopo aver lemmatizzato sia le annotazioni manuali che gli output dei modelli (v. [sez. 3.2.2](#)). Le metriche di valutazione classiche sono riportate nella tabella 6²⁶. Contrariamente alle aspettative, i risultati sono più bassi rispetto a quelli riportati per l'EXPLICIT dataset e tutti inferiori a 0,2.

Modello	<i>Precision (weighted)</i>	<i>Recall (weighted)</i>	<i>F-measure (weighted)</i>
BERT-multi	0,1	0,1	0,08
BERT-it	0,2	0,2	0,17
UmBERTo	0,15	0,16	0,13
BERTino	0,11	0,14	0,1
ELECTRA	0,18	0,16	0,13

Tabella 6. Risultati delle metriche di valutazione classiche sui punteggi dei modelli.

²⁶ Queste metriche sono calcolate facendo una media dei valori sul primo e sul secondo annotatore.

Anche i *similarity scores* dei modelli sono decisamente più bassi rispetto a quelli sull'EXPLICIT dataset. Mentre prima si andava da un minimo di 0,41 con BERT-multi ad un massimo di 0,67 con UmBERTo, qui si va da un minimo di 0,32, sempre per BERT-multi, ad un massimo di 0,46 con BERT-it. La differenza di punteggio tra i modelli è meno ampia, con l'eccezione ancora di BERT-multi, che si posiziona decisamente più in basso rispetto agli altri quattro modelli. Questa volta è un modello più grande a offrire migliori prestazioni, BERT-it, mentre BERTino raggiunge un punteggio leggermente inferiore degli altri (0,41).

Nella figura 3 vengono mostrati i risultati dei singoli modelli in comparazione con il punteggio di *similarity* tra i due annotatori (0,74). L'annotazione manuale viene qui trattata alla stregua di quella automatica per facilitare la comparazione tra le due.

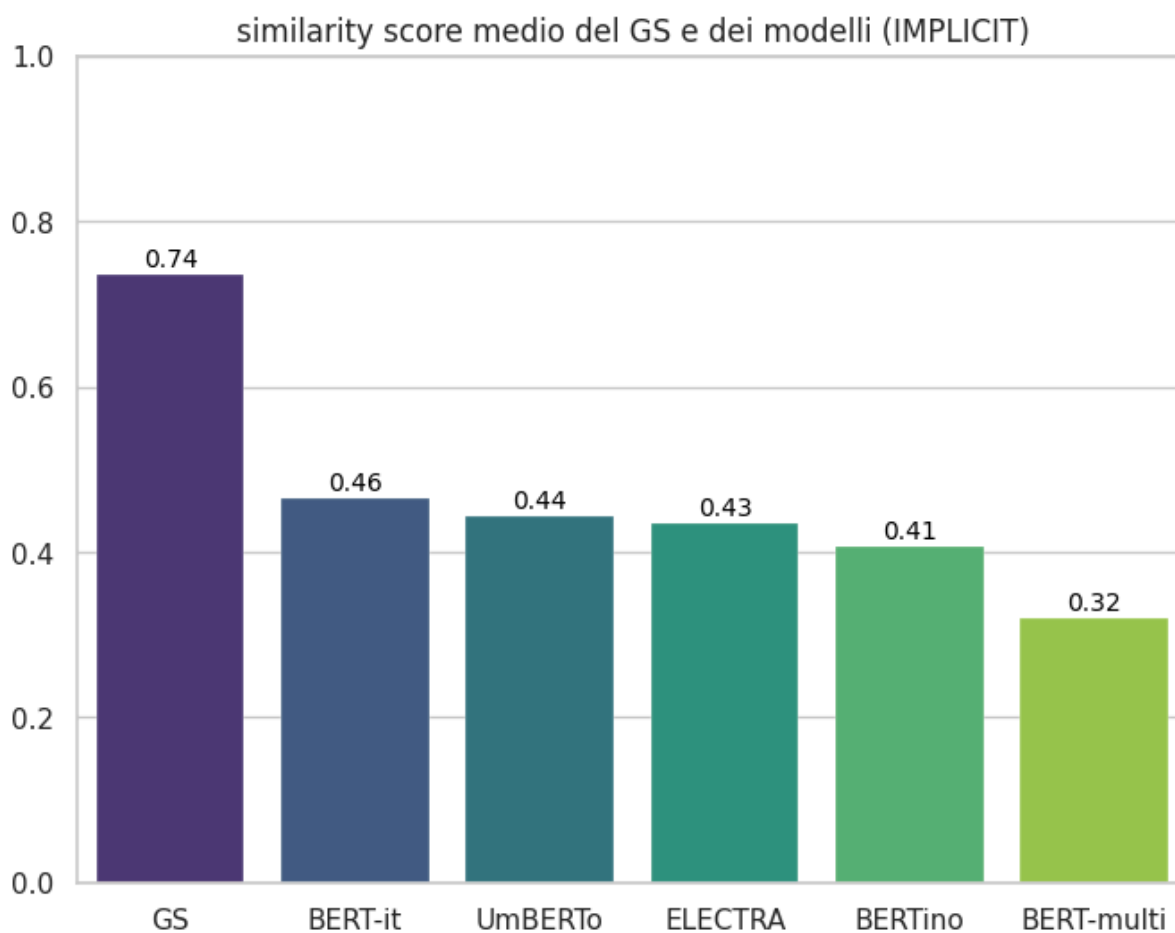


Figura 3. Risultati dei singoli modelli comparati con l'annotazione manuale (GS).

Nella tabella 7 viene riportata una statistica descrittiva dei risultati dei vari modelli e dell'annotazione GS sull'IMPLICIT dataset.

Modello	Moda	Mediana	Media	Deviazione Standard	Minimo	Massimo	Campo di Variazione
GS	1,0	1,0	0,735	0,317	-0,06	1,0	1,06
BERT-multi	1,0	0,27	0,319	0,281	-0,23	1,0	1,23
BERT-it	1,0	0,4	0,464	0,304	-0,1	1,0	1,1
UmBERTo	1,0	0,4	0,443	0,294	-0,08	1,0	1,08
BERTino	1,0	0,38	0,406	0,292	-0,23	1,0	1,23
ELECTRA	1,0	0,39	0,434	0,298	-0,17	1,0	1,17

Tabella 7. Statistica descrittiva delle annotazioni manuali (GS) e dei *similarity scores* dei modelli.

Nonostante la Media dei modelli si abbassi rispetto a quanto osservato sull'EXPLICIT dataset, il valore della Moda è 1 in tutti i casi; significa che, anche in questo dataset, lo *score* ottenuto più volte è 1. Questo *score* indica un'identità tra l'output del modello lemmatizzato e le parole previste come GS dagli annotatori, anch'esse lemmatizzate e identiche tra loro. Il numero di risposte corrette è indicato nella tabella 8. Nel caso dell'annotazione manuale, si è raggiunto un accordo perfetto in più della metà dei casi (53,5%), mentre BERT-it, UmBERTo e ELECTRA hanno superato il 10% di risposte corrette. Si tratta comunque di un buon risultato, considerate la possibilità di completare lo spazio con un range di parole molto ampio e il design dell'esperimento (v. [sez. 3.2.2.](#)).

Modello	Risposte corrette	Percentuale	Accuracy
GS	321/600	53,5%	0,54
BERT-multi	45/600	7,67%	0,07
BERT-it	89/600	14,83%	0,15
UmBERTo	71/600	11,83%	0,12
BERTino	58/600	9,33%	0,09
ELECTRA	70/600	11,67%	0,12

Tabella 8. Risposte esatte per ogni modello.

Continuando a commentare la tabella 7, il campo di variazione di BERTino (1,23), che nell'EXPLICIT dataset era il più basso (1,0), qui è il più alto insieme a quello di BERT-multi, che dimostra come questo valore sia scollegato da una tendenza dei modelli ma dipenda dal funzionamento dei vettori per il calcolo della *similarity* e dal design dell'esperimento. La Media dei modelli, presa come riferimento per valutarne le *performance*, è parzialmente

influenzata dai punteggi di *similarity* negativi, che possono considerarsi, come già detto, errori di computazione. Per questo, una variazione di alcuni decimali nella media non va considerata come uno scarto significativo nel risultato dei modelli.

Nella tabella 9 si riportano gli errori di computazione per ogni modello su questo dataset. Il numero di errori si alza, soprattutto per BERT-multi, con punteggi di *similarity* sotto lo 0 in 36 frasi, e BERTino, che restituisce *unknown* in 31 casi.

Modello	Output troncato	<i>unknown</i>	<i>Similarity</i> negativa
BERT-multi	27	29	36
BERT-it	2	17	6
UmBERTo	0	21	5
BERTino	6	31	11
ELECTRA	2	19	5

Tabella 9. Errori di computazione per ogni modello.

Il *similarity score* medio raggiunto per i singoli verbi anche in questo dataset è stato calcolato sulle 20 frasi di ogni verbo unendo i risultati di tutti i modelli su quelle frasi, senza tenere in considerazione i punteggi della *similarity* tra annotatori. I risultati sono riportati nella figura 4.

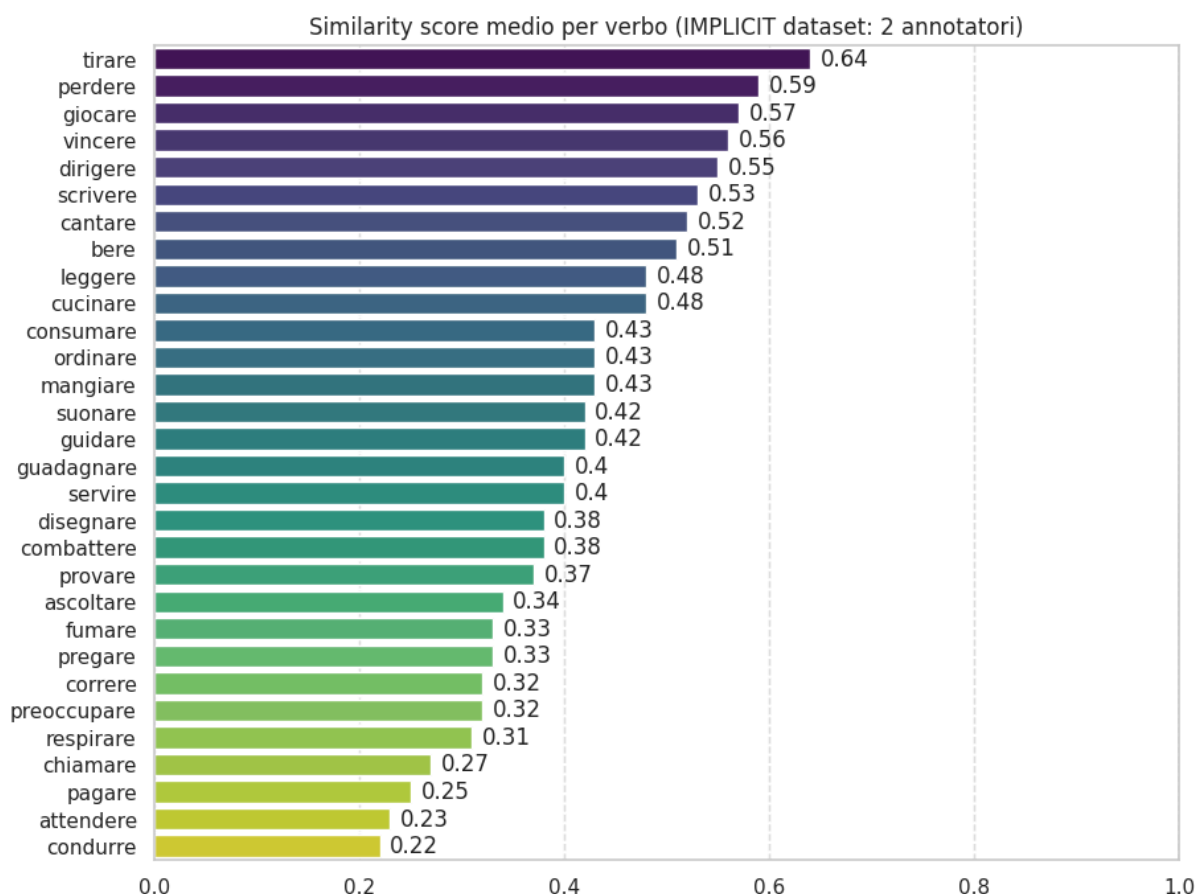


Figura 4. Risultati dei modelli sui singoli verbi.

La tabella 10 riassume i punteggi medi di *similarity* per ogni verbo dell'Ontologia ottenuti dai modelli sull'EXPLICIT e sull'IMPLICIT dataset. Vengono riportati inoltre i punteggi di *similarity* tra le due annotazioni manuali, sempre divisi per verbo. I verbi sono classificati in base alle classi semantiche di Levin (1993) discusse nella [sezione 2.1](#). Questa tabella permette di osservare più facilmente le differenze e le similarità nel trattamento di ogni verbo e di ogni classe semantica, confrontando tra loro l'elaborazione automatica e manuale, nonché i risultati sui due dataset.

Classe semantica	Verbo	EXPLICIT	IMPLICIT	
		Modelli	Modelli	Annotatori
<i>Verbs of Creation and Transformation</i>	cantare	0,45	0,52	0,7
	cucinare	0,57	0,48	0,77
	dirigere	0,66	0,55	0,8
	disegnare	0,5	0,38	0,79
	giocare	0,63	0,57	0,91
	provare	0,55	0,37	0,83
	scrivere	0,56	0,53	0,76
	suonare	0,5	0,42	0,69
<i>Verbs of Ingesting</i>	bere	0,6	0,51	0,89
	mangiare	0,61	0,43	0,89
<i>Verbs Involving the Body</i>	fumare	0,49	0,33	0,58
	respirare	0,57	0,31	0,94
<i>Verbs of Motion</i>	correre	0,41	0,32	0,6
	guidare	0,51	0,42	0,87
<i>Psych-Verbs</i>	pagare	0,62	0,25	0,35
	preoccupare	0,53	0,32	0,43
<i>Verbs of Change of Possession</i>	guadagnare	0,77	0,4	0,65
	ordinare	0,5	0,43	0,86
	perdere	0,51	0,59	0,93
	servire	0,58	0,4	0,66
	vincere	0,66	0,56	0,9
<i>Verbs of Existence</i>	attendere	0,52	0,23	0,55
	consumare	0,6	0,43	0,93
<i>Verbs of Communication</i>	chiamare	0,65	0,27	0,5
<i>Verbs of Desire</i>	pregare	0,47	0,33	0,74
<i>Verbs of Perception</i>	ascoltare	0,8	0,34	0,74
<i>Verbs of Social Interaction</i>	combattere	0,8	0,38	0,71
<i>Learn Verbs</i>	leggere	0,52	0,48	0,7
<i>Verbs of Throwing</i>	tirare	0,72	0,64	0,95
<i>Verbs with Predicative Complements</i>	condurre	0,57	0,22	0,43

Tabella 10. Confronto dei punteggi di *similarity* ottenuti per ciascun verbo sui due dataset.

Capitolo 5. Discussione

In questo capitolo si discutono i risultati presentati nel capitolo 4, rimandando anche alle osservazioni sull’annotazione ([sez. 2.3](#)). Nella prima sezione si cerca di rispondere alle domande presentate all’inizio del lavoro; nella [sezione 5.2](#), si analizzano dal punto di vista linguistico gli output dei modelli sulle varie classi verbali.

5.1. Panoramica

In questa prima sezione si affrontano le domande di ricerca presentate nell’introduzione:

- Ci sono differenze tra il completamento di un Oggetto Implicito e di uno Esplicito?
- Quali modelli producono output più simili al Gold Standard?
- Qual è il ruolo del contesto?
- Su quali classi verbali si ottengono risultati migliori?
- I modelli riconoscono il tipo semantico dell’Oggetto?

Iniziamo con la questione delle differenze tra l’EXPLICIT e l’IMPLICIT dataset. La decisione di includere anche un dataset con Oggetti Espliciti nell’esperimento, come accennato nella [sez. 1.2.2](#), è basata su uno studio di Ye et al. (2022) in cui si osservava che, sul task di completamento di un token implicito in presenza di metonimia logica, il modello *fine-tuned* su frasi implicite funzionava meglio di quello *fine-tuned* su frasi esplicite. Si ipotizzava che questa differenza fosse dovuta alla maggiore ricchezza contestuale delle frasi implicite, che permetteva di ricavare più informazioni semantiche.

