



GPT-3.5: la linguistica nelle mani dell'AI

Esperimenti di analisi sintattica su Guglielmo IX e Raimbaut d'Aurenga

Mariagrazia Staffieri*

*Sapienza Università di Roma
mariagrazia.staffieri@uniroma1.it

1 Premessa

Uno studio sintattico ad ampio spettro, condotto su una mole considerevole di testi – siano essi in poesia o in prosa – oppure su un unico testo di lunghezza notevole, o ancora su un intero genere letterario, deve comprensibilmente prevedere più fasi di analisi: nello specifico, sarà necessario *in primis* un esame generico del macrotesto e, successivamente, la verifica sul piano microtestuale dei fenomeni sintattici individuati. Da questo punto di vista, se il materiale sottoposto all'analisi risulta difficilmente gestibile in virtù della sua cospicuità, è chiaro che un'eventuale automazione dello spoglio sintattico testuale, anche parziale, consentirebbe una non indifferente riduzione del carico di lavoro, giungendo, presumibilmente, a risultati soddisfacenti e migliori rispetto a quelli che si potrebbero ottenere svolgendo l'analisi manualmente, in tempistiche più brevi (a patto che, logicamente, al software scelto per e preposto a questa procedura siano state fornite *a priori* istruzioni precise sulle modalità mediante cui operare).

In quest'ottica, i metodi computazionali presenti nel campo delle *Digital Humanities* (d'ora in avanti *DH*) per l'analisi del testo – in continuo sviluppo e potenziamento – hanno fornito e continuano a fornire un supporto notevole allo studio, anche linguistico, dei testi letterari medievali sia in poesia sia in prosa. Si vedano, a titolo esemplificativo, gli strumenti impiegati in ambito stilometrico che, mediante indagini statistiche sugli usi linguistici, tentano «di “misurare” lo stile degli autori, discernendo così le loro “impronte autoriali” latenti»¹.

¹ Cioffi 2023, p. 108. Lo studioso offre una rassegna dei metodi computazionali maggiormente impiegati nell'ambito delle *Digital Humanities*, accompagnata da un'efficace introduzione esplicativa (*ivi*, pp. 91-113). Per approfondimenti, si veda anche Canettieri 2013.

Fra questi, basti ricordare la “distanza Delta” e l’“analisi Zeta”², utili all’esplorazione di ampi corpora testuali.

Se si vuole proseguire su questa falsariga, non si può non riflettere su eventuali futuri sviluppi dei grandi modelli di linguaggio generativo (*LLM*), i quali, benché approdati da poco nel campo delle *DH*, sembrano già mostrare le loro potenzialità. Quale ruolo può svolgere, ad esempio, il recentissimo modello linguistico *GPT*, sviluppato da *OpenAI*? E quali possono essere i futuri sviluppi della linguistica romanza in tal senso?

2 Introduzione

Alla domanda “quale categoria sintattica può essere posta sotto la lente analitica di *GPT*?” non vi sono una risposta corretta e una errata: l’intelligenza artificiale può prendere in esame qualsiasi fenomeno sintattico; tuttavia, non è garantito che la sua analisi dia sempre risultati attendibili e coerenti. Ad ogni modo, quel quesito è stato la base per l’avvio della ricerca qui esposta: lo scopo ultimo non è consistito solamente nel trovare la categoria sintattica che potesse essere studiata con più semplicità e immediatezza da *GPT-3.5*, bensì anche nel rintracciare un fenomeno che fosse difficilmente o non indagabile da altri modelli computazionali di umanistica digitale ma che, sottoposto all’esame sintattico dell’*AI*, potesse invece fornire risultati incoraggianti. La rosa di tratti sintattici eleggibili, accuratamente e primariamente delineata, comprende le macrocategorie di: sintassi del periodo (coordinazione, asindetica e polisindetica, e subordinazione); sintassi verbale (costruzioni perifrastiche verbali, in particolar modo); struttura frasale e ordine sintattico (eventuali alterazioni e inversioni delle principali unità proposizionali di soggetto, verbo e oggetto). Prima di procedere con la spiegazione relativa alla scelta della componente sintattica adeguata, è però necessaria una premessa fondamentale.