Nel presente esperimento, l’ipotesi è stata applicata al completamento dell’Oggetto Implicito in italiano. In questo caso, si prevedeva un migliore funzionamento dei modelli sul dataset implicito, grazie alla maggiore ricchezza informativa del contesto. I risultati ottenuti smentiscono l’ipotesi. Il punteggio di *similarity* dei modelli è decisamente più basso sull’IMPLICIT dataset per ogni modello (tabella 1).

Modello	Similarity media	
	EXPLICIT dataset	IMPLICIT dataset
BERT-multi	0,410	0,319
BERT-it	0,594	0,464
UmBERTo	0,668	0,443
BERTino	0,655	0,406
ELECTRA	0,576	0,434

Tabella 1. Confronto dei risultati dei modelli sui due dataset.

Teniamo in considerazione, tuttavia, alcuni punti: 1) l'EXPLICIT dataset contiene un contesto morfosintattico molto più ricco dell'IMPLICIT dataset; 2) il task differisce da quello di Ye et al. (2022), dove si completava lo spazio mascherato con un verbo, poiché prevede un completamento di un Nome, una classe più ampia; 3) il design dell'esperimento comprende una strategia per trattare il doppio [MASK]; 4) i risultati sono stati lemmatizzati; 4) La *similarity* è una metrica indicativa che potrebbe non riflettere la correttezza linguistica degli output. Inoltre, per due verbi del dataset l'ipotesi è verificata, poiché il loro punteggio di *similarity* è più alto sull'IMPLICIT dataset (*cantare* EXP: 0,45; IMP: 0,52 e *perdere* EXP: 0,51; IMP: 0,59), e per altri verbi si ottengono punteggi molto vicini nei due dataset, con una variazione minima (es. *scrivere* EXP: 0,56; IMP: 0,53 e *leggere* EXP: 0,52; IMP: 0,48). Per tutte queste ragioni è necessaria un'analisi linguistica che consideri i verbi separatamente, per cui si rimanda alla [sezione 5.2](#).

Per quanto riguarda le prestazioni dei diversi modelli, la classifica cambia da un dataset all'altro: sul dataset EXPLICIT funzionano meglio i modelli italiani allenati su piccoli corpora, UmBERTo e BERTino, mentre sul dataset IMPLICIT funziona meglio un grande modello, BERT-it, affiancato sempre da UmBERTo e subito dopo da ELECTRA, mentre la *performance* di BERTino si abbassa di molto. I punteggi dei modelli UmBERTo e BERTino sono quelli che diminuiscono di più dall'uno all'altro dataset. Nel caso di BERTino, esso è anche il modello con più output vuoti o *unknown*, forse perché privilegia altre categorie grammaticali, diverse dai nomi, per il completamento della posizione mascherata.

Per quanto riguarda i due tipi di *Defaulting* annotati nell'IMPLICIT dataset, l'uno dipendente dalla semantica verbale (*Lexical*) e l'altro dal contesto (*Pragmatic*) (v. [sez. 1.1.1](#)), il coefficiente di Pearson riporta una correlazione tra il tipo di *Defaulting* e l'ampiezza del range di Oggetti GS (v. [sez. 2.3.2](#)): un verbo con *Defaulting* prevalentemente lessicale tenderà ad avere un range di Oggetti più ridotto (es. *guadagnare*, *tirare*, *respirare*), e

viceversa, un verbo con *Defaulting* pragmatico tenderà ad avere un range più ampio (es. *condurre, attendere, ascoltare*).

Rimane da verificare l'effetto del tipo di *Defaulting* sui punteggi di *similarity*. Per farlo, è stato calcolato il coefficiente di Pearson tra il numero di volte in cui il *Lexical Defaulting* è stato annotato per un certo verbo dai due annotatori (cfr. figura 6, [sezione 2.3.2.](#)) e il punteggio medio di *similarity* per quel verbo (cfr. figura 4, [sezione 4.2.](#)). Il valore della metrica è 0,287 e indica una correlazione positiva, ma il *p-value* uguale a 0,125 indica che la correlazione positiva tra queste due variabili non è statisticamente significativa. L'influenza del tipo di *Defaulting* potrebbe essere circoscritta a quei verbi fortemente orientati verso uno o l'altro tipo di omissione, per cui anche in questo caso si rimanda ad un'analisi dei singoli verbi.

Risulta essere una correlazione negativa statisticamente significativa invece quella tra l'ampiezza del range di Oggetti GS e il punteggio di *similarity*: -0,593, *p-value* 0,000558. Il valore negativo indica che, al crescere del range di Oggetti GS, il punteggio di *similarity* medio dei modelli si abbassa.

Per valutare se c'è una differenza statisticamente significativa nei punteggi delle varie classi verbali, è stato usato il test *oneway ANOVA*, che rileva un'eventuale differenza tra le medie di almeno tre gruppi. Il test considera come *features* di raggruppamento le classi semantiche, suddivise in base a Levin (1993), e il tipo azionale dei verbi, in base alla divisione di Vendler (1957) (cfr. [sez. 2.1.](#), pp. 26-27). I valori sui due dataset sono mostrati nella tabella 2.

Dataset	Classe semantica		Tipo azionale	
	<i>F-value</i>	<i>p-value</i>	<i>F-value</i>	<i>p-value</i>
IMPLICIT	2,915	0,023	1,355	0,278
EXPLICIT	2,231	0,065	0,668	0,583

Tabella 2. Risultati del test ANOVA per calcolare l'effetto delle classi verbali sul punteggio medio di *similarity*.

Il test dà risultati significativi solo per la relazione tra la classe semantica e i punteggi di *similarity* ottenuti sull'IMPLICIT dataset (*F-value* = 2,915; *p-value* = 0,023). In altre parole, la classe semantica dei verbi sembra influenzare in modo significativo la variazione del punteggio di *similarity* solo nell'IMPLICIT dataset; invece, nell'EXPLICIT dataset, il valore di *p-value* per questa classe si alza leggermente al di sopra della soglia di significatività e

quindi la correlazione non è da considerarsi statisticamente significativa ($p\text{-value} = 0,065$). Per quanto riguarda il tipo azionale dei verbi, la correlazione con il punteggio di *similarity* non è significativa né per l'IMPLICIT dataset né per l'EXPLICIT dataset.

Per verificare la differenza tra i punteggi delle classi nell'IMPLICIT dataset, è stato applicato il test di Tukey²⁷. Nel caso dei tipi azionali (figura 1), osserviamo che gli *Achievements* hanno un punteggio medio migliore degli altri tipi azionali, mentre gli *Activites* sono la classe più omogenea nella distribuzione dei punteggi. Le differenze tra le classi non sono statisticamente significative.

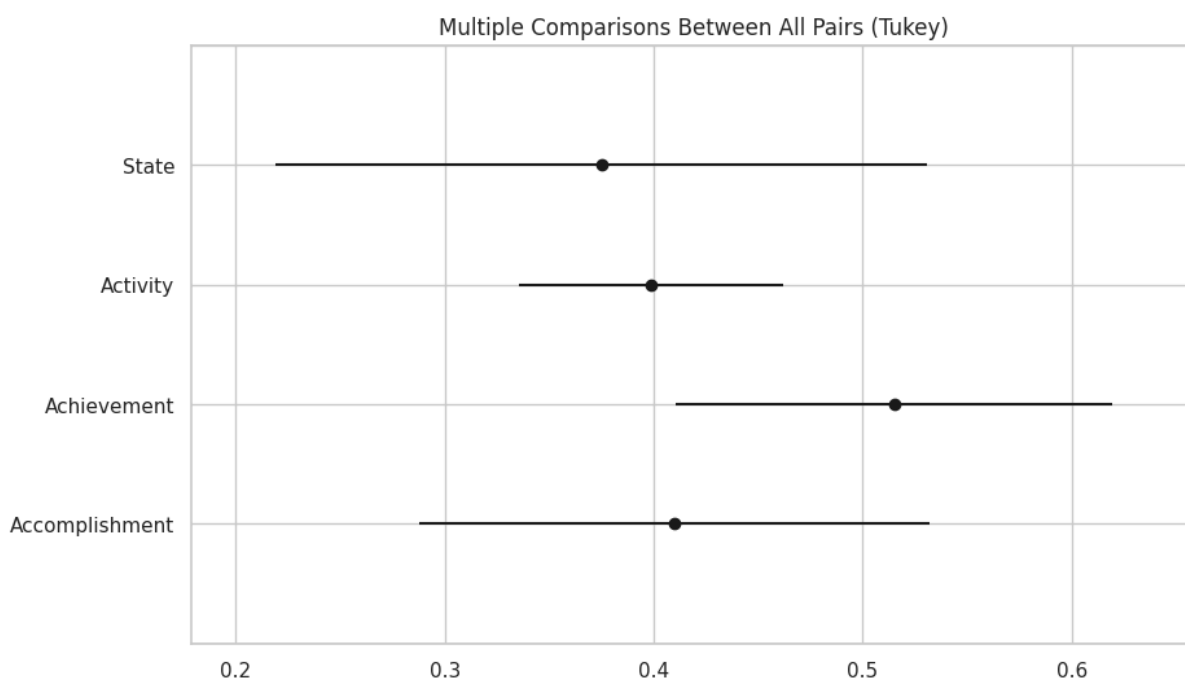


Figura 1. Comparazione tra i punteggi medi dei modelli sui tipi azionali (IMPLICIT dataset).

La figura 2 invece mostra i punteggi medi delle classi semantiche. Anche in questo caso, il test di Tukey non rileva differenze significative tra le classi, ma la figura ci può dare un'idea della distribuzione dei punteggi.

²⁷ Il test confronta tra loro tutte le possibili coppie di gruppi (es. i tipi azionali *States* e *Activities*).

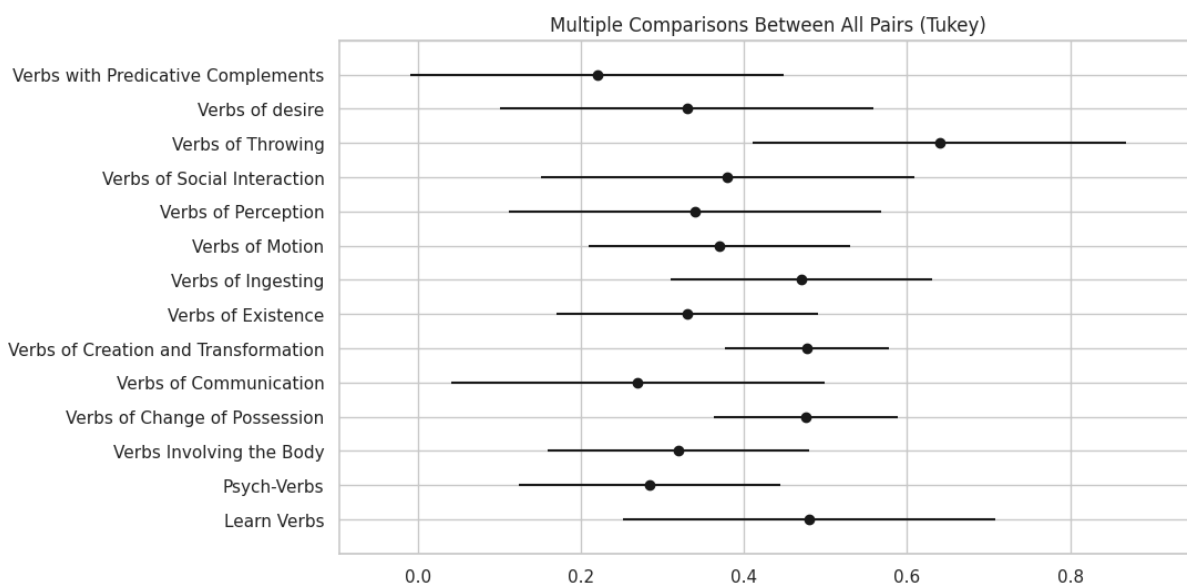


Figura 2. Comparazione tra i punteggi medi dei modelli sulle classi semantiche (IMPLICIT dataset).

La classe con risultati migliori è *Verbs of Throwing*, che comprende l'*Achievement tirare* nel suo significato calcistico. In questo pattern l'Oggetto è molto poco variabile e si riduce a *palla, pallone e pallina*. Inoltre, il verbo è tra quelli con più annotazioni di *Lexical Defaulting*, poiché l'Oggetto tende ad essere codificato nel verbo. Il punteggio indica che il modello ha riconosciuto il pattern semantico di riferimento e ha ricostruito con facilità gli Oggetti Impliciti a partire dalla radice verbale, basandosi sulle collocazioni più frequenti verbo-Oggetto. Nella [sez. 1.2.2](#), è stato accennato che la conoscenza del mondo di BERT si basa soprattutto su *n-grams* e collocazioni, piuttosto che su vere e proprie inferenze logiche e capacità pragmatiche. Questo modo di lavorare per associazioni probabilistiche è del tutto in linea con l'abilità nel risolvere un *Lexical Defaulting* e con il risultato raggiunto per questo verbo.

La classe con punteggio minore invece è *Verbs with Predicative Complements*²⁸, che comprende *condurre* nel senso di *presentare una trasmissione* o *un incontro*. Questo verbo è speculare a *tirare* in quanto a tipo di omissione; infatti, abbiamo visto che è il verbo con il maggior numero di annotazioni di *Pragmatic Defaulting* e anche uno tra quelli con un range di Oggetti più ampio, poiché l'Oggetto Implicito si deve (quasi) sempre dedurre dal contesto pragmatico (v. [sez. 2.3.2](#)). Un punteggio minore su questo verbo potrebbe indicare una

²⁸ Questa classe comprende verbi che reggono un complemento predicativo. Possono essere parafrasati nella struttura "NP V NP come NP" (Levin, 1993), per esempio "Marco conduce *Amici Miei* come presentatore". Per ulteriori spiegazioni cfr. [sez. 5.2.8](#), pp. 95-96.

minore capacità dei modelli a ricostruire l'Oggetto partendo dal contesto, un risultato che sarebbe in linea con le scarse capacità pragmatiche e inferenziali di BERT (v. [sez. 1.2.2.](#)). Per la discussione su *tirare* e *condurre* si rimanda alla [sezione 5.2.8.](#)

L'ultima domanda di ricerca tra quelle elencate all'inizio di questa sezione riguarda la capacità dei modelli di riconoscere il tipo semantico dell'Oggetto. Data la grande varietà di contesti e fattori in gioco, il modo migliore di rispondere è un'esplorazione di tipo descrittivo all'interno delle singole classi semantiche, che viene proposta nella prossima sezione. Tale esplorazione servirà anche ad osservare più da vicino gli *output* dei modelli e commentare il trattamento dei singoli verbi rispetto alle domande precedenti.

5.2. Esplorazione delle classi verbali

In questa sezione vengono discussi i risultati dei modelli sulle varie classi semantiche, con esempi tratti dall'IMPLICIT dataset. Per ogni sotto-sezione si propone una breve introduzione dove: 1) vengono presentati i verbi della classe semantica e i loro pattern; 2) viene data qualche informazione sui verbi, tra cui il tipo azionale, il tipo di alternanza transitiva e, quando presente, il tipo di NI in base alla divisione di Cennamo (2017) (v. [sez. 1.1.3.](#)); 3) quando sono rilevanti, vengono discusse le annotazioni manuali e i punteggi ottenuti dai modelli. Dopo l'introduzione generale della classe, verrà esaminato in dettaglio ciascun verbo. Per ognuno, vengono commentati gli output dei modelli sull'IMPLICIT dataset e vengono evidenziati i punti critici del completamento, facendo attenzione in particolare alla capacità di indovinare il giusto tipo semantico dell'Oggetto e il pattern di riferimento. Questa esplorazione permette di valutare il funzionamento dei modelli al di là del punteggio di *similarity* e tenendo in considerazione tutti i fattori linguistici in gioco, tra cui il contesto sintattico, le relazioni di significato tra parole e la presenza di inferenze logiche.

5.2.1. *Verbs of Creation and Transformation*

Questa classe comprende i verbi *cantare*, *cucinare*, *dirigere* (nel significato di *guidare un gruppo nell'esecuzione di una performance* oppure *nella creazione di un film*), *disegnare*, *giocare* (nel significato sportivo di *disputare*), *provare* (nel significato di *eseguire le prove di una performance*), *scrivere* e *suonare* (compreso il significato di *suonare il campanello* o *il citofono*).