Eventuali tentativi di automazione dell’analisi sintattica devono necessariamente rivolgere l’attenzione all’esame dei dati sul piano quantitativo-statistico (statistica inferenziale), mediante metodi di *Text Mining* (o *Text Analytics*) e di *Machine Learning*. Si prendano ad esempio, nell’ambito della sintassi del periodo, i legami di due o più proposizioni mediante polisindeto. Ebbene, se si vogliono rintracciare le congiunzioni para e ipotattiche in un testo, sia esso in poesia o in prosa, per ricavarne una “rappresentazione vettoriale”, è necessario stilare, in una fase preliminare, una lista ordinata (“dizionario”) di elementi da ricercare. Si tratta di un’operazione indubbiamente laboriosa, che richiede non poco tempo e che porta con sé alcuni limiti non indifferenti. Basti pensare alle congiunzioni per le quali sussistono grafie differenti (per la lirica occitana, si veda la più semplice *e*, sovente alternata a *ez* oppure a *et*): è indispensabile che tutte queste varianti grafiche siano inserite nella lista, per ottenere dei risultati attendibili. La pratica di selezione, ad ogni modo, malgrado alcuni vincoli e con un po’ di attenzione, può essere portata a termine, e l’indagine relativa a paratassi e ipotassi può proseguire. Il risultato finale sarà un *DTM* (*Document-Term Matrix*), vale a dire un documento che racchiude una rappresentazione numerica (con frequenze assolute) delle ricorrenze dei termini del dizionario nei testi posti in analisi.

² Cfr. Burrows 2002.

Tuttavia, non tutte le categorie sintattiche sono passibili di automazione o, almeno, non attraverso le più comuni metodologie di *Data Analytics*, proprio perché non è sempre possibile elaborare un “vocabolario” fisso come per le congiunzioni para e ipotattiche. È il caso della coordinazione asindetica oppure delle sovversioni del regolare ordine sintattico soggetto-verbo-oggetto (con conseguenti anastrofi, inversioni e alterazioni dell’*ordo verborum*), o ancora delle strutture perifrastiche. Ed è qui che interviene *GPT-3.5*: nei paragrafi seguenti vedremo se e fino a che punto sia possibile, per queste componenti sintattiche, trovare un’applicazione automatica mediante un modello di linguaggio generativo; se e come l’*AI* può colmare quelle lacune lasciate dagli strumenti computazionali più diffusi nel campo delle *DH*; e se e come il chatbot è in grado di condurre un’analisi sintattica su un *corpus* lirico provenzale, con eventuali limitazioni, potenzialità e possibili sviluppi futuri.

3 Considerazioni preliminari

3.1 La scelta della categoria sintattica

Se escludiamo coordinazione polisindetica e subordinazione, che sembrano poter intraprendere senza troppe difficoltà un embrionale cammino verso la strada dell’automazione, tre sono le categorie sintattiche che potrebbero giungere all’attenzione scrutatrice di *GPT*: coordinazione per asindeto, sovversioni dell’ordine sintattico, costruzioni perifrastiche verbali. Per quanto riguarda la prima categoria, recenti analisi da me effettuate su un campione di *corpora* trobadorici provenzali hanno provato che si tratta in generale di un fenomeno poco diffuso nella lirica occitana (per Raimbaut d’Aurenga, ad esempio, si annoverano solamente 80 casi su un totale di 2258 versi)³. Data la sua ridotta frequenza, quindi, la categoria in questione è stata in questa sede esclusa. Le due restanti, invece, funzionano entrambe sotto la lente di *GPT-3.5*; tuttavia, il riconoscimento delle variazioni dell’ordine sintattico non è sempre istantaneo, e spesso il chatbot presenta risultati poco efficaci e/o completamente errati. Di seguito verrà presentato un breve test a conferma di quanto appena sostenuto.

Quando forniamo al modello linguistico di *AI* una proposizione – preferibilmente una principale –, questi sembra in grado di riconoscere le alterazioni dell’*ordo verborum* in lingua provenzale, anche in lirica. Vediamo un esempio. Abbiamo chiesto a *GPT* se nella frase «Amigu’ai ieu» – che apre il v. 25 («Amigu’ai ieu, no sai qui s’es») della celeberrima *Farai un vers de dreyt nien* di Guglielmo IX (*BdT* 183, 7)⁴ – vi fossero sovversioni rispetto al comune ordine sintattico soggetto-verbo-oggetto. La risposta del chatbot è stata la seguente:

³ In generale, la coordinazione per asindeto sembra essere un fenomeno poco diffuso in ambito provenzale, anche in prosa. Da un ulteriore, ma sempre recentissimo, spoglio sintattico da me condotto sulle *vidas* trobadoriche (edizione di Boutière–Schutz–Cluzel 1964), è risultato che questa categoria rappresenta solamente il 2% del totale, se si considera la divisione dei singoli testi in proposizioni (59 casi su 2996 proposizioni complessive), e lo 0,3% del totale delle parole di cui si compongono i testi (59 casi su 20.199 parole complessive).

⁴ Edizione di riferimento: Pasero 1973.

- (1a) *Nella frase «Amigu'ai ieu», l'ordine sintattico sembra essere leggermente diverso da quello che potrebbe essere considerato l'ordine standard delle lingue romanze moderne come l'italiano. Questa frase sembra essere in antico-occitano o in un dialetto simile.*

L'ordine sintattico tradizionale dell'antico-occitano potrebbe variare rispetto all'ordine Soggetto-Verbo-Oggetto delle lingue romanze moderne.

Nella frase data: «Amigu'ai ieu» (Amico io ho), il soggetto ("ieu" che significa "io") segue il verbo e l'oggetto ("Amigu'ai" che significa "ho amico"). L'ordine è spesso descritto come OVS (Oggetto-Verbo-Soggetto), dove il verbo precede il soggetto e succede l'oggetto.

Feedback positivo. Proseguendo, ci rendiamo però subito conto che più complessa è la costruzione frasale che affidiamo a GPT, e meno attendibili e precisi sono i risultati che ci vengono forniti: è il caso del v. 7 di BdT 183, 3, sempre di Guglielmo IX («Dos cavals ai a ma sselha, ben e gen»). In questo e in molteplici ulteriori casi più elaborati, l'intelligenza artificiale, ammettendo di non conoscere appieno le norme grammaticali dell'antico-occitano in merito a questa macrocategoria sintattica, risponde così:

- (2a) *Nella frase «Dos cavals ai a ma sselha, ben e gen», sembra che l'ordine sintattico sia rispettato, almeno in termini generali. Tuttavia, è importante notare che questa frase potrebbe appartenere a una lingua o a un dialetto che ha un ordine sintattico diverso rispetto alle lingue romanze moderne come l'italiano. Quindi, se questa frase non è in italiano moderno, potrebbe essere scritta secondo le regole dell'ordine sintattico della lingua a cui appartiene.*

Dal contesto sembra che sia in antico-occitano o in un dialetto simile. Senza conoscere la grammatica specifica di questa lingua o dialetto, potrebbe essere difficile valutare con precisione se ci sono sovversioni dell'ordine sintattico.