Proprio perché coinvolgono un processo di creazione o trasformazione, questi verbi sono durativi, dinamici e atelici (*Activities*)²⁹. I cosiddetti *Active Accomplishments* come *disegnare* possono diventare telici quando l'Oggetto è specifico e codifica al suo interno la fine del processo (es. *‘‘ho cantato in due minuti’’ vs ‘‘ho cantato la canzone in due minuti’’). L'Oggetto di questi verbi è *affected* (Levin, 1993) o incrementale, perché aumenta o diminuisce nel corso dell'evento.

Questi verbi esibiscono un'alternanza di tipo *Unspecified Object* (Levin 1993), poiché nell'uso intransitivo con INI implicano spesso un Oggetto generico. Tuttavia, così come i *Verbs of Ingesting* (v. [sez. 5.2.2.](#)), alternano liberamente tra Oggetti generici e specifici e talvolta le due strategie di *Defaulting* sembrerebbero addirittura sovrapporsi (Cennamo, 2017; Jezek, 2018) (cfr. esempio 6, [sez. 1.1.1.](#)).

Per quanto riguarda l'annotazione del *Defaulting* nell'IMPLICIT dataset, è stata distribuita equamente, senza tendenze spiccate verso l'uno o l'altro tipo, il che conferma la loro elasticità da questo punto di vista. C'è una leggera prevalenza di *Lexical Defaulting* in *dirigere*, *disegnare* e *suonare* (più di 25 annotazioni). Questa annotazione si trova spesso con una lettura aspettuale imperfettiva, quando il verbo descrive una propensione o abitudine del Soggetto, o semplicemente quando l'Oggetto è generico e corrisponde al rappresentante della categoria semantica di riferimento (es. tabella 3).

id	Frase	Defaulting		Oggetto GS	
		1	2	1	2
204	Per una dozzina di anni Keaton si aggira come un fantasma per gli studios di Hollywood, <i>dirigendo</i> [MASK][MASK], sceneggiando, interpretando, nell'anonimato o quasi.	<i>Lexical Defaulting</i>	<i>Lexical Defaulting</i>	film	film

Tabella 3. Esempio di *Lexical Defaulting* in una frase con *dirigere*³⁰.

In generale, i modelli completano con successo gli Oggetti di questi verbi, con un punteggio di *similarity* che si aggira intorno allo 0,5.

Iniziamo la discussione con il verbo *cantare*. Esso è uno dei due verbi su cui i modelli raggiungono un punteggio di *similarity* più alto nelle frasi implicite che in quelle esplicite. Il

²⁹ Con l'eccezione di *suonare* quando regge *il citofono* o *il campanello*, che è un *Achievement*. Questo pattern occorre solo tre volte nel dataset, quindi il verbo è stato considerato come un *activity*.

³⁰ Le annotazioni manuali e gli output dei modelli che vengono presentati in questo capitolo non sono lemmatizzati, mentre i punteggi di *similarity* riportati nelle tabelle sono stati calcolati dopo la lemmatizzazione.

punteggio basso sull'EXPLICIT dataset è dovuto al grande numero di sinonimi e parole metonimiche tra gli output dei modelli, come nelle due frasi riportate nella tabella 4.

id	Frase	Oggetto GS	UmBERTo		BERTino	
			<i>out</i>	<i>sim</i>	<i>out</i>	<i>sim</i>
67	Qui chi non <i>canta</i> la [MASK] anche se stupida, anche se sciocca, sparisce tosto senza una traccia, non ha più voce, non ha più faccia.	filastrocca	canzone	0,59	canzone	0,59
69	io l'anno scorso dopo il concerto di lucca non smettevo più di <i>cantare</i> il [MASK]!	ritornello	rock	0,49	brano	0,63

Tabella 4. Esempio di frase con *cantare* (EXPLICIT dataset).

Nell'IMPLICIT dataset i modelli riescono ad indovinare quasi sempre il tipo semantico [Musical Composition] e la *similarity* è più alta perché spesso indovinano il rappresentante prototipico della categoria, *canzone* (37%)³¹. Solo in poche eccezioni viene restituita una parola in relazione metonimica, come per esempio uno strumento musicale (es. tabella 5). Negli altri casi, la *similarity* si abbassa semplicemente per la possibilità di scegliere tra vari sinonimi, come *brano*, *pezzo*, *canzone* e *canto*. Infine, quattro frasi contengono come Oggetto *coro* in riferimento al coro da stadio; i modelli hanno fatto molta fatica a indovinare il contesto e solo una volta BERT-multi ha indovinato il lemma (es. #65, tabella 5).

id	Frase	Oggetto GS		UmBERTo		ELECTRA	
		1	2	<i>out</i>	<i>sim</i>	<i>out</i>	<i>sim</i>
76	Achille era già uomo, ma ancora temeva la verga, quando <i>cantava</i> [MASK][MASK] sui monti della sua patria, e mai avrebbe riso della coda che aveva il citaredo suo maestro; oggi i discepoli pigliano a bastonate Rufo e gli altri maestri; Rufo, che pure è spesso detto il Cicerone allòbrogo.	canti	canzoni	tromba	0,37	canzone	0,71
65	Stavolta però in qualsiasi angolo della Curva <i>si è cantato</i> [MASK][MASK] e l'effetto canoro e coreografico hanno dimostrato ancora una volta che il nostro tifo è al di sopra di tanti altri che professano i migliori del mondo.	cori	coro	tifo	0,15	ritmo	0,25

Tabella 5. Esempio di frase con *cantare*.

³¹ Le percentuali riportate vicino agli output, es. *canzone* (37%), indicano la frequenza con cui i modelli hanno restituito il lemma. Ognuno dei cinque modelli completa 20 frasi per verbo, quindi il numero totale di output è 100. Per esempio, il 37% indica che il lemma è stato restituito esattamente 37 volte.

Un altro verbo con un ottimo risultato è *giocare* (*similarity* media di 0,57); anche qui i modelli hanno riconosciuto il pattern e completato la parola mascherata con la giusta categoria semantica, [Competition]. I lemmi più frequenti sono *partita* (45%), *match* (10%) e *campionato* (7%). Questo verbo è affine agli *Achievements* *tirare*, *vincere* e *perdere* per la sfera semantica sportiva, e, come per quelli, il completamento del suo Oggetto Implicito risulta facile per i modelli. Questo potrebbe accadere perché l'uso del pattern è limitato ad un solo contesto settoriale.

Si trova qualche volta il completamento della parola mascherata con un'espressione temporale, che il filtro per isolare solo i sostantivi utilizzato nello script non ha potuto eliminare. Per esempio, nella frase #264 (tabella 6), il focus è stato messo sul fatto che la squadra *tornerà a giocare la sera*. In questo caso non si tratta di un errore. Esempi come questo ricorrono in tutto il dataset, mostrando come il punteggio di *similarity* per valutare questo task sia solo indicativo della correttezza dei completamenti.

id	Frase	Oggetto GS		BERT-multi		BERTino	
		1	2	out	sim	out	sim
264	Il Settebello tornerà a <i>giocare</i> [MASK][MASK] domenica sera alle 21 con la Serbia & Montenegro.	partita	partita	sera	0,4	sera	0,4

Tabella 6. Esempio di frase per *giocare* con output temporale.

Le frasi con il verbo *dirigere* appartengono a due pattern semantici, con Oggetti di tipo [Musical Performance] o [Movie]. Il primo pattern è stato completato facilmente. Alcune volte è stata indovinata la categoria semantica [Musical Performance] (es. *concerto*, *spettacolo*, *performance*), ma spesso l'output è metonimico (es. *banda*, *coro*, *violino*, *pianista*). Per quanto riguarda il secondo pattern, spesso è stato restituito il lemma prototipico del tipo semantico [Movie], *film* (29%). Anche qui sono state proposte parole affini, come iponimi (es. *commedia*, *musical*) o parole metonimiche (es. *pellicola*); alcune volte l'output è di tipo [Human] e si riferisce agli attori (es. *personaggio*, *attore*, *cast*).

In varie frasi, la posizione mascherata è stata completata con nomi predicativi riferiti al Soggetto (es. l'output di BERTino per la frase #215, tabella 6). Questo fenomeno si osserva anche in altri verbi, in particolare *condurre* (v. [sez. 5.2.8.](#)). Infine, i modelli hanno fatto confusione con un altro pattern di *dirigere*, dove il verbo ha il significato di *coordinare*, *sovrintendere* e regge i tipi semantici [Institution] o [Activity]. Tuttavia, ciò si verifica solamente quando il contesto è troppo povero per dedurre il pattern corretto (es. UmBERTo

per la frase #215, tabella 7). In generale, il comportamento dei modelli con questo verbo rivela un'abilità nel riconoscere il giusto pattern grazie agli indizi contestuali.

id	Frase	Oggetto GS		UmBERTo		BERTino	
		1	2	<i>out</i>	<i>sim</i>	<i>out</i>	<i>sim</i>
215	<i>Dirige</i> [MASK][MASK] MICHAEL MOORE.	film	film	collezione	0,14	regista	0,64

Tabella 7. Esempio di frase con *dirigere*.

Per quanto riguarda il verbo *scrivere*, i risultati sono ottimi. Per lo più gli output rientrano nell'ampia categoria [Document]. I più frequenti sono *articolo* (20%), *libro* (11%), *lettera* (9%) e *testo* (9%). Alcune volte i risultati rientrano nella sfera della programmazione (es. *codice*, *programma*) e in quella della messaggistica online (es. *messaggio*, *mail*), coerentemente con il contesto della frase.

Cucinare ha ottenuto un punteggio nella media (0,48). Così come *mangiare*, questo verbo ha molta elasticità tra *Lexical* e *Pragmatic Defaulting* e può implicare un'entità specifica o generica. Le parole generiche *cibo* e *piatto* coprono rispettivamente il 12% e l'11% degli output; le parole più specifiche *pasta* e *pane* corrispondono in totale al 17% degli output, *carne* e *pesce* al 12%. Tutti gli output appartengono ai due tipi semantici indicati nel pattern di riferimento, [Food] e [Meal], con pochissime eccezioni.

È molto interessante osservare gli output dei modelli per vedere in cosa consiste la loro conoscenza del mondo, che deriva dai dati con cui sono stati allenati. Consideriamo per esempio la frase #182 (tabella 8). Il contesto della frase non permette di ricostruire un'entità specifica perché non dà informazioni a riguardo; è un caso di *Pragmatic Defaulting* in cui l'Oggetto è codificato nel contesto extra-frasale. Il primo annotatore supera questo ostacolo annotando un'entità generica, *piatto*, mentre il secondo annotatore annota un elemento presente nella frase, *birra*; i modelli invece restituiscono una parola specifica arbitraria in base alla forza della collocazione *cucinare* + Oggetto. In casi analoghi, i grandi modelli BERT-multi e BERT-it tendono a restituire *carne* e *pollo*, mentre UmBERTo e BERTino privilegiano *pasta*, *pane* e *pizza* (cfr. anche *servire*, [sez. 5.2.6.](#), pag. 84).

id	Frase	Oggetto GS		BERT-multi		BERT-it		UmBERTo		BERTino		ELECTRA	
		1	2	out	sim	out	sim	out	sim	out	sim	out	sim
182	Ovviamente potete gustarla bevendo la medesima birra utilizzata per <i>cucinare</i> [MASK][MASK].	piatto	birra	carne	0,45	pollo	0,45	pasta	0,52	pane	0,47	pasta	0,52

Tabella 8. Esempio di frase con *cucinare*.

Il verbo *suonare* comprende tre pattern. Nel primo regge il tipo semantico [Musical Instrument], nel secondo [Musical Composition]. Nel terzo pattern, assume il significato di *segnalare il proprio arrivo* e regge tipicamente i lemmi *citofono* e *campanello*. La *similarity* su questo verbo si abbassa leggermente a causa della sua polisemia (0,42).

Un risultato molto interessante è il mancato riconoscimento del pattern idiomatico. Nelle tre frasi con questo pattern, che vengono riportate nella tabella 9, solo BERTino riconosce il tipo semantico corretto e completa la parola mascherata con l’Oggetto GS (frase #577). La frase #548 contiene il famoso titolo “Il postino suona sempre due volte”; in questo caso, i modelli completano l’Oggetto con *pianoforte/piano*, non essendo in grado di capire cosa suoni il *postino*, cioè di cogliere la restrizione semantica del Soggetto sull’Oggetto.

id	Frase	Oggetto GS		BERT -multi	BERT -it	UmBERTo	BERTino	ELECTRA
		1	2					
548	dopo film come Cinque pezzi facili, Il postino <i>suona</i> [MASK][MASK] sempre due volte e La vedova nera, affronta l'avventura.	campanello	campanello	unknown	piano	pianoforte	pianoforte	pianoforte
549	Signora mia mi scusi <i>se ho suonato</i> [MASK][MASK] ma non so sche cosa sia la voglia di parlarle che mi ha preso è stato come un fuoco che si è acceso	citofono	campanello	pianoforte	sveglia	telefono	canzone	microfono
557	Dopo un un bel po' <i>ho suonato</i> [MASK][MASK] a più di 10 case, ho guadagnato solo qualche spicciolo e mentre cercavo di contarli sono passato davanti a un grande edificio con delle bandiere e con una scritta molto grossa che non so nemmeno cosa sia.	campanello	campanello	pianoforte	ritmo	sveglia	campanello	cori

Tabella 9. Esempi di frasi con *suonare* (pattern 3)³².

Gli ultimi due verbi di questa classe, *disegnare* e *provare*, sono i più problematici. Per quanto riguarda *disegnare*, il verbo è molto polisemico; nel pattern di riferimento regge i tipi semantici [Image] e [Physical Entity], ma in altri contesti può reggere come Oggetto il documento creato nel processo di disegnare, di tipo [Document], o può assumere il significato di *progettare*. Il punteggio si è abbassato perché gli annotatori hanno usato soprattutto il lemma *immagine* come rappresentante generico prototipico del tipo [Image], mentre i modelli hanno privilegiato *disegno* (18%), *fumetto* (10%) e *quadro* (8%), riferendosi quindi ad un pattern affine.

³² Per mancanza di spazio, i punteggi di *similarity* dei modelli vengono riportati qui di seguito nello stesso ordine della tabella. Frase #548: 0; 0,37; 0,44; 0,44; 0,44; frase #549: 0,37; 0,3; 0,51; 0,16; 0,53; frase 557: 0,44; 0,22; 0,41; 1; 0,05.

L'ultimo verbo della classe è *provare*. Anch'esso è un verbo molto polisemico; nel pattern scelto ha il significato di *eseguire le prove* e regge i tipi semantici [Artwork] e [Performance]. Nella maggior parte dei casi, i modelli riconoscono la sfera semantica dell'esibizione artistica, specialmente musicale, e restituiscono parole come *musica* (23%), *musical*, *concerto*, *spettacolo*, *scena*, *canzone* o termini metonimici come *disco*. Altri output rimandano a pattern diversi, come *costume*, con cui il verbo significa *indossare*, oppure *piano* e *chitarra*, con cui significa *sottoporre a prova*. Altre volte il significato si discosta maggiormente dal pattern scelto e il verbo viene inteso con il significato di *esperire* (es. *emozione*, *passione* e *vita*) oppure quello di *fare un tentativo* (es. *giro*, *tiro* e *gioco*). Un esempio con due pattern diversi è fornito nella tabella 10.

id	Frase	Oggetto GS		UmBERTo		BERTino	
		1	2	out	sim	out	sim
466	Poi forse Zappi propose di fare a voce tutto il concerto. <i>Provammo</i> un po' [MASK][MASK] nei camerini e salimmo sul palco tentando a fare a voce i nostri pezzi, ognuno imitando il proprio strumento.	concerto	concerto	sensazione	0,16	chitarra	0,38

Tabella 10. Esempio di frase con *provare*.

In conclusione, gli output per questa classe, in parte corrispondono ai rappresentanti prototipici delle categorie semantiche, in parte consistono in parole che si trovano in relazioni metonimiche con questi. Alcune volte i modelli hanno interpretato erroneamente il significato del verbo e non hanno riconosciuto il giusto pattern. Ciò si è verificato in due casi: 1) quando il contesto della frase non era abbastanza ricco e quindi era presente un *Pragmatic Defaulting* con rimando extra-frasale; 2) quando era necessaria una comprensione più profonda delle relazioni tra le parole e delle restrizioni sull'Oggetto. In particolare, i modelli hanno avuto difficoltà a completare l'Oggetto dei verbi polisemici nel caso di pattern meno frequenti, come *suonare il campanello* e *provare un concerto*, ignorando le restrizioni che il Soggetto applica sull'Oggetto e privilegiando invece le collocazioni verbo-Oggetto più frequenti.

5.2.2. *Verbs of Ingesting*

I due *Verbs of Ingesting* *bere* e *mangiare* si comportano in modo simile a quelli della classe precedente. Infatti sono *Activities* che reggono un tema incrementale e possono implicare un Oggetto generico oppure uno più specifico. Su entrambi, i modelli ottengono un punteggio di

similarity intermedio (rispettivamente 0,51 e 0,43). Nelle annotazioni manuali, *bere* ha una forte tendenza verso il *Pragmatic Defaulting*, perché tende a riferirsi a entità specifiche piuttosto che essere usato in modo generico; *mangiare*, d'altra parte, presenta una distribuzione equa dei due tipi di *Defaulting*.

Per quanto riguarda *bere*, il verbo comprende due pattern. Nel primo pattern, regge il tipo semantico [Beverage], mentre nel secondo pattern, dove ha il significato di *assumere bevande alcoliche*, regge principalmente i lemmi *alcol*, *alcolici* e *drink*. In linea con la prevalenza di *Pragmatic Defaulting*, il nome generico rappresentante la categoria è stato restituito una sola volta (*bevanda*), mentre in tutti gli altri casi i modelli hanno fatto riferimento a entità specifiche, in particolare *vino* (31%), *acqua* (13%) e *caffè* (13%). Un errore comune è stato la lemmatizzazione errata di *latte* in *latta* (4%), che ha abbassato leggermente il punteggio. In alcuni casi, è stato restituito il contenitore della bevanda (es. *bicchiere*, *bottiglia*), è stato quindi coinvolto un pattern metonimico.

Notevolmente, i modelli faticano a riconoscere la categoria [Alcoholic Beverage] quando il verbo significa *assumere bevande alcoliche*. Quando c'è un indizio contestuale, per esempio *bottiglie di vino* nella frase #58 (tabella 11), riescono ad interpretare il pattern, mentre quando non ci sono parole del *frame* e l'interpretazione richiede capacità deduttive, i modelli completano la parola mascherata con altri liquidi in base alle collocazioni più frequenti (es. #59, tabella 11).

id	Frase	Oggetto GS		BERT -multi		BERT-it		UmBERTo		BERTino		ELECTRA	
		1	2	out	sim	out	sim	out	sim	out	sim	out	sim
58	L'anno scorso gli inquirenti avevano cercato le impronte del ragazzo su alcune bottiglie di vino trovate nel ranch, al fine di sostenere le accuse secondo le quali Jackson avrebbe costretto il giovane a <i>bere</i> [MASK][MASK] per poi violentarlo.	vino	vino	vino	1	vino	1	vino	1	bicchieri	0,62	vino	1
59	Le ricerche dicono che i ragazzi <i>bevono</i> [MASK][MASK] per essere più socievoli, disinibiti ed euforici.	alcol	alcolico	casa	0,08	acqua	0,37	acqua	0,37	acqua	0,37	caffè	0,37

Tabella 11. Esempi di frasi con *bere*.

Il verbo *mangiare* per molti aspetti è analogo a *cucinare*; infatti i verbi reggono gli stessi due tipi semantici, [Food] e [Meal]. Come per *cucinare*, una piccola percentuale degli output corrisponde al rappresentante della categoria semantica: *cibo* (9%) e *pasto* (4%). Troviamo poi entità più specifiche come *panino* (13%), *carne* (10%), *pesce* (9%), *pane* (9%), *pranzo* (4%), *pizza* (5%), che purtroppo è stato lemmatizzato in *pizzare*, e *latte* (3%), lemmatizzato sempre in *latta*. In pochi casi è stata restituita una bevanda (*birra*, *vino*). In due casi, lo spazio mascherato è stato riempito con il Soggetto in posizione post-verbale (es. tabella 12), un fenomeno che osserviamo anche con altri verbi del dataset (per esempio *preoccupare*, v. [sez. 5.2.5.](#)).

id	Frase	Oggetto GS		BERT-multi		UmBERTo		BERTino	
		1	2	out	sim	out	sim	out	sim
342	Non solo, ma allorchè mi caricai la capra adulta sulle spalle per portarla via, il piccolo mi seguì fino al recinto; allora deposi a terra la madre, presi in braccio il capretto e lo portai all' interno, oltre la palizzata, nella speranza di poterlo allevare, ma si rifiutò di <i>mangiare</i> [MASK] e quindi mi vidi costretto a ucciderlo e a mangiarmelo.	erba	cibo	madre	0,15	cucciolo	0,31	vitello	0,3

Tabella 12. Esempio di frase con *mangiare*.

In conclusione, i modelli hanno saputo completare con successo le posizioni mascherate di *bere*, nonostante contenessero molte istanze di *Pragmatic Defaulting*. Per il verbo *mangiare*, hanno restituito risultati molto simili a quelli già descritti per la classe *Verbs of Creation and Transformation* (cfr. [sez. 5.2.1](#)). Hanno avuto qualche difficoltà a riconoscere il pattern di *bere* in cui il verbo significa *assumere bevande alcoliche*, perché la sua interpretazione richiede spesso un’elaborazione logica e inferenziale della frase e una comprensione profonda dei rapporti semantici tra le parole.

5.2.3. *Verbs Involving the Body*

I verbi di questa classe sono *fumare* e *respirare*. Sono stati assegnati a questa classe perché coinvolgono una parte del corpo, nello specifico descrivono l’inalazione attraverso i polmoni di qualche sostanza. Per quanto riguarda l’annotazione, *respirare* è molto tendente al *Lexical Defaulting* e ha un range di Oggetti di soli 3 elementi, di tipo semantico [Vapour]: *aria*, *ossigeno* e *odore*. Anche *fumare* ha un range di Oggetti abbastanza limitato, comprendente elementi della categoria [Drug] (*tabacco*, *sostanza*, *erba* e *droga*) e gli oggetti da cui si inala il fumo (*sigaretta* e *sigaro*). La distribuzione equa dei tipi di *Defaulting* sembra indicare che l’Oggetto di *fumare* è meno codificato nella radice verbale rispetto a quello di *respirare*. Nonostante il range limitato, entrambi i verbi hanno ottenuto dei punteggi più bassi rispetto a quelli delle classi precedenti.

Nel caso di *respirare*, gli output dei modelli comprendono elementi del tipo semantico di riferimento, come *respiro* (17%), *aria* (5%) e *odore* (4%). *Respiro* è uno *Shadow Argument* del verbo, ovvero è un Argomento incorporato nella semantica della radice verbale che non può essere espresso a meno che non venga ulteriormente specificato; i modelli tendono a

restituire Argomenti di questo tipo, come succede anche con altri verbi, mostrando una scarsa sensibilità a tale limite linguistico (v. ad esempio *tirare* e *pregare*, [sez. 5.2.8.](#)).

Sono stati restituiti inoltre elementi appartenenti ad altri tipi semantici, come *corpo* (13%), *terra*, *mare*, *città* e *mondo* (es. #488, tabella 13). In questi ultimi casi, il verbo può reggere gli Oggetti proposti dai modelli grazie ad un meccanismo di *semantic coercion* (Pustejovsky e Jezek, 2008), in cui il vapore è sostituito con la parte del corpo o il luogo che lo emette.

id	Frase	Oggetto GS		BERT-it		UmBERTo	
		1	2	out	sim	out	sim
488	I quali vorrebbero trasformare in realtà il mito dell'aria sempre pulita, o meglio, l'ipotesi di una mastodontica tenda ad ossigeno in cui serrare l'intera città, nel tentativo, lodevole per quanto utopistico, di farci <i>respirare</i> [MASK][MASK].	ossigeno	aria	città	0,11	pianeta	0,25

Tabella 13. Esempio di frase con *respirare*.

Per quanto riguarda *fumare*, la *similarity* si è abbassata per l'errore di lemmatizzazione del token *sigarette*, che è stato ricondotto al verbo *sigarettare*. Questo errore ha restituito un punteggio di *similarity* uguale a 0 perché il token *sigarettare* non ha un vettore associato nel dizionario di *SpaCy*, e ha interessato il 24% degli output (per esempio l'output di UmBERTo nella frase #253, tabella 14). Il lemma *sigaretta* è stato annotato nella maggior parte dei casi (27%), seguito da *tabacco* (25%) e *fumo* (6%). Se eliminiamo gli errori dovuti alla lemmatizzazione, i risultati sono molto positivi; gli output al di fuori della categoria semantica [Drug] sono pochi e sono stati prodotti soprattutto da BERT-multi. È interessante l'output *caffè* (5%), che per frequenza di collocazione è associato all'atto di *fumare*, anche se non può costituire il suo Oggetto (es. BERT-it nella frase #253, tabella 14).

id	Frase	Oggetto GS		BERT-it		UmBERTo	
		1	2	out	sim	out	sim
253	I tre amici stavano <i>fumando</i> [MASK][MASK] in auto, parcheggiata in una zona periferica di Pavullo quando i militari dell'Arma, nel loro consueto giro di pattugliamento della zona, li hanno scoperti.	erba	droga	caffè	0,2	sigarette	0

Tabella 14. Esempio di frase con *fumare*.

In conclusione, i bassi punteggi di *similarity* per i due verbi di questa classe sono dovuti a: 1) nel caso di *respirare*, l'esplicitazione errata dello *Shadow Argument*; 3) nel caso di *fumare*, errori di lemmatizzazione di alcune parole al plurale. La tipicità degli Oggetti di *fumare*, nello specifico *sigaretta* e *tabacco*, si riflette nell'annotazione di *Lexical Defaulting* e nel range ridotto di elementi dell'annotazione manuale, nonché sulla buona *performance* dei modelli, che completano la posizione mascherata con facilità. Per *respirare*, ciò non si verifica, perché i modelli tendono a replicare il meccanismo di *semantic coercion* (Pustejovsky e Jezek, 2008) e restituire parole di altri tipi semantici.

5.2.4. *Verbs of Motion*

La classe dei *Verbs of Motion* è rappresentata nell'Ontologia da *correre* e *guidare*. Il verbo *correre* nel suo uso transitivo ha il significato di *gareggiare* e regge il tipo semantico [Competition], con un uso limitato all'ambito sportivo. *Guidare* è usato nel significato frequente di *condurre* e regge il tipo semantico [Road Vehicle]. Il punteggio di *similarity* è particolarmente scarso per *correre* in entrambi i dataset (EXP = 0,41; IMP = 0,32) e potrebbe indicare il mancato riconoscimento del pattern, mentre *guidare* ha ottenuto un punteggio più alto (EXP = 0,51; IMP = 0,42).

Per quanto riguarda *correre*, gli Oggetti GS annotati manualmente sono *gara*, *competizione*, *corsa*, *distanza*, *trials*, *meeting* e *campionato*. Gli output dei modelli comprendono lemmi della categoria prevista dal pattern, con prevalenza di *gara* (9%) e *campionato* (5%). In altri casi, è stata indicata la distanza percorsa durante la competizione (es. *miglio*), il luogo (es. *strada*, *pista*) oppure il premio in palio (es. *trofeo*, *titolo*, *traguardo*), coinvolgendo pattern metonimici rispetto a quello di riferimento. Spesso i modelli hanno completato lo spazio mascherato con un'espressione temporale che indica il momento in cui si svolge la competizione (es. *domenica* (8%), *giorno*). In quattro casi, il senso del verbo è stato interpretato come idiomatico ed è stato restituito l'output *rischio* (es. #161, tabella 15). Anche nell'EXPLICIT dataset i modelli hanno restituito espressioni temporali e la parola *rischio* (6%).

id	Frase	Oggetto GS		UmBERTo	
		1	2	out	sim
161	Ora i francesi dovranno <i>correre</i> [MASK][MASK] senza un partner locale spalleggiati dai soli olandesi di Redevco, società specializzata nel business immobiliare.	gara	competizione	rischio	0,21

Tabella 15. Esempio di frase con *correre*.

Gli output dei modelli per *guidare* invece appartengono soprattutto alla categoria semantica di riferimento, [Road Vehicle]. Comprendono le parole generiche *veicolo* (17%) e *vettura* (3%) e iponimi più specifici come *macchina* (17%), che viene privilegiata rispetto al sinonimo *auto* (2%), e poi *moto* (3%), *taxi* (3%), *aereo*, *camion* e *treno*. Talvolta viene restituito un completamento che si riferisce al Soggetto post-verbale, come si osserva anche in altri verbi (per esempio *preoccupare*, [sez. 5.2.5](#), e *condurre*, [sez. 5.2.8](#)).

In alcuni casi, il verbo viene interpretato nel suo significato di *capeggiare*, come nell'esempio #312 (tabella 16), dove UmBERTo e ELECTRA completano lo spazio mascherato con *mondo*. In questo caso i modelli non indovinano la categoria dell'Oggetto, nonostante gli indizi contestuali che rimandano all'ambito delle auto da corsa (*Schumacher*, *pilota*, *vincerebbe*, *gara*) e compiono un errore logico, coinvolgendo un pattern semantico irrilevante, sulla base di una collocazione verbo-Oggetto frequente.

id	Frase	Oggetto GS		UmBERTo		ELECTRA	
		1	2	out	sim	out	sim
311	<i>Guidava</i> [MASK][MASK] sotto gli effetti degli stupefacenti appena assunti mettendo a repentaglio la sua incolumità, ma soprattutto quella di chi lo incrociava.	veicolo	veicolo	ragazzo	0,11	ragazzo	0,11
312	Schumacher è il pilota più forte del mondo, ma se dovesse <i>guidare</i> [MASK][MASK] con una mano sola non vincerebbe neppure una gara.	veicolo	auto	mondo	0,14	mondo	0,14

Tabella 16. Esempio di frasi con *guidare*.

In conclusione, i modelli hanno saputo riconoscere il giusto pattern di *guidare*, nonostante la sua polisemia, con alcune eccezioni dove si è verificata l'influenza del pattern in cui il verbo assume il significato di *capeggiare*. Per *correre* hanno dato punteggi di *similarity* peggiori perché hanno spesso coinvolto delle relazioni metonimiche e in vari casi hanno riempito la posizione mascherata con un'espressione temporale.

5.2.5. Psych-Verbs

Questa classe comprende i verbi *preoccupare* e *pagare*. *Preoccupare* è un verbo anomalo nell’Ontologia. Innanzitutto, esibisce il secondo tipo di INI individuata da Cennamo (2017), che è possibile solo in contesti imperfettivi. Inoltre, è l’unico verbo del dataset che regge sempre un Oggetto generico, tipicamente plurale e [+umano] e con il ruolo tematico di paziente o esperiente. Nella classificazione di Levin (1993), il verbo esibisce un’alternanza transitiva chiamata *PRO-arbitrary Object Alternations*, che indica l’interpretazione arbitraria dell’Oggetto nei contesti intransitivi, dove spesso si riferisce a entità generiche come *noi*, *uno* o *le persone* (v. [sez. 1.1.3.](#)). Nell’annotazione manuale, mostra una tendenza verso il *Lexical Defaulting*, perché gli annotatori hanno annotato spesso un rappresentante del tipo semantico [Human] (il primo annotatore *persone* (10/20 frasi), il secondo *popolazione* (6/20 frasi) o *interessati* (9/20 frasi)). *Pagare* è inteso nel significato particolare e meno comune di *espiare*, *scontare* e regge come Oggetto il tipo [Abstract Entity] (es. *conseguenza*, *debito*, *errore*).

I punteggi di *similarity* (rispettivamente 0,32 e 0,25) sono molto bassi e indicano una difficoltà dei modelli nel completare l’Oggetto per questi verbi. Anche l’annotazione manuale è stata problematica, infatti i punteggi di *similarity* tra annotatori sono tra i più bassi del dataset (0,43 e 0,35). Notiamo che sull’EXPLICIT dataset invece i modelli hanno ottenuto un buon punteggio per questi verbi (0,53 e 0,62), riconoscendo i pattern semantici grazie alla ricchezza del contesto; un esempio è riportato nella tabella 17.

id	Frase	Oggetto GS	UmBERTo		BERTino	
			<i>out</i>	<i>sim</i>	<i>out</i>	<i>sim</i>
390	Agamennone <i>pagò</i> il [MASK] del viaggio con la vita della propria figlia: trovò il suo letto occupato da un altro uomo, l'anima uscì dal suo cuore attraverso la ferita aperta dal pugnale stretto nella mano che credeva la più fidata.	prezzo	prezzo	1	prezzo	1

Tabella 17. Esempio di frase con *pagare* (EXPLICIT dataset).