In generale, il modello di OpenAI risulta avere più difficoltà quando nella frase da analizzare, oltre a soggetto, verbo e oggetto, sono presenti elementi ulteriori che ne pregiudicano la corretta verifica (aggettivi, avverbi, congiunzioni – quando la proposizione in questione non è una principale, bensì una coordinata o una subordinata – e così via). È questa una delle possibili motivazioni sottese all'errato esame sintattico del verso «Dos cavals ai a ma sselha, ben e gen». Se infatti proviamo a rimuovere il materiale "in eccesso" – e probabilmente percepito "di disturbo" dal chatbot –, fornendo a GPT le sole unità proposizionali principali («Dos cavals ai»), noteremo che l'analisi migliora, seppure non sia ancora da ritenersi pienamente accettabile come quella di (1a):

- (3a) *La frase «Dos cavals ai» sembra essere in una forma che non corrisponde all'ordine sintattico tipico dell'italiano standard, ma è più coerente con la struttura sintattica dell'antico-occitano. In italiano, l'ordine sintattico tipico sarebbe SVO. Quindi, la frase potrebbe essere tradotta in italiano come "Ho due cavalli" o "Ci sono due cavalli", a seconda del contesto.*

Se i versi devono essere, nella maggior parte dei casi, "frammentati" per poter essere analizzati correttamente dall'AI, la situazione migliora di gran lunga nel caso di proposizioni di un testo non poetico. Il modello linguistico, infatti, sembra trovarsi maggiormente a suo agio in contesti prosastici, e quindi quando, benché vi siano sovversioni delle principali unità proposizionali, non vi sono comunque eccessive alterazioni, e le proposizioni risultano nel complesso di più

facile comprensione sul piano semantico. Prendiamo ad esempio la principale «Mout duret lor amors longa sason» tratta dalla *vida* di Bernart de Ventadorn⁵: notiamo che *GPT*, nonostante qualche errore di traduzione, la analizza correttamente, com'era avvenuto per (1a).

(4a) «*Mout duret lor amors*» (Molto dura il loro amore): *verbo e soggetto sono invertiti rispetto all'ordine SVO*.

Nella lirica, le variazioni legate all'*ordo verborum* sono da considerarsi un primo punto di contatto fra andamento sintattico e andamento versificatorio, poiché spesso la scelta di invertire alcune parti fondamentali del discorso poetico è relazionabile al rispetto di determinati criteri metrici (fra cui ortometria versale e schema rimico). Di conseguenza, in quest'ambito troviamo comprensibilmente casi più articolati di sovversioni dell'ordine sintattico, in confronto a quanto osservabile, ad esempio, in una prosa semplice come quella delle *vidas*.

L'ultima categoria sintattica è rappresentata dalle costruzioni perifrastiche verbali. Si tratta di un campo, quello delle perifrasi, che si presenta allo stesso modo e senza eccessivi mutamenti nel passaggio dalla poesia alla prosa, e che non sembra essere particolarmente condizionato dalle scelte versificatorie dell'autore. Ciò nonostante, questo fenomeno si presta a sperimentazioni di vario genere circa le forme ausiliari e le combinazioni ausiliare-infinito, le quali danno luogo a costruzioni sempre differenti e non esauribili interamente in una lista (com'è possibile fare invece con le congiunzioni para e ipotattiche). Benché il numero di accostamenti ausiliare-infinito sia incalcolabile – considerato anche che il verbo principale non è mai lo stesso –, è comunque possibile stilare un elenco basilare, necessario come punto di riferimento e di partenza per la ricerca qui esposta.

I modelli perifrastici che il chatbot dovrà individuare sono i seguenti:

- *ausar* + infinito: «ni no **m'aus traire** adenan» (*BdT* 183,1: v. 10);
- *anar* + gerundio: «**van domnas duptan**» (*BdT* 389,4: v. 19);
- *poder, voler, dever* + infinito: «**puesc ne traïr** lo vers» (*BdT* 183, 2: v. 6);
- *saper, fazer* + infinito: «**farai tornar** tal nevei» (*BdT* 389, 9: v. 11);
- *laisser* + infinito: «que·s **laissez morir** de sei» (*BdT* 183, 4: v. 21);
- *esser* + participio presente: «e d'aquel eus **serai jausentz**» (*BdT* 389, 2: v. 12).