Iniziamo la discussione con *preoccupare*. Per questo verbo, i modelli hanno restituito molti output appartenenti al tipo [Human]: in particolare, nomi generici rappresentanti la categoria (*gente* (5%) e *popolazione*), nomi riferiti a gruppi più ristretti di persone (es. *consumatori*, *medici*, *ricercatori*) oppure nomi collettivi (es. *governo* (5%), *società* (3%), *paese* (3%), *città*, *mondo*, *giuria*). I risultati sono buoni e rispettano il pattern; tuttavia, la *similarity* si è abbassata perché in quasi metà dei casi i modelli hanno completato lo spazio con un Soggetto

post-verbale come *rischio*, *criminalità*, *fenomeno*. A volte il completamento è corretto dal punto di vista sintattico, ma talvolta risulta agrammaticale perché il Soggetto è già espresso altrove (es. frasi #454 e #460, tabella 18). Questa tendenza è dovuta probabilmente alla frequenza di costruzioni con Soggetto post-verbale del tipo “preoccupa la criminalità” nei corpora di training (per lo stesso problema si veda anche la discussione su *condurre* nella [sezione 5.2.8.](#)).

id	Frase	Oggetto GS		BERT-multi		BERT-it		BERTino	
		1	2	out	sim	out	sim	out	sim
454	Ciò che <i>preoccupa</i> maggiormente [MASK][MASK] in tutti i paesi sono soprattutto tre tipi di reati: i reati contro i minori, i reati contro la morale pubblica, i reati contro il patrimonio.	persone	popolazione	violenza	0,52	criminalità	0,56	fenomeno	0,37
460	<i>Preoccupa</i> invece [MASK][MASK], nella capitale, il raduno previsto nel pomeriggio del giorno precedente dai neofascisti di Base autonoma nel quartiere Appio.	politici	autorità	unknown	0	situazione	0,51	preoccupazione	0,55

Tabella 18. Esempio di frasi con *preoccupare*.

Il verbo *pagare* ha un range di Oggetti tra i più ampi nell’annotazione manuale (22 lemmi). I modelli sono stati in grado di riconoscere la categoria semantica, restituendo parole astratte come *prezzo* (16%), *pena* e *danno* e nomi eventivi come *perdita*, *trasferimento* e *risarcimento*. Altre volte, l’output si riferisce a pagamenti veri e propri di tipo [Money], coinvolgendo un altro pattern del verbo (es. *soldi*, *pizzo*, *biglietto*, *multa*). Notiamo tuttavia che, nella maggior parte dei contesti, la sfumatura tra un pattern e l’altro è debole ed entrambi sono accettabili (es. tabella 19).

id	Frase	Oggetto GS		UmBERTo		ELECTRA	
		1	2	out	sim	out	sim
386	Non possiamo accettare l'assioma secondo il quale quando lo Stato va in default, per scelte di politica economica sballate, a <i>pagare</i> [MASK][MASK] deve essere sempre e solo la gente.	scelte	default	tassa	0,22	debito	0,24

Tabella 19. Esempio di frase con *pagare*.

In conclusione, i bassi punteggi di *similarity* ottenuti per questi verbi sono dovuti a: 1) nel caso di *preoccupare*, la confusione tra Oggetto e Soggetto, per cui spesso i modelli hanno completato lo spazio mascherato con un Soggetto post-verbale; 2) nel caso di *pagare*, la sovrapposizione con altri pattern più frequenti del verbo diversi dall'uso figurato, che coinvolgono entità concrete di tipo [Money].

5.2.6. *Verbs of Change of Possession*

All'interno di questa categoria sono stati inseriti i cinque verbi: *guadagnare*, *ordinare*, *servire*, *vincere* e *perdere*. Sono tutti verbi telici, che codificano al loro interno la conclusione dell'evento verbale. *Guadagnare*, *ordinare* e *servire* descrivono un passaggio di proprietà di soldi e cibo attraverso un evento durativo. *Vincere* e *perdere* invece indicano un evento simultaneo in cui il passaggio di proprietà è figurato. Come era già stato osservato da Fillmore (1986), l'uso intransitivo degli *Achievements* in italiano è un'eccezione possibile solo in contesti idiomatici o settoriali (cfr. [sez. 1.1.1.](#)). Infatti, i pattern di riferimento per questi due verbi sono usati in ambito sportivo o politico, dove reggono i tipi [Competition] e [War].

Tutti i verbi di questa classe sono stati annotati con un range limitato di Oggetti GS. In particolare, *guadagnare* regge solo *soldi* o *denaro*; il verbo è fortemente orientato verso il *Defaulting* lessicale (39/40 annotazioni) poiché non implica quasi mai un Oggetto ricostruibile dal contesto. Tra gli altri verbi, *perdere* è orientato verso il *Defaulting* pragmatico (35/40 annotazioni) per l'abbondanza di parole che rimandano allo stesso *frame* dell'Oggetto implicito e di riferimenti anaforici nel contesto. I punteggi di *similarity* su questa classe sono intermedi per i primi tre verbi (0,4-0,43), e molto alti per gli ultimi due (0,56-0,59). Inoltre, *perdere* è uno dei due verbi per cui i modelli ottengono un punteggio più alto nell'IMPLICIT dataset che nell'EXPLICIT dataset, insieme a *cantare* (v. [sez. 5.2.1.](#)).

Iniziamo la discussione di questa classe da *guadagnare*. Nel pattern proposto, il verbo regge Oggetti di tipo [Money]. I modelli hanno restituito diverse volte i due nomi codificati

nella radice verbale e proposti nell’annotazione, *soldi* (14%) e *denaro* (15%). Talvolta gli output indicano la valuta del denaro (es. *euro*, *bitcoin*), la quantità (es. *percentuale*, *centesimo*, *metà*) o la forma in cui il denaro si guadagna (es. *stipendio*, *salario*, *profitto*). Alcune volte, il verbo è stato ricondotto ad un altro pattern in cui assume il significato *ottenere*, *conquistare come risultato di un lavoro o di un impegno* e regge il tipo semantico [Attitude] o [State]; in questi casi, l’output può indicare entità astratte come *reputazione*, *possibilità* e *tempo* o concrete come *titolo* e *bonus*. Come osservato anche nel caso di *pagare*, un altro verbo che può coinvolgere uno scambio di denaro oppure essere usato in modo figurato con entità astratte (v. [sez. 5.2.5](#)), i due pattern sono spesso interscambiabili (es. tabella 20).

id	Frase	Oggetto GS		BERTino	
		1	2	out	sim
295	Come <i>si guadagna</i> [MASK][MASK] con il software libero?	soldi	denaro	reputazione	0,18

Tabella 20. Esempio di frase con *guadagnare*.

Notiamo che la polisemia è spesso un fattore di confusione negli output dei modelli, come è stato osservato finora per alcuni pattern problematici dei verbi *cantare*, *dirigere*, *suonare*, *disegnare* e *provare* ([sez. 5.2.1.](#)), *bere* ([sez. 5.2.2.](#)), *respirare* ([sez. 5.2.3.](#)) e *guidare* ([sez. 5.2.4.](#)). Talvolta gli errori interessano solo degli *shift* metonimici, altre volte vengono coinvolti pattern completamente incoerenti con la sfera semantica della frase. Questa differenza è legata alla produzione di frasi logicamente coerenti e deve essere considerata attentamente nella valutazione degli output.

Riprendendo la discussione, il verbo *ordinare* comprende due pattern semantici: nel primo regge il tipo [Artifact] e si usa soprattutto per le prenotazioni *online*, nel secondo regge i tipi [Food], [Beverage] e [Meal] e si trova in contesti di ristorazione. I modelli hanno saputo riconoscere il primo pattern in tutte le 8 frasi che lo contenevano, completandole nella maggior parte dei casi con il lemma *prodotto* (14%). Per quanto riguarda il secondo pattern, i risultati sono simili a quelli già discussi per i verbi *mangiare* e *bere* (cfr. [sez. 5.2.2.](#)) e comprendono sia nomi generici come *cibo*, *piatto* e *pasto*, sia nomi specifici come *caffè* (addirittura il 20% degli output) e *drink*.

Notevolmente, i modelli hanno distinto quasi sempre i due pattern di *ordinare*. Un caso interessante è la frase #367 (tabella 19); dalla logica della frase si capisce che vengono descritte delle azioni sequenziali: 1) sedersi al tavolo; 2) ordinare qualcosa; 3) ordinare il

conto, ed è ovvio che nel secondo passaggio si effettua un'ordinazione di cibi o bevande. Tuttavia i modelli non riescono a cogliere questa sequenzialità e completano con la parola *conto*. Solo BERTino completa l'Oggetto correttamente, riferendosi ad un'azione precedente, e restituisce *menu*.

id	Frase	Oggetto GS		UmBERTo		BERTino	
		1	2	out	sim	out	sim
367	Tralasciando per ragioni di spazio e di tempo il codice necessario per sedersi al tavolo ed <i>ordinare</i> [MASK][MASK], vediamo cosa succede quando diciamo: Cameriere, il conto!	pasto	pasto	conto	0,19	menu	0,37

Tabella 19. Esempio di frase con *ordinare*.

Il verbo *servire* regge gli stessi tipi semantici di *ordinare*, [Food] e [Meal], ed è usato in contesti di ristorazione. Anche in questo caso, sono stati riportati sia nomi generici come *piatto* (18%) e *cibo* (12%), sia nomi specifici come *vino* (5%) e *pollo*. In generale, i modelli e gli annotatori privilegiano il tipo semantico [Food] rispetto a [Beverage]; BERT-multi e BERT-it hanno una tendenza maggiore a restituire *carne* e *pesce* laddove gli altri modelli tendono a completare con *pasta*, *pane* e *pizza* (un'osservazione già avanzata a proposito di *cucinare*, [sez. 5.2.1.](#)). Il punteggio di *similarity* per questo verbo si abbassa leggermente perché spesso i modelli restituiscono output di tipo [Human], come *cliente* (5%), *famiglia* e *bambino* (es. tabella 20). Anche in questo caso, l'output non è da considerarsi sbagliato poiché i pattern sono interscambiabili.

id	Frase	Oggetto GS		BERT-multi		BERT-it		ELECTRA	
		1	2	out	sim	out	sim	out	sim
528	In America le attrici non si vergognano di <i>servire</i> [MASK][MASK] ai tavoli.	cibo	cibo	carne	0,52	clienti	0,29	bambini	0,38

Tabella 20. Esempio di frase con *servire*.

Gli ultimi due verbi di questa classe sono *vincere* e *perdere*. Nonostante i due pattern siano semanticamente speculari, i verbi sono diversi per due motivi principali: prima di tutto, *perdere* è più orientato verso il *Pragmatic Defaulting*, mentre *vincere* tende a codificare l'Oggetto nella radice lessicale; in secondo luogo, *perdere* comprende pattern più frequenti rispetto a quello di riferimento, quindi il rischio di sovrapposizione è maggiore.

Per quanto riguarda le frasi con *vincere*, i modelli restituiscono soprattutto output appartenenti ai due tipi semantici del pattern, tra cui *partita* (15%), *gara* (9%), *campionato* (8%) e *gioco* (6%) per il tipo [Competition] e *guerra* (6%) e *battaglia* (6%) per il tipo [War]. Alcune volte si trova *premio* (9%) che appartiene al tipo [Artifact], ma le frasi completate in questo modo sono corrette dal punto di vista logico. In generale, i risultati su questo verbo sono ottimi.

Per quanto riguarda il verbo *perdere*, i modelli ottengono un punteggio di *similarity* molto alto perché il 37% degli output corrisponde al lemma *partita*, il rappresentante prototipico della categoria [Competition], usato spesso anche nell’annotazione manuale (29/40 annotazioni). Per la stessa categoria, si trovano *derby* (3%), *campionato* (4%) e *match*. Tuttavia, l’alta polisemia del verbo ha creato alcuni problemi ai modelli; infatti essi hanno restituito anche parole che esulano dai contesti del pattern, come *treno*, *cappotto* e *soldi* oppure *concentrazione* e *memoria*, o ancora *primato*, *consenso*, *controllo* e *leadership*. Alcune volte hanno privilegiato espressioni idiomatiche che corrispondono a collocazioni verbo-Oggetto molto frequenti, come *perdere la testa* (3%) e *perdere la bussola*. Un esempio ambiguo, in cui anche l’annotatore 2 non ha saputo proporre un Oggetto, è la frase #420, riportata nella tabella 21; in questo caso, vista la mancanza di indizi contestuali, i modelli hanno restituito le collocazioni verbo-Oggetto più frequenti, che richiamano, appunto, usi idiomatici del verbo.

id	Frase	Oggetto GS		BERT-it		UmBERTo		BERTino		ELECTRA	
		1	2	out	sim	out	sim	out	sim	out	sim
420	l'ironia non è una via d'uscita: è solo un modo educato di <i>perdere</i> [MASK][MASK].	dibattito	unknown	controllo	0,15	testa	0,04	Bussola	0,05	memoria	0,1

Tabella 21. Esempio di frase con *perdere*.

Anche nell’EXPLICIT dataset i punteggi si sono abbassati per via dell’alta polisemia del verbo e della forza delle collocazioni in altri pattern, come *perdere le tracce* e *perdere la vita* (esempi nella tabella 22).

id	Frase	Oggetto GS	BERT-multi		BERT-it		UmBERTo		BERTino		ELECTRA	
			out	sim	out	sim	out	sim	out	sim	out	sim
402	Sapete a volte perchè <i>perdevo</i> le [MASK] per pochi secondi?	gare	parti	0,41	cose	0,38	tracce	0,31	puntate	0,55	parti	0,41
404	Odifreddi: Gli americani <i>persero</i> la [MASK].	guerra	guerra	1	vita	0,37	vita	0,37	vita	0,37	vita	0,37

Tabella 22. Esempio di frasi con *perdere* (EXPLICIT DATASET)

In conclusione, i modelli esibiscono comportamenti differenti per i cinque verbi di questa classe. Nel caso di *guadagnare*, si è verificata una sovrapposizione con un altro pattern, nel quale il verbo assume il significato di *ottenere* e viene utilizzato per riferirsi a entità astratte. Per quanto riguarda *servire* e *ordinare*, i modelli hanno saputo riconoscere con successo l’ambito semantico, mostrando solo talvolta qualche problema a ricostruire le entità specifiche attraverso un ragionamento logico-deduttivo. Infine, i verbi *perdere* e *vincere* sono stati affrontati con un ottimo risultato; tuttavia, nel caso di *perdere*, la polisemia del verbo ha causato talvolta la confusione con pattern idiomatici come *perdere la testa*.

5.2.7. Verbs of Existence

I *Verbs of Existence* comprendono i verbi *consumare* e *attendere*. *Consumare* è usato nel senso di *impiegare*, *sfruttare* e regge le categorie [Energy], [Gas] e [Inanimate]. *Attendere* è usato con il significato di *aspettare* e regge la categoria [Event]³³. Il punteggio di *similarity* dei modelli è medio per *consumare* (0,43), mentre è molto basso per *attendere* (0,23). Per quanto riguarda l’annotazione manuale, il range di Oggetti GS di *consumare* è limitato a sei elementi (*risorse* (20/40 annotazioni) *energia*, *calorie*, *benzina*, *combustibile* e *carburante*) mentre il range di *attendere* è il più ampio in tutto il dataset e comprende 24 lemmi, un dato correlato con l’annotazione prevalente di *Pragmatic Defaulting* (35/40 annotazioni).

Iniziamo la discussione da *consumare*, per cui i modelli hanno funzionato molto bene. Tra gli output più frequenti troviamo *energia* (19%), *carburante* (12%), *denaro* (6%) e *motore* (4%). L’unico output che potrebbe rimandare ad un altro pattern è *cibo*, con cui il verbo può assumere il significato di *mangiare*, tuttavia questo lemma è stato restituito poche volte (es. frase #153, tabella 23).

³³ Su T-PAS questo pattern regge anche la subordinata oggettiva ‘che + [Event]’, che nella nostra Ontologia è stata esclusa perché il task prevede di completare lo spazio mascherato con un solo sostantivo.

id	Frase	Oggetto GS		UmBERTo	
		1	2	out	sim
153	Un invito a risparmiare e <i>consumare</i> meno [MASK][MASK] per tutta la giornata.	risorse	risorse	cibo	0,23

Tabella 23. Esempio di frase con *consumare*.

Per quanto riguarda il verbo *attendere*, molte volte non è stato trovato un sostantivo per completare la parola mascherata ed è stato restituito il token *unknown* (14%). Potrebbe essere perché i nomi eventivi sono più difficili da prevedere rispetto ad altri nomi. In molti casi, i modelli hanno restituito lemmi come *giorno* (9%), *minuto* (8%), *attimo* (3%) e *tempo*, che costituiscono Complementi di Tempo piuttosto che Complementi Oggetto, un problema già visto a proposito di altri verbi (es. *giocare*, [sez. 5.2.4.](#)). Altre volte, sono stati restituiti i tipi semantici [Vehicle] (es. *pullman* e *treno*), [Human] (es. *medico* e *personaggio*) e [Physical Entity] (es. *sole*); in tutti questi casi, il verbo può reggere un sostantivo appartenente a un diverso tipo semantico perché esercita una *semantic coercion* su di esso, reinterpretando l'Oggetto come [Event] grazie ad alcune caratteristiche codificate nella sua semantica (Pustejovsky e Jezek, 2008). I modelli tendono a replicare un'operazione di questo tipo, mentre gli annotatori privilegiano nomi eventivi deverbali come *avvio*, *caricamento* e *pagamento*, perché sono più prototipici del tipo semantico indicato nel pattern (es. frase #22, tabella 24).

id	Frase	Oggetto GS		UmBERTo		BERTino		ELECTRA	
		1	2	out	sim	out	sim	out	sim
22	A seguito della riforma del CNR e del conseguente commissariamento, la Commissione decise peraltro di <i>attendere</i> [MASK][MASK] prima di ascoltare il professor De Maio onde consentirgli di avviare la fase di transizione.	avvio	transizione	giorno	0,2	minuto	0,11	riunione	0,41

Tabella 24. Esempio di frase con *attendere*.

In conclusione, i modelli hanno raggiunto risultati diversi su questi due verbi; mentre per *consumare* hanno riconosciuto il tipo semantico corretto, per *attendere* hanno privilegiato meccanismi di *semantic coercion* laddove gli annotatori hanno annotato nomi deverbali, causando un abbassamento del punteggio di *similarity*.

5.2.8. Altri verbi

Le rimanenti sette classi semantiche comprendono un solo verbo ciascuna, quindi, per comodità, sono raggruppate sotto questa sezione. I verbi saranno discussi in ordine di punteggio di *similarity*, dal più alto al più basso. L'elenco è riportato qui di seguito: *tirare* (*Verbs of Throwing*) nel significato di *calciare la palla*; *leggere* (*Lean Verbs*); *combattere* (*Verbs of Social Interaction*); *ascoltare* (*Verbs of Perception*); *pregare* (*Verbs of Desire*); *chiamare* (*Verbs of Communication*) nel significato di *contattare attraverso apparecchio telefonico* e infine *condurre* (*Verbs with Predicative Complements*) nel significato di *essere presentatore di*.

Per quanto riguarda l'annotazione manuale, il verbo *tirare* si distingue per la tendenza verso il *Lexical Defaulting* e l'ampiezza del range, poiché regge solamente tre Oggetti GS: *palla* (36/40 annotazioni), *pallina* e *pallone*. L'IAA su questo verbo è il più alto (0,95) grazie alla prototipicità del lemma *palla* rispetto alla categoria [Ball]. Anche il punteggio dei modelli è il più alto (0,64) e dimostra che il pattern di riferimento è stato riconosciuto facilmente. All'estremo opposto, troviamo *condurre*, che ha una fortissima tendenza verso il *Pragmatic Defaulting* (39/40 annotazioni) e un range di Oggetti di 22 elementi. Gli annotatori raggiungono un IAA abbastanza basso sull'annotazione dell'Oggetto (0,43) (cfr. anche [sez. 2.3.2.](#) per la discussione dell'annotazione) e i modelli funzionano peggio che su ogni altro verbo (0,22).

Iniziamo allora dal verbo su cui i modelli hanno raggiunto il punteggio più alto, *tirare* (*Verbs of Throwing*). Si tratta di un verbo altamente polisemico, per cui T-PAS riporta 33 pattern; in particolare, il verbo è utilizzato frequentemente in espressioni idiomatiche (es. *tirare avanti*) e in costruzioni frasali (es. *tirarsi su*). Nelle frasi del dataset ha sempre il significato tecnico di *calciare la palla* e regge il tipo semantico [Ball]. Insieme a *vincere* e *perdere*, appartiene alla classe azionale degli *Achievements*, su cui i modelli hanno raggiunto i punteggi di *similarity* più alti.

Tra gli output dei modelli, si registrano soprattutto *palla* (20%), *pallone* (26%) e *pallina* (1%). Si trovano poi delle parole affini al tipo [Ball], come *tiro* (18%), *gol* (4%), *rigore* (2%), *testa* (2%) e *calcio* (5%). In particolare, *tiro* è uno *Shadow Argument*, cioè un Argomento codificato nella semantica del verbo che non può essere esplicitato senza ulteriori specificazioni (Jezek, 2018): risulta infatti agrammaticale la frase **“tira un tiro”*, ma è più accettabile *“tira un tiro lungo”*. I modelli non sembrano essere sensibili a tale limite

linguistico, come discusso anche per altri verbi, come *respirare* (*respiro* (18%), v. [sez. 5.2.3.](#)) e *disegnare* (*disegno* (17%), v. [sez. 5.2.1.](#)).

Infine, per le frasi #561 e #563 (tabella 25), sono stati restituiti output che rimandano ad usi idiomatici del verbo: *cinghia* (*tirare la cinghia*), *fiato* (*tirare il fiato*), *corda* (*tirare la corda*) e *somma* (*tirare le somme*). L'output di BERT-it (*tira*) *il fiato* nella frase #563 potrebbe essere logicamente corretto nel caso in cui la prima frase fosse una conseguenza della seconda: la partita è stata molto seguita, quindi l'azzurro *tira il fiato*. Negli altri casi, i modelli non colgono il tipo semantico implicato dal verbo e completano lo spazio mascherato solo sulla base della frequenza della collocazione. Tuttavia, la *performance* dei modelli su questo verbo in generale è molto buona, grazie all'uso fortemente settoriale del pattern e al numero ridotto di elementi che può implicare in questo uso.

id	Frase	Oggetto GS		BERT-multi		BERT-it		UmBERTo		BERTino	
		1	2	out	sim	out	sim	out	sim	out	sim
561	Gara chiusa, che si rassegna a <i>tirare</i> [MASK][MASK] al 90' senza eccessivi sforzi?	palla	palla	gara	0,39	fiato	0,29	rigore	0,42	corda	0,53
563	Eppure, nonostante tutto, l'azzurro <i>tira</i> [MASK][MASK]: la partita di mercoledì è stata seguita da quasi 12 milioni e mezzo di tele-spettatori.	palla	palla	calcio	0,38	fiato	0,29	Cinghia	0,22	somma	0,17

Tabella 25. Esempio di frasi con *tirare*.

Leggere (*Lean Verbs*) è un *Activity* dal comportamento analogo ai *Verbs of Creation and Transformation* (v. [sez. 5.2.1.](#)). Infatti, nell'annotazione manuale è perfettamente equilibrato tra i due tipi di *Defaulting* e ha un range di Oggetti medio. Il rappresentante adottato indipendentemente dagli annotatori per il tipo [Document] è *testo* (15/40), poi si trovano soprattutto *libro* (9/40) e *messaggio* (3/40).

I modelli hanno restituito parole del tipo [Document] nella maggior parte dei casi, tra cui *libro* (19%), *giornale* (11%), *testo* (9%), *articolo* (4%) e *brano* (3%). Dove c'era un riferimento alla sfera della messaggistica online, i modelli hanno restituito output come *commento*, *blog*, *messaggio* e *sito*. Poche volte gli output non rientrano nella categoria, ma sono collegate ad essa attraverso relazioni metonimiche, per esempio *Latino* e *dialetto*

descrivono la lingua del documento, *profeta* indica l'autore del documento, *riga* e *parola* (5%) indicano parti del documento (es. tabella 26).

id	Frase	Oggetto GS		BERT-multi		BERTino	
		1	2	out	sim	out	sim
328	E mentre la maestra <i>leggeva</i> [MASK][MASK]: ' Il vento si alzò forte nel cielo, fino a toccare le nuvole e a spostarle con un lieve movimento della sua bocca...'	storia	racconto	Latino	0,08	parole	0,18

Tabella 26. Esempio di frase con *leggere*.

Passiamo al verbo *combattere* (*Verbs of Social Interaction*). In Levin (1993), *fight* è indicato come verbo con *Understood Reciprocal Object Alternation*, poiché nella sua lettura intransitiva è usato frequentemente con il significato di *combattersi l'un l'altro*, con un Soggetto plurale, come nella frase “i soldati combattono” (cfr. anche [sez. 2.1.](#)). Nel nostro dataset, invece, è utilizzato con il significato di *affrontare* e regge i tipi semantici [War] e [Competition]. Nell'annotazione manuale, il primo tipo semantico è rappresentato dai lemmi *guerra* (13%), *battaglia* (9%), *duello* (3%) e *scontro* (2%), il secondo da *incontro* (5%), *match* e *competizione*; inoltre, in due occasioni, il secondo annotatore ha annotato *nemico* e *avversario*, con riferimento ad un altro pattern del verbo.

I modelli hanno avuto difficoltà a completare la posizione mascherata con un sostantivo e hanno restituito un numero alto di *unknown* (10%). Hanno riconosciuto con più facilità il primo tipo semantico, restituendo output come *guerra* (14%), *battaglia* (11%), *combattimento* (4%), *conflitto* (4%) e *lotta* (4%), mentre non hanno riconosciuto quasi mai il secondo tipo semantico, [Competition]. In entrambi i casi, si è verificata l'influenza di altri pattern del verbo, dovuta alla sua polisemia. Talvolta i modelli hanno restituito output di tipo [Human], come *nemico* (8%), *campione*, *alieno* e *signore*, altre volte hanno completato lo spazio mascherato con parole astratte di tipo [Eventuality] o [Concept], con le quali il verbo viene interpretato nel significato di *contrastare*, per esempio *cancro*, *paura*, *problema*, *panico*, *violenza* e *terrorismo* (es. tabella 27). L'alta polisemia del verbo ha provocato errori specialmente quando la comprensione del pattern richiedeva capacità logiche di elaborazione delle informazioni.

id	Frase	Oggetto GS		UmBERTo		ELECTRA	
		1	2	out	sim	out	sim
111	Scontri anche nell'isola di Mindanao, dove si combatte [MASK][MASK] ormai da una settimana.	guerra	guerra	terrorismo	0,54	violenza	0,5

Tabella 27. Esempio di frase con *combattere*.

Il prossimo verbo, *ascoltare* (*Verbs of Perception*), nell'annotazione manuale è fortemente orientato verso il *Pragmatic Defaulting* (33/40 annotazioni) e ha un range di Oggetti abbastanza alto (17 elementi). Comprende due pattern: nel primo regge i tipi [Sound], [Musical Performance] e [Speech Act]; nel secondo regge il tipo [Human] e significa *avere capacità di ascolto non solo uditivo, ma anche di comprensione nei confronti di*. La varietà di Oggetti annotati è grande, anche per via dei molti tipi semantici implicati, e include soprattutto *suono* (10%), rappresentante prototipico della categoria [Sound], e *musica* (6%). Nelle annotazioni troviamo anche nomi come *radio* e *tv*, che sono in relazione metonimica con la categoria [Sound] e coinvolgono il meccanismo di *semantic coercion* già accennato per altri verbi (es. *attendere*, [sez. 5.2.7.](#), e *respirare*, [sez. 5.2.3.](#)).

Gli output dei modelli includono parole delle categorie semantiche di riferimento, come *musica* (9%), *brano* (12%), *discorso* (6%); in alcuni casi evidenziano relazioni metonimiche, come per esempio nella frase #17 (tabella 27), dove tutti i modelli (tranne BERT-multi) hanno completato la parola mascherata con *pianoforte*. Notiamo che l'output *pianoforte*, accettabile dal punto di vista logico e semantico, ha ricevuto un punteggio di *similarity* medio-alto (0,54), un dato che dimostra come la misura sia adatta a cogliere la vicinanza degli output dei modelli alla classe semantica di riferimento. Nei casi di *Pragmatic Defaulting* dove la comprensione dell'Oggetto richiede abilità logiche, i modelli hanno favorito le collocazioni più frequenti e ottenuto risultati logicamente incoerenti (es. #16, tabella 28).

id	Frase	Oggetto GS		BERT-it		UmBERTo	
		1	2	out	sim	out	sim
17	Se c'è un pianista che vale la pena di ascoltare, insomma, qualsiasi cosa voglia suonare, ebbene è Keith Jarrett, capace di commuovere e di esaltare, in grado di inventare inedite figure musicali ed al tempo stesso in grado di far viaggiare chi <i>ascolta</i> [MASK][MASK] in una straordinaria macchina sonora che non conosce limiti di tempo e di spazio.	musica	musica	pianoforte	0,54	pianoforte	0,54
16	Hang appoggiò un orecchio sul petto di Romero ed <i>ascoltò</i> [MASK][MASK] con profondo raccoglimento.	respiro	battito	discorso	0,41	discorso	0,41

Tabella 28. Esempio di frasi con *ascoltare*.

Continuando la discussione dei verbi in ordine di punteggio, ci occupiamo del verbo *pregare* (*Verbs of Desire*). Tale verbo nell'Ontologia comprende due pattern semantici: nel primo significa *rivolgersi a* e regge come Oggetto il tipo semantico [Deity], nel secondo significa *invitare gentilmente* e regge il tipo semantico [Human]. Nel secondo significato, è utilizzato spesso nella costruzione impersonale *si prega* [Human] *di*; inoltre, analogamente a *preoccupare* (v. [sez. 5.2.5.](#)), esibisce una *PRO-arbitrary Object Alternation* con interpretazione dell'Oggetto arbitraria, l'omissione del quale è possibile solo in contesti imperfettivi (cfr. [sez. 1.1.3.](#), pag. 14).

In generale questo verbo ha dato problemi ai modelli, che hanno restituito *unknown* nel 20% dei casi. Anche sull'EXPLICIT dataset la *similarity* per questo verbo è tra le più basse (0,47). Nelle frasi contenenti il primo pattern del verbo, gli annotatori hanno proposto come GS soprattutto *Dio* (13/40 annotazioni) e *divinità* (7/40 annotazioni). I modelli hanno invece privilegiato *Signore* (13%), raramente *Dio* (3%) e *Vergine* (1%). Dove gli annotatori hanno proposto le parole *YHWH* (il dio ebraico) e *Allah* (il dio islamico), i modelli non hanno colto la sfumatura semantica (v. tabella 29).

Nella maggior parte dei casi, i modelli hanno completato la posizione mascherata con altri tipi semantici. Talvolta gli output consistono in parole di ambito religioso, come il mezzo usato per pregare (es. *preghiera* (7%), di nuovo uno *Shadow Argument*) o il motivo per cui si prega (es. *misericordia*, *salute*, *pace*, *scuse*). Altre volte, i modelli hanno restituito un'espressione temporale come *giorno* (5%), *domenica* o *minuto*. Nonostante il mancato riconoscimento del tipo semantico di riferimento, [Deity], l'ambito discorsivo è stato

riconosciuto in tutti i casi, con pochissime eccezioni in cui si è verificata una sovrapposizione con il secondo pattern (es. l'output di ELECTRA per la frase #432, *gente*, nella tabella 29).

id	Frase	Oggetto GS		BERT-it		UmBERTo		ELECTRA	
		1	2	out	sim	out	sim	out	sim
430	Non ricordo che gli altri Ebrei del campo <i>pregassero</i> [MASK][MASK], nè io, che pure avevo avuto precedentemente una importante esperienza religiosa (protestante evangelica) nel campo non ho mai pregato	Dio	YHWH	Signore	0,31	unknown	0	domenica	-0,06
432	Qui il Profeta fece disporre un muro a secco tutt 'in giro a una palma, al fine di separare 'da ciò che è impuro 'la gente venuta ad ascoltarlo e, quindi, a meditare, e infine a <i>pregare</i> [MASK][MASK] con lui in quel recinto che, se vogliamo, fu la prima moschea dell'islàm.	Allah	Allah	Dio	0,5	pace	0,21	gente	0,08

Tabella 29. Esempio di frasi con *pregare*.

Per quanto riguarda il secondo pattern di *pregare*, gli annotatori hanno annotato principalmente *persone* (4/40) e *utenti* (4/40). Il tipo semantico [Human] è stato riconosciuto quasi sempre dai modelli; tra gli output più frequenti si registrano *cliente* (5%), *sindaco* (5%), *docente* (3%), *cittadinanza* (3%) e *lettore* (3%). Le frasi con questo pattern non sono mai state ricondotte ad un pattern diverso, principalmente grazie alla tipicità della struttura impersonale (es. tabella 30).

id	Frase	Oggetto GS		BERT-it		UmBERTo	
		1	2	out	sim	out	sim
436	Data dell'informazione: Fino al 1984 AVVERTENZA: <i>Si prega</i> [MASK][MASK] di notare che questa fonte è sensibile.	persone	utente	direzione	0,33	lettore	0,5

Tabella 30. Esempio di frase con *pregare*.