Lo spoglio in questione fa riferimento alle combinazioni più comunemente riscontrabili nella lirica provenzale, e chiaramente già individuate nei due *corpora*-campione. Come già notato da vari studiosi, alcune di queste perifrasi (*anar* + gerundio ed *esser* + participio presente, ad

⁵ Versione tradata da *N²*, per cui cfr. Boutière–Schutz–Cluzel 1964, pp. 26-28.

esempio) si riscontrano senza troppa fatica anche nella lingua poetica italiana pre-stilnovista, in continuità con le scelte sintattiche provenzali in materia lirica.⁶

3.2 Eventuali fattori di agevolazione e di rischio

Individuare e registrare tutte le forme perifrastiche presenti in una frase oppure in un testo rappresenta una sfida, insomma, per *GPT-3.5*. Tuttavia, quest'operazione potrebbe rivelarsi potenzialmente più semplice rispetto alla segnalazione di sovversioni dell'ordine sintattico. L'eventuale facilitazione nella ricerca delle perifrasi rispetto all'altra categoria menzionata è legata a due fattori:

- a) la ricerca si ripete allo stesso modo in tutti i casi. Qualunque sia la proposizione sottopostagli, il chatbot dovrà rintracciare sempre le stesse componenti: un verbo ausiliare e un verbo all'infinito. Ciò al contrario, ad esempio, delle alterazioni delle unità proposizionali, ove le categorie grammaticali coinvolte sono molteplici.
- b) La ricerca primaria condotta da *GPT* sarà sempre il verbo all'infinito. Come si vedrà dai risultati delle analisi, è probabile che a favorire la comprensione e conseguentemente il lavoro svolto dal modello linguistico sia la prioritaria individuazione del verbo all'infinito, che dà l'input iniziale per riconoscere una perifrasi. Questo si nota, come vedremo, tanto più nelle costruzioni in cui l'ausiliare è sottinteso, ma che vengono ugualmente riconosciute come strutture perifrastiche (si veda l'esempio «*saup dir e far*», laddove *far* rappresenta una costruzione a sé stante, seppur sostenuta dal medesimo ausiliare *saup*).

Accanto ai motivi che possono favorire la comprensione, e quindi il corretto svolgimento del lavoro, di *GPT-3.5* relativamente alle costruzioni perifrastiche verbali, sono stati previsti anche due possibili limiti che il chatbot avrebbe potuto riscontrare relativamente alla registrazione di queste strutture, compromettendone parzialmente o interamente l'analisi. Queste due componenti di possibile rischio sono:

- a) inversione dei due elementi costituenti la struttura perifrastica, quindi infinito + ausiliare anziché ausiliare + infinito, riscontrabile soprattutto nella più complessa sintassi di Raimbaut d'Aurenga. Si vedano, a scopo illustrativo, i casi di *BdT* 389, 27, v. 29: «s' **amar volc** de guiza genta» e di *BdT* 389, 38a, v. 11: «**Amar la dei**».
- b) Interposizione di elementi terzi fra verbo ausiliare e verbo principale. Se ne segnalano occorrenze anche in Guglielmo IX, come in *BdT* 183, 11, v. 5: «ben **deu** chascus lo joi **jauzir**» oppure in *BdT* 183, 8: v. 4: «ben **dei**, si puesc, al meils **anar**». Generalmente questi fattori di interferenza sono l'oggetto e/o il soggetto del verbo, come avviene nel primo esempio qui riportato, oppure avverbi e locuzioni, o ancora ulteriori proposizioni, come si legge nel secondo esempio.

⁶ Imprescindibile il lavoro di Corti 2005, pp. 67-155. Per approfondimenti, vd. anche Brambilla 1964; Squartini 2011; Colella 2020a; 2020b.

Tenendo ben presenti questi presupposti, è ora possibile osservare con una lente d'ingrandimento come lavora *GPT-3.5* in campo linguistico.

4 La sperimentazione

4.1 Premessa metodologica

Al fine di assicurare il corretto svolgimento della sperimentazione, si è deciso di optare, in uno stadio preparatorio, per una divisione in più passaggi e per uno sviluppo del procedimento *step by step*. L'impiego di questa metodologia, infatti, ha permesso di tastare gradualmente e accuratamente il lavoro di *GPT* e di verificare, di volta in volta, i risultati raggiunti dal chatbot. In una fase preliminare, l'obiettivo è stato capire se l'intelligenza artificiale fosse in grado di rilevare una costruzione perifrastica in un testo in versi in lingua antico-occitana. Proprio per questo, è stata elaborata una griglia di valutazione comprendente i seguenti criteri:

- Riconoscimento della costruzione perifrastica verbale complessiva (verbo ausiliare + verbo principale).
- Analisi sintattico-descrittiva delle componenti della perifrasi individuata, a un livello superficiale (identificazione di – e distinzione fra – rispettivamente, ausiliare e infinito) e a un livello più specifico (verifica di modo, tempo e persona verbali dell'ausiliare).
- Traduzione in italiano moderno e interpretazione del significato della perifrasi.

Prendendo le mosse da questo sintetico schema, e prima di procedere con l'effettiva analisi, si è resa necessaria una suddivisione tipologica delle costruzioni perifrastiche secondo quanto segue:

- a) Costruzioni perifrastiche regolari, nelle quali cioè il verbo principale segue coerentemente il verbo ausiliare e non vi sono ulteriori elementi da considerare all'interno della struttura (fatta eccezione per eventuali pronomi clitici).
- b) Costruzioni perifrastiche irregolari o variabili, le quali possono presentare alterazioni nella loro struttura interna, come vedremo in §4.4.

Queste sottocategorie verranno esaminate mediante tre differenti modalità:

- I. Analisi della perifrasi sul singolo verso: verrà sottoposto all'attenzione di *GPT-3.5* solamente il verso contenente la perifrasi.
- II. Analisi della perifrasi su due versi: verranno sottoposti all'attenzione del chatbot il verso contenente la perifrasi + il verso precedente o il successivo (la scelta dell'uno o dell'altro terrà conto dell'inizio e della fine della proposizione, al fine di assicurarne la corretta e completa resa).

III. Analisi della perifrasi in un testo: verrà sottoposto all'attenzione del modello linguistico il verso contenente la perifrasi contestualizzato nella propria strofa.

Si inizierà con l'analisi delle costruzioni perifrastiche regolari e irregolari sul singolo verso, per poi convogliare entrambe le tipologie nell'esame su due versi e su intere stanze. Prima di procedere con l'esperimento però, verrà esposto un breve test preliminare svolto sul primo *corpus* lirico, avente lo scopo di "educare" il chatbot, e correggere quindi già dal principio eventuali errori verificatisi durante procedura di analisi. Il test si riferisce esclusivamente alle strutture perifrastiche regolari esaminate sul singolo verso.

4.2 Test preliminare

Il *corpus*-campione utilizzato per questo test preliminare – che sarà poi oggetto anche della sperimentazione in sé – appartiene a Guglielmo IX. Considerando esclusivamente la lirica, la scelta del primo trovatore da sottoporre a *GPT* è stata quantomai ovvia: la raccolta di versi del duca d'Aquitania, infatti, si contraddistingue per una generale semplicità sintattico-stilistica, muovendosi ampiamente su un andamento discorsivo paratattico⁷ (i rapporti di coordinazione costituiscono circa il 45,7% del totale, vale a dire 214 casi su 468 versi complessivi, contro i 143 casi di subordinazione, rappresentanti il 30,5% del totale). Sulla falsariga della sintassi del periodo, negli ambiti della sintassi verbale e della costruzione frasale, il *corpus* del trovatore annovera rispettivamente il 17,7% di costruzioni perifrastiche e solamente il 23,2% di sovversioni del regolare ordine sintattico soggetto-verbo-oggetto.