Passando al verbo *chiamare* (*Verbs of Communication*), esso è utilizzato nel dataset con il significato di *contattare mediante apparecchio telefonico* e regge i tipi semantici [Human] e

[Institution]. La *similarity* sull'IMPLICIT dataset è molto bassa (0,27), mentre sull'EXPLICIT dataset è decisamente più alta (0,65) per la maggiore ricchezza informativa del contesto frasale. L'annotazione manuale mostra una tendenza verso il *Pragmatic Defaulting* (31/40 annotazioni) e un range di Oggetti medio-alto (16 elementi).

Il punteggio di *similarity* tra annotatori per questo verbo è tra i più bassi (0,5). Ciò è dovuto a tre motivi principali: 1) l'Oggetto Implicito di *chiamare* spesso deve essere ricostruito dal contesto extra-frasale; 2) l'Oggetto Implicito spesso consiste in un pronome personale (es. “oggi (li) chiamo e organizziamo”) piuttosto che in un sostantivo; 3) quando l'Oggetto Implicito appartiene al tipo semantico [Human], è particolarmente complicato trovare nomi generici e plurali che fungano da Oggetto, sia quando la semantica del verbo richiede un completamento arbitrario (es. per i verbi *preoccupare*, [sez. 5.2.5](#) e *pregare* nel secondo pattern, [sez. 5.2.8](#), che hanno IAA molto basso), sia quando il Complemento Oggetto Implicito è riferito ad un'entità specifica, come in questo caso.

Gli annotatori hanno cercato di ovviare a questi problemi annotando parole come *persona* (5/40 annotazioni), *negozio* (4/40 annotazioni), *servizio* (4/40 annotazioni) e *numero* (3/40 annotazioni); in particolare, quest'ultima parola esula dal tipo semantico di riferimento ma svincola dalla necessità di ricostruire un'entità specifica. Anche i modelli hanno adottato la stessa strategia, restituendo parole come *telefono* (15%), *numero* (5%), *linea* e *segreteria*. Oltre a questi, tra gli output dei modelli si registrano nomi specifici, perlopiù assegnati arbitrariamente, come *medico*, *dottore*, *ragazzo*, *padre*, *dottore*, *venditore*, *gestore* e, per il secondo tipo semantico, *polizia*, *ministero*, *sito*, *concessionario*. Infine, alcuni output appartengono al tipo [Vehicle], come *taxi* (4%) e *macchina* (3%). Alcuni esempi sono riportati nella tabella 31.

id	Frase	Oggetto GS		BERT-it		UmBERTo		BERTino	
		1	2	out	sim	out	sim	out	sim
82	A decine <i>hanno chiamato</i> [MASK][MASK] per esprimergli solidarietà e anche per offrirgli ospitalità nelle più belle zone di mare della Sardegna.	amico	persona	comune	0,15	familiare	0,4	sindaco	0,2
86	Da Orrigoni prova a <i>chiamare</i> [MASK][MASK] e ti danno un preventivo.	negozio	negozio	professionista	0,28	rivenditore	0,6	taxi	0,31
90	Puntuale, il primo <i>chiamò</i> [MASK][MASK] e subito le disse che voleva portarla a ballare.	donna	donna	ragazzo	0,6	ragazzo	0,6	ragazzo	0,6

Tabella 31. Esempio di frasi con *chiamare*.

L'ultimo verbo del dataset è *condurre* (*Verbs with Predicative Complements*), usato nel significato di *essere presentatore di* con i tipi semantici [TV Program] e [Activity]. Il verbo è quello su cui i modelli hanno ottenuto una *similarity* più bassa e anche quello con il maggior numero di *Pragmatic Defaulting*. Anche l'IAA agreement è basso (0,43), mentre i punteggi si alzano leggermente sull'EXPLICIT dataset (0,57).

Il problema principale è stato che nella maggior parte dei casi i modelli hanno completato lo spazio mascherato con un Soggetto post-verbale, probabilmente a causa dell'alta frequenza di questa costruzione sintattica con il pattern in questione (es. "conduce Mario Rossi")³⁴. Questo è un problema già visto per altri verbi (es. *bere*, [sez. 5.2.2.](#), *guidare*, [sez. 5.2.4.](#) e *preoccupare*, [sez. 5.2.5.](#)). In tutte le frasi era presente un Soggetto post-verbale; per questo motivo, i modelli hanno restituito quasi sempre un predicato nominale che indica la professione del Soggetto³⁵. Si registrano quindi moltissimi output come *giornalista* (29%), *conduttore* (12%), *maestro* (5%), *regista* (4%) e *professore* (3%) (es. tabella 32). In pochissimi casi è stato restituito un Oggetto della categoria indicata nel pattern, come *dibattito* (8%), *serata* (3%) e *trasmissione* (2%).

³⁴ È possibile che il pattern si trovi solo con questa costruzione.

³⁵ La classe semantica *Verbs with Predicative Complements* in Levin (1993) è caratterizzata proprio dal complemento predicativo.

id	Frase	Oggetto GS		BERT-it		UmBERTo		BERTino	
		1	2	out	sim	out	sim	out	sim
131	In onda alle 23.15, <i>condurre</i> [MASK][MASK] Maurizio Mannoni	programma	programma TV	giornalista	0,23	conduttore	0,28	regista	0,19

Tabella 32. Esempio di frase con *condurre*.

La discussione sul verbo *condurre* dimostra che il basso punteggio di *similarity* in questo caso non dipende dal tipo di *Defaulting* ma dalla struttura sintattica associata al pattern, dove il Soggetto tende a trovarsi in posizione post-verbale, e nello specifico da una tendenza dei modelli a completare la posizione precedente un nome proprio con un predicato nominale che ne indica la professione.

Conclusioni

Il presente lavoro di tesi ha permesso di avanzare alcune osservazioni di natura linguistica sul comportamento di cinque modelli BERT italiani nel task di completamento dell'Oggetto Implicito. Per verificare l'ipotesi iniziale, secondo cui i contesti impliciti sarebbero più ricchi dal punto di vista informativo (v. [sez. 1.2.2.](#)), è stato istituito un EXPLICIT dataset di confronto. I risultati ottenuti smentiscono l'ipotesi di partenza, poiché si registrano punteggi di *similarity* peggiori nei contesti impliciti. Questo dimostrerebbe che i modelli ricostruiscono con più successo un Oggetto originariamente espresso rispetto ad uno Implicito. Tuttavia, questo risultato potrebbe essere influenzato dalla ricchezza morfo-sintattica dell'EXPLICIT dataset.

Un risultato interessante emerge dall'annotazione manuale, dove si dimostra una correlazione significativa tra il range di Oggetti Impliciti dei verbi e i due tipi di omissione dell'Oggetto, *Lexical Defaulting* e *Pragmatic Defaulting* (Jezek, 2018) (cfr. [sez. 1.1.1.](#)). È stato dimostrato che i verbi tendenti al *Lexical Defaulting* hanno un range di Oggetti più ristretto, in quanto codificano l'Oggetto Implicito all'interno della radice lessicale, mentre i verbi più orientati al *Pragmatic Defaulting* hanno un range di Oggetti più ampio, poiché l'Argomento è codificato nel contesto pragmatico e deve essere ricavato da questo di volta in volta (v. [sez. 2.3.2.](#)).

I modelli indovinano con più facilità l'Oggetto quando è presente un *Lexical Defaulting*, poiché possono affidarsi alla forza della collocazione verbo-Oggetto, come per esempio nel caso di *tirare la palla* oppure *respirare l'aria*. Il modello che completa più facilmente lo spazio mascherato con un Oggetto prototipico del tipo semantico di riferimento è BERT-it, motivo per cui è anche il modello con il risultato migliore sull'IMPLICIT dataset. Al contrario, il più leggero, BERTino, che raggiungeva un punteggio molto buono sull'EXPLICIT dataset, funziona peggio degli altri modelli nei contesti impliciti, forse per la tendenza a completare un Oggetto *lexically defaulted* con più variazione.

Tra i verbi scelti per l'Ontologia, quelli per cui i modelli e gli annotatori raggiungono risultati migliori sono verbi settoriali, usati in contesti molto marcati: *tirare*, *vincere*, *perdere*, *giocare* e *dirigere*. In questi verbi, la sovrapposizione con altri pattern si è verificata di meno grazie alla specificità del contesto discorsivo in cui vengono usati. I verbi come *mangiare* e *scrivere*, di tipo azionale *Activities* con duplice interpretazione dell'Oggetto Implicito (v. [sez. 1.1.3.](#)), hanno ottenuto punteggi medio-alti e costituiscono una classe uniforme per quanto riguarda il trattamento automatico dell'Oggetto.

I modelli hanno dimostrato in generale ottime capacità di riconoscere il contesto d'uso del verbo, con la sola eccezione di alcuni pattern meno frequenti a cui sono stati sovrapposti pattern più comuni, come si verifica per *bere* nel significato di *assumere bevande alcoliche*, per esempio nella frase “i ragazzi bevono [acqua] per essere più socievoli, disinibiti ed euforici.” (v. [sez. 5.2.2.](#), tabella 11).

Spesso hanno restituito come completamento della posizione mascherata parole in relazioni metonimiche con il Gold Standard, come per esempio *bicchiere* per *vino* (v. [sez. 5.2.2.](#)) oppure *disegno* per *immagine* (v. [sez. 5.2.1.](#)). Interessante la tendenza a riprodurre il meccanismo di *semantic coercion* (Pustejovsky e Jezek, 2008), cioè a restituire parole di un tipo semantico diverso da quello indicato nel pattern, che il verbo può reggere grazie ad una forzatura semantica, facendo emergere alcune informazioni intrinseche al sostantivo, per esempio nell'espressione *respirare il mondo*, dove la parola *mondo* contiene un'informazione sul tipo [Vapour] (v. [sez. 5.2.3.](#)).

Un errore notevole e ricorrente dei modelli è l'esplicitazione di uno *Shadow Argument* codificato nel verbo che, non essendo ulteriormente specificato, rende la frase agrammaticale. Ciò si verifica con vari verbi, tra cui *respirare (un respiro)*, *disegnare (un disegno)*, *fumare (il fumo)*, *pregare (una preghiera)*, *tirare (un tiro)* e dimostra un'insensibilità dei modelli BERT a questo limite linguistico.

Laddove i modelli non potevano indovinare l'Oggetto poiché non era deducibile dal contesto frasale, hanno privilegiato le collocazioni più frequenti, facendo emergere dei bias intrinseci nella loro conoscenza del mondo; per esempio, è stata osservata la tendenza dei grandi modelli BERT-it e BERT-multi a restituire i lemmi *carne* e *pesce* nei contesti di alimentazione o ristorazione, dove invece BERTino e UmBERTo privilegiano i lemmi *pane*, *pasta* e *pizza*.

Dal comportamento dei modelli emerge anche una tendenza a confondere il pattern di riferimento con altri pattern idiomatici dove la forza della collocazione è più forte, come per esempio in *tirare il fiato* e *perdere la testa*, con il risultato che vengono create frasi illogiche. Anche dove è necessario identificare relazioni più complesse tra i partecipanti all'evento verbale, per esempio quando il Soggetto applica delle restrizioni semantiche sull'Oggetto, i modelli tendono a basarsi semplicemente sulla tipicità delle collocazioni verbo-Oggetto senza comprendere appieno queste relazioni e restituendo delle frasi logicamente incoerenti, per esempio nella frase “il postino suona [il pianoforte] sempre due volte” (v. [sez. 5.2.1.](#), tabella 9).

In conclusione, i risultati confermano le osservazioni già avanzate in studi precedenti, secondo cui i modelli BERT tendono ad affidarsi alla frequenza delle collocazioni e degli *n-grams* anche nelle frasi dove sarebbe necessaria una comprensione più profonda e un'elaborazione semantica più complessa, con il rischio di produrre dei risultati illogici, violare dei limiti linguistici o alimentare dei bias intrinseci (cfr. [sez. 1.2.2.](#)).

A livello metodologico, la discussione sui risultati ha dimostrato che il punteggio di *similarity* è una misura approssimativa per valutare questo task, tuttavia in molti casi riesce a descrivere la vicinanza dell'output alla categoria semantica di riferimento.

Alcuni errori dei modelli sono dovuti al design dell'esperimento, in particolare, i completamenti di natura temporale, come *la sera*, e i soggetti o complementi predicativi post-verbali, come *il giornalista*. Tali errori si sono verificati perché non è stato impostato un filtro per isolare la relazione sintattica dell'output con il verbo, un miglioramento che potrebbe affinare l'esperimento. Un altro limite metodologico è stato la lemmatizzazione, che in vari casi ha prodotto output scorretti come *sigarettare* per *sigarette*, abbassando il punteggio di *similarity*.

Infine, è importante menzionare che altre famiglie di modelli rilasciate dopo BERT potrebbero completare con successo questo task. Una possibile e interessante estensione di questo progetto riguarderebbe l'applicazione dell'esperimento ai modelli della famiglia GPT.

Riferimenti bibliografici

- Albertin, G., Miaschi, A. e Brunato, D. (2021). On the Role of Textual Connectives in Sentence Comprehension: A New Dataset for Italian. In *Proceedings of the Eighth Italian Conference on Computational Linguistics CLiC-it 2021*.
- Apidianaki, M. e Soler, A. G. (2021). ALL Dolphins Are Intelligent and SOME Are Friendly: Probing BERT for Nouns' Semantic Properties and their Prototypicality. In *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, Punta Cana, Dominican Republic: Association for Computational Linguistics, pp. 79-94.
- Baker, C.F. (2014). FrameNet: A Knowledge Base for Natural Language Processing. In *Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929-2014)*, Baltimore, MD, USA: Association for Computational Linguistics, pp. 1-5.
- Baroni, M. e Kilgariff, A. (2006). Large linguistically-processed web corpora for multiple languages. In *EACL '06: Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations*, Trento, Italia: Association for Computational Linguistics, pp. 87-90.
- Cennamo, M. (2017). Object omission and the semantics of predicates in Italian in a comparative perspective. In L. Hellan, A. L. Malchukov e M. Cennamo (eds), *Contrastive Studies in Verbal Valency*, Amsterdam: Benjamins, pp. 251-273.
- Cheng, P. e Erk, K. (2018). Implicit Argument Prediction with Event Knowledge. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, New Orleans, Louisiana: Association for Computational Linguistics, pp. 831-840.
- Chiari, I. (2002). La procedura cloze, la ridondanza e la valutazione della competenza della lingua italiana. In *Italica*, 79/4, pp. 525-540.
- Devlin, J., Chang, M.W., Lee, K. e Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, Minneapolis, Minnesota: Association for Computational Linguistics, pp. 4171-4186.
- Ettinger, A. (2019). What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. In *Transactions of the Association for Computational Linguistics*, 8, pp. 34-48.

- Fillmore, C.J. 1982. Frame Semantics. In *Cognitive Linguistics. Basic Readings*, 34, pp. 373-400.
- Fillmore, C. J. (1986). Pragmatically Controlled Zero Anaphora. In *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, pp. 95-107.
- Firth, J. R. (1957). Modes of Meaning. In *Papers in Linguistics, 1934-1951*, London: Oxford University Press, pp. 190-215.
- Gagliardi, G. (2018). Inter-Annotator Agreement in linguistica: una rassegna critica. In *Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-it 2018*.
- Gerber, M. e Joyce, C. (2010). Beyond NomBank: A study of implicit arguments for nominal predicates. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, pp. 1583-1592.
- Goldberg, A.E. (2005). Argument realization: The role of constructions, lexical semantics and discourse factors. In J. Östman e F. Mirjam (eds.), *Construction Grammars. Cognitive Grounding and Theoretical Extensions [Constructional Approaches to Language 3]*, Amsterdam: John Benjamins, pp. 17-43.
- Harris, Z. S. (1954). Distributional Structure. In *Word*, 10/2-3, pp. 146-162.
- Hovy, E. e Lavid, J. (2010). Towards a ‘science’ of corpus annotation: A new methodological challenge for corpus linguistics. In *International Journal of Translation Studies*, 22, pp. 13-36.
- Jezek, E. (2003). *Classi di verbi tra semantica e sintassi*. Pisa: ETS edizioni.
- Jezek, E. (2018). Partecipanti impliciti nella struttura argomentale dei verbi. In S. Dallabrida, e P. Cordin (eds.), *La Grammatica delle Valenze*, Firenze: Franco Cesati, pp. 55-71.
- Jezek, E., Magnini, B., Feltracco, A., Bianchini, A. e Popescu, O. (2014). T-PAS; A resource of Typed Predicate Argument Structures for linguistic analysis and semantic processing. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC '14)*, Reykjavik, Iceland: European Language Resources Association (ELRA), pp. 890-895.
- Jezek, E. e Sprugnoli, R. (2023). *Linguistica computazionale. Introduzione all'analisi automatica dei testi*. Bologna: Il Mulino.
- Podkorytov, M., Biś, D. e Liu, X. (2021). How Can the [MASK] Know? The Sources and Limitations of Knowledge in BERT. In *Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN)*, 18-22, Shenzhen, China, pp. 1-8.