Il test preliminare è quindi consistito nel sottoporre a *GPT* i versi singoli dell'autore in questione, contenenti le costruzioni perifrastiche: si è scelto di seguire l'ordine cronologico delle perifrasi, a partire quindi dalle prime che si incontrano svolgendo l'analisi sintattica completa dei componimenti di Guglielmo IX (adottando la sequenza numerica dei testi e l'identificazione repertoriale presenti in *BdT*). Le costruzioni perifrastiche regolari in questo *corpus* sono 52 su 83 complessive (62,6%): fra queste, riporteremo qui due dei casi più significativi.

Abbiamo chiesto all'intelligenza artificiale se sapesse trovare le strutture perifrastiche verbali nel verso «ni no **m'aus traire** adenan» (*BdT* 183,1: v. 10), e la risposta del chatbot è stata la seguente:

- (1) Nella frase «ni no m'aus traire adenan», è presente una costruzione perifrastica verbale. La costruzione perifrastica è "no m'aus traire", in cui "aus" funge da ausiliare, accompagnato dal clitico "m", e "traire" è il verbo principale all'infinito. Insieme, formano la perifrasi verbale per indicare l'azione di "osare" o "avere il coraggio" di fare qualcosa: "non ho il coraggio di tirare".

Nel complesso, l'analisi condotta da *GPT* è accettabile. Riprendendo la nostra griglia di valutazione, notiamo infatti che il modello linguistico è stato in grado di riconoscere la costruzione perifrastica completa (verbo ausiliare + verbo principale); al contempo, è riuscito a

⁷ Gli indici di frequenza da qui in avanti riportati provengono da un lavoro di analisi statistica della sintassi provenzale da me svolto tanto in ambito poetico quanto in ambito prosastico (relativamente solo alle *vidas* e *razos* biografiche trobadoriche).

condurre un'analisi sintattico-descrittiva precisa delle unità della perifrasi, identificando anche modo, tempo e persona verbali dell'ausiliare. La componente in merito alla quale l'intelligenza artificiale si è rivelata in parte carente è la traduzione in italiano moderno, e quindi la corretta resa del significato. Com'è evidente, la difficoltà di base si pone a livello dell'infinito: se l'interpretazione dell'ausiliare è infatti buona, per il verbo principale sembra invece che il chatbot sia stato tratto in inganno dalla parziale omografia fra il verbo in antico-occitano "traire" e l'italiano "tirare", conseguentemente a cui vi è stata una traduzione per analogia grafica. Il plausibile errore di comprensione non ha comunque compromesso l'esame sintattico nel suo complesso, il quale ha fornito un risultato incoraggiante.

Vediamo un altro caso.

(2) «*bos metges er, si·m pot guerir*» (BdT 183,7: v. 23)

Nel verso fornito, è presente una costruzione perifrastica: "m pot guerir". Questa costruzione perifrastica è formata dal verbo ausiliare "pot" (potere) seguito dal verbo all'infinito "guerir" (guarire) e dalla particella pronominale "m" (mi). Insieme, formano la costruzione perifrastica che esprime il concetto di possibilità o capacità: "se mi posso guarire".

Anche in questo caso, GPT-3.5 ha svolto un buon lavoro, migliore per alcuni aspetti in confronto a (1) poiché, oltre a un corretto riconoscimento della costruzione perifrastica verbale completa e a un'ottima analisi sintattico-descrittiva delle componenti della perifrasi, il modello linguistico ha fornito anche una traduzione soddisfacente e maggiormente afferente al reale significato della frase, rispetto al caso precedente. Inoltre, al pari di (1), il chatbot chiarisce quale tipologia di struttura perifrastica sta esaminando: una perifrasi esprime «il concetto di possibilità o capacità» grazie al verbo servile "potere". La sola incorrettezza (che, come vedremo, non sarà un *unicum*) riguarda la persona verbale alla quale è declinato l'ausiliare, reso alla prima singolare "posso" in luogo della terza singolare "può".