- Landis, J.R. e Koch, G.G. (1977). The measurement of observer agreement for categorical data, in *Biometrics*, 33/1, pp. 159-174.
- Lo Duca, M.G. (2000). Proprietà valenziali e criteri di descrizione lessicografica: Un caso di alternanza argomentale. In *Studi Italiani di Linguistica Teorica e Applicata*, 29/2, pp. 219-242.
- Lombardi, A. e Lenci, A. (2021). Agentività e telicità in GilBERTo: implicazioni cognitive. In *Proceedings of the Eighth Italian Conference on Computational Linguistics CLiC-it 2021*.
- Muffo, M. e Bertino, E. (2020). BERTino: an Italian DistilBERT model. In *Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-it 2020*.
- Nuccorini, S. (2001). *Il cloze test in inglese. Ricerca, metodologia, didattica*. Carocci editore.
- Parisi, L., Francia, S. e Magnani, P. (2020). *Umberto: An Italian language model trained with whole word masking*.
- Polignano, M., Basile, P., De Gemmis, M., Semeraro, G. e Basile, V. (2019). AIBERTo: Italian BERT Language Understanding Model for NLP Challenging Tasks Based on Tweets. In *Proceedings of the Sixth Italian Conference on Computational Linguistics CLiC-it 2019*.
- Puccinelli, D., Demartini, S. e Ferrari, P. L. (2021). Tackling Italian University Assessment Tests with Transformer-Based Language Models. In *Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-it 2018*.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.
- Pustejovsky J. e Jezek, E. (2008). Semantic Coercion in Language: Beyond Distributional Analysis. In *Italian Journal of Linguistics*, 20/2.
- Ravasio, G. e Di Perna, L. (2020). *GilBERTo: An Italian pretrained language model based on RoBERTa*.
- Rogers A., Kovaleva, O. e Rumshisky, A. (2020). A Primer in BERTology: What We Know About How BERT Works. In *Transactions of the Association for Computational Linguistics*, 8, pp. 842-866.
- Roth, M. e Frank, A. (2013). Automatically Identifying Implicit Arguments to Improve Argument Linking and Coherence Modeling. In *Proceedings of the Second Joint Conference on Lexical and Computational Semantics (*SEM)*, 1, Atlanta, Georgia, USA: Association for Computational Linguistics, pp. 306-316.
- Ruppenhofer, J., Sporleder, C., Morante, R., Baker, C. e Palmer, M. (2009). SemEval-2010 Task 10: Linking Events and Their Participants in Discourse. In *Proceedings of the*

- Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009)*, Boulder, Colorado: Association for Computational Linguistics, pp. 106-111.
- Ruppenhofer, J. e Michaelis, L. (2010). A constructional account of genre-based argument omissions. In *Constructions and Frames*, 2/2, pp. 158-184
- Shannon, C.E. (1948). A Mathematical Theory of Communication. In *Bell System Technical Journal*, 27, pp. 379-423.
- Squartini, M. (2013). *Il Verbo*. Carocci editore.
- Tamburini, F. (2020). How "BERTology" Changed the State-of-the-Art also for Italian NLP. In *Proceedings of the Seventh Italian Conference on Computational Linguistics CliC-it 2020*.
- Tonelli, S. e Del Monte, R. (2011). Desperately Seeking Implicit Arguments in Text. In *Proceedings of the ACL 2011 Workshop on Relational Models of Semantics*, Portland, Oregon, USA: Association for Computational Linguistics, pp. 54-62.
- Tu, J., Rim, K., Holderness, E. e Pustejovsky, J. (2022). Dense Paraphrasing for Textual Enrichment. In *Proceedings of the 15th International Conference on Computational Semantics*, Nancy, France: Association for Computational Linguistics, pp. 39-49.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. e Polosukhin, I. (2017). Attention Is All You Need. In *Advances in Neural Information Processing Systems*, pp. 5998-6008.
- Vendler, Z. (1957). Verbs and Times. In *The Philosophical Review*, 66/2, pp. 143-160.
- Ye, B., Tu, J., Jezek, E. e Pustejovsky, J. (2022). Interpreting Logical Metonymy through Dense Paraphrasing. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44.
- Zhang, Z., Kong, X., Liu, Z., Ma, X. e Hovy, E. (2020). A Two-Step Approach for Implicit Event Argument Detection. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, pp. 7479-7485.

Sitografia

BERT-it

<https://huggingface.co/dbmdz/bert-base-italian-cased>

BERT-multi

<https://huggingface.co/google-bert/bert-base-multilingual-cased>

BERTino

<https://github.com/indigo-ai/BERTino>

ELECTRA-it

<https://huggingface.co/dbmdz/electra-base-italian-xxl-cased-generator>

FastText

<https://fasttext.cc/>

FrameNet

<https://framenet.icsi.berkeley.edu/>

GitHub

Pagina principale: <https://github.com>

Repository del progetto: https://github.com/agnesedaff/Implicit_obj_completion

Google

Google Colab: <https://colab.research.google.com/>

Google Drive: https://www.google.com/intl/it_it/drive/

Hugging Face

Pagina principale: <https://huggingface.co>

Introduzione a BERT: https://huggingface.co/docs/transformers/model_doc/bert

Introduzione a ELECTRA: https://huggingface.co/docs/transformers/model_doc/electra

Introduzione al NLP: <https://huggingface.co/learn/nlp-course/chapter0/1?fw=pt>

Funzione *pipeline*: https://huggingface.co/transformers/v3.0.2/main_classes/pipelines.html

Libreria *Transformers*: <https://huggingface.co/docs/transformers/index>

I-CAB = Italian Content Annotator Bank

https://ontotext.fbk.eu/Publications/magnini_silfi-pb06.pdf

Italian Wikipedia

https://it.wikipedia.org/wiki/Pagina_principale

ItWac reduced = Italian corpus from the .it domain

<https://www.sketchengine.eu/itwac-italian-corpus/>

Matplotlib

Modulo *pyplot*: https://matplotlib.org/3.5.3/api/_as_gen/matplotlib.pyplot.html

NomBank

<https://catalog.ldc.upenn.edu/LDC2008T23>

NumPy

Funzione *mean*: <https://numpy.org/doc/stable/reference/generated/numpy.mean.html>

NVdB = Nuovo Vocabolario di Base

<https://dizionario.internazionale.it>

OPUS corpus

<https://opus.nlpl.eu/>

OSCAR corpus

<https://oscar-project.org/>

Paisà corpus

<https://www.cnr.it/it/banche-dati-istituti/banca-dati/1024/corpus-paisa>

Pandas

<https://pandas.pydata.org/>

PropBank

<https://propbank.github.io>

Python

<https://docs.python.org/3/>

Scikit-learn

Modulo *sklearn.metrics*:

<https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics>

Funzione *cohen_kappa_score*:

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.cohen_kappa_score.html

Funzione *precision_recall_fscore_support*:

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html

SciPy

Modulo *stats*: <https://docs.scipy.org/doc/scipy/reference/stats.html>

Funzione *pearsonr*:

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html>

Seaborn

<https://seaborn.pydata.org/>

SemEval 2010

Task 10: https://www.coli.uni-saarland.de/projects/semEval2010_FG/

Sketch Engine

<https://www.sketchengine.eu/>

SpaCy

Pagina principale: <https://spacy.io>

Pipeline italiana *it_core_news_lg*: <https://spacy.io/models/it>

Statsmodels

Funzione *pairwise_tukeyhsd*:

https://www.statsmodels.org/stable/generated/statsmodels.stats.multicomp.pairwise_tukeyhsd.html

T-PAS = Typed Predicate-Argument Structures

<https://tpas.unipv.it>

<https://tpas.sketchengine.eu>

UD Italian ParTUT

https://universaldependencies.org/treebanks/it_partut/index.html

UD Italian-ISDT

https://universaldependencies.org/treebanks/it_isdt/index.html

UmBERTo

<https://huggingface.co/Musixmatch/umberto-commoncrawl-cased-v1>

WIKINER

<https://metatext.io/datasets/wikiner>

Ringraziamenti

Vorrei ringraziare la mia relatrice, la professoressa Jezek, per avermi incoraggiata a coltivare questo progetto, introducendomi anche alla comunità scientifica in occasione di CliC-it 2023 e proponendomi un tirocinio presso la Fondazione Bruno Kessler.

Ringrazio la mia correlatrice, la professoressa Combei, per la sua disponibilità e per il tempo dedicato alla lettura della tesi.

Un ringraziamento speciale va a Paolo Boccalini, che si è gentilmente proposto per annotare il corpus come secondo annotatore.

Ringrazio infine la mia famiglia per avermi permesso di seguire le mie passioni e tutte le persone che mi sono state vicino in questo periodo importante della mia vita.

Abstract

La tesi descrive un esperimento sul completamento automatico dell'Oggetto Implicito in italiano. Il task è strutturato come un *fill-mask* o *cloze-task* da applicare a cinque modelli BERT italiani, sfruttando appieno le loro capacità bidirezionali. In primis, a partire da un'Ontologia scelta di 30 verbi (37 pattern semantici della risorsa T-PAS), viene creato un corpus di 1.200 frasi. Il corpus è diviso in due dataset, chiamati EXPLICIT e IMPLICIT. Il secondo dataset, contenente Oggetti Impliciti, viene annotato manualmente da due esperti con un sostantivo *Gold Standard* (GS) e il tipo di omissione, intesa come una strategia di *Defaulting* che può avvenire a livello lessicale o pragmatico (Jezek, 2018). Dall'annotazione manuale emerge una correlazione significativa tra il tipo di *Defaulting* e il range di possibili completamenti per ogni verbo. Successivamente, tale risorsa viene utilizzata per applicare l'esperimento e i risultati sono valutati calcolando la *cosine similarity* tra l'output dei modelli e il completamento manuale GS. Si dimostra che il modello *bert-base-italian-xxl-cased* funziona meglio dei modelli più leggeri nel task, grazie alla capacità di indovinare le collocazioni più frequenti nei contesti di *Defaulting* lessicale. Viene confermato quanto osservato in studi precedenti, ovvero che i modelli BERT tendono a privilegiare la frequenza degli *n-grams*, con alcune difficoltà a completare l'Oggetto laddove è necessaria una comprensione più profonda delle relazioni semantiche (es. output = "Il postino suona [il pianoforte] sempre due volte"). Si osserva inoltre la tendenza dei modelli a restituire parole in relazione metonimica con il GS, replicando il meccanismo di *semantic coercion* (Pustejovsky e Jezek, 2008), e la scarsa sensibilità ai limiti linguistici nell'esplicitazione degli *Shadow Arguments*.

Abstract (English)

This thesis describes an experiment on automatic Implicit Object completion in Italian. The task is structured as a *fill-mask* or *cloze-task* to be applied to five Italian BERT models, fully exploiting their bidirectional capabilities. Firstly, starting from a selected Ontology of 30 verbs (37 semantic patterns from the T-PAS resource), a corpus of 1.200 sentences is created. The corpus is divided into two datasets, called EXPLICIT and IMPLICIT. The second dataset, containing Implicit Objects, is manually annotated by two experts with both a *Gold Standard* (GS) Noun and the type of omission occurring, understood as a *Defaulting* strategy that can apply either lexically or pragmatically (Jezek, 2018). The manual annotation shows a significant correlation between the type of *Defaulting* and the range of possible completions for each verb. Subsequently, the experiment is applied and the results are evaluated by calculating the *cosine similarity* between the model's output and the manual GS completion. It is demonstrated that the model *bert-base-italian-xxl-cased* performs better than lighter models in the task, thanks to its ability to guess the most frequent collocations in *Lexical Defaulting* contexts. It is confirmed what has been observed in previous studies, namely that BERT models tend to favor the frequency of *n-grams*, with some difficulty in completing the Object when a deeper understanding of semantic relationships is required (e.g., output = "Il postino suona [il pianoforte] sempre due volte"). Furthermore, it is observed that the models tend to return words in metonymic relation to the GS, replicating the mechanism of *semantic coercion* (Pustejovsky and Jezek, 2008), and possess limited sensitivity to linguistic boundaries in the explication of *Shadow Arguments*.

APPENDICE A

Linee guida per l'annotazione dell'IMPLICIT dataset

Introduzione

In questo dataset ci sono 600 frasi con Oggetto Implicito estratte dal corpus *ItWac reduced* (Baroni e Kilgarriff, 2006). I verbi scelti sono 30 e per ognuno di essi ci sono 20 frasi.

L'annotazione prevede due passaggi:

- 1) Annotazione del tipo di omissione in atto.
- 2) Annotazione di un Complemento Oggetto Implicito.

Come annotare il tipo di omissione

- 1) Questa annotazione serve per capire di che tipo è l'omissione del Complemento Oggetto. Definiamo *Lexical Defaulting* l'omissione in virtù del significato intrinseco del verbo e *Pragmatic Defaulting* l'omissione in virtù del contesto (Jezek, 2018).

«[Nel *Lexical Defaulting*, ndr] l'oggetto inespresso è genericamente inteso come una classe di entità, che abbiamo chiamato precedentemente tipo semantico (cibo (mangiare), documenti (leggere, scrivere), veicoli (parcheggiare), selezionato dal predicato. [...]. Se una specifica istanza della classe deve essere ricostruita (la pizza, il biglietto di auguri, la nuova BMW, ecc.), il meccanismo attivo non sarà il defaulting lessicale, ma quello pragmatico.»

(Jezek, 2018: 62)

- 2) Come individuare il tipo di *Defaulting*:
 - a) Il Complemento Oggetto è menzionato altrove nella frase?
Sì → *Pragmatic Defaulting*; **No/non so** → avanti
 - b) Il Complemento Oggetto si può ricavare da qualcosa di dato nel contesto, c'è qualche indizio contestuale riferito ad una specifica istanza?
Sì → *Pragmatic Defaulting*; **No/non so** → avanti
 - c) Il Complemento Oggetto si può ricavare dal contesto extra-frasale o deittico, c'è qualche indizio contestuale riferito ad una specifica istanza?
Sì → *Pragmatic Defaulting*; **No/non so** → avanti

- d) Il Complemento Oggetto è un'entità generica (una classe di istanze) contenuta nella semantica del verbo? (es. *ascoltare [un suono]*, *bere [un liquido]*, *guidare [un veicolo]*...)

Sì → *Lexical Defaulting*; **No/non so** → *unknown*

Come annotare il Complemento Oggetto Implicito

- 1) Obiettivo: annotare la parola migliore per completare lo spazio destinato al Complemento Oggetto, indicato con [MASK].
- 2) L'annotazione deve essere un nome singolare o plurale. Può essere preceduto da un articolo che non andrà annotato (es. *il vino* → *vino*).
- 3) Come individuare la parola da annotare:
 - a) Il Complemento Oggetto è menzionato altrove nella frase?
Sì → inserirlo (ignorare eventuali ripetizioni); **No** → avanti
 - b) Il Complemento Oggetto è un'istanza che si può ricavare dal contesto o dal significato del verbo?
Sì → inserirlo; **No** → avanti
 - c) Guardare la classe semantica dell'Oggetto indicata nel pattern (es. SOUND = suono, FOOD = cibo). Questa classe di entità si può inserire come Oggetto?
Sì → Inserirla; **No** → avanti
 - d) Se non si trova un completamento, annotare *unknown*

Esempi:

Frase	Tipo di Defaulting	Complemento Oggetto Implicito
Dopo essersi versata il vino la donna beve [MASK] con piacere e ringraziò.	<i>Pragmatic Defaulting</i>	vino
Marta aveva in mano un bicchiere e beveva [MASK] mentre il marito la guardava con amore.	<i>Lexical Defaulting</i>	bevanda