4.3 Analisi delle costruzioni perifrastiche regolari sul singolo verso

Considerata la buona riuscita del test preliminare, è stato possibile proseguire l'esperimento relativo all'analisi delle costruzioni perifrastiche regolari sul singolo verso, prendendo in esame il secondo *corpus*-campione scelto per questa sperimentazione: quello di Raimbaut d'Aurenga⁸.

Anche in questo caso, la selezione relativa all'autore da sottoporre alla lente analitica di GPT trova specifiche motivazioni. Com'è noto, l'aurengate si rende fautore di un linguaggio enigmatico, dallo stile oscuro e complesso, mostrandosi come un sincero innovatore anche nel campo della sperimentazione sintattica⁹. Le analisi statistiche da me condotte su questo *corpus* non fanno altro che confermare l'ermeticità stilistica dell'autore, il quale sviluppa il discorso poetico su una linea prettamente ipotattica, con livelli di subordinazione che arrivano – com'era stato con la paratassi per Guglielmo IX – al 44,9% (1015 casi su 2258 versi totali), accompagnati

⁸ Edizione di riferimento: Pattison 1952.

⁹ Approfondimenti ulteriori circa l'arte metrico-stilistica rambaldiana in Pattison 1952, pp. 33-62; Picone 1977, pp. 28-37; Frasca 1992, pp. 19-62; Milone 1998, pp. 11-34; Di Girolamo 2021, pp. 100-119.

da una percentuale di sovversioni dell'ordine sintattico standard SVO del 36% e dal 12,3% di costruzioni perifrastiche verbali (279 casi). È bene sottolineare che, benché il numero di strutture perifrastiche si trovi a un livello piuttosto discreto in ambito poetico, ciò che interessa per l'analisi in questione è la semplicità (nel caso del duca d'Aquitania) oppure la complessità (nel caso di Raimbaut d'Aurenga) delle perifrasi, per tastare il *modus operandi* dell'intelligenza artificiale su casistiche più e meno elaborate.

In questo *corpus* poetico, le costruzioni perifrastiche regolari sono 143 sul totale di 279 (51,3%). L'ordine seguito rimane quello presente su *BdT*, tenendo comunque ben presente la linea cronologica con cui è stata effettuata l'analisi sintattica.

Al pari del caso precedente, qui esporremo i due esempi più significativi. Di seguito il primo:

(3) «*qe·l mons totz li deu servir*» (BdT 389, 7: v. 55)

Nel verso «qe·l mons totz li deu servir», è presente una costruzione perifrastica verbale. La costruzione perifrastica è “deu servir”, in cui il verbo “deu” funge da ausiliare e “servir” è il verbo principale. Insieme, formano la perifrasi verbale per indicare un'azione futura o un obbligo: “deve servire”.

Com'era avvenuto per Guglielmo IX, *GPT-3.5* non sembra aver riscontrato problemi circa il riconoscimento della perifrasi e la conseguente analisi sintattico-descrittiva. Anche qui, infatti, i problemi rilevati riguardano la traduzione e, nello specifico, la resa della persona verbale, con la consueta confusione fra prima e terza singolari, al pari di (2).

Proseguiamo con il secondo caso:

(4) «*lo ben qe·m degratz offrir*» (BdT 389, 7: v. 67)

Nel verso «lo ben qe·m degratz offrir», è presente una costruzione perifrastica verbale. La costruzione perifrastica è “degratz offrir”, in cui il verbo “degratz” funge da ausiliare e “offrir” è il verbo principale. Insieme, formano la perifrasi verbale per indicare un'azione che deve essere compiuta: “dovete offrire”.

I risultati soddisfacenti di quest'analisi pongono anche l'esempio (4) sulla falsariga dei precedenti. La difficoltà legata alla traduzione, questa volta, non riguarda la persona, bensì il tempo e il modo verbali: il verbo che funge da ausiliare è declinato al presente condizionale, ma il modello linguistico lo traduce al presente indicativo, pur preservando la persona (seconda plurale).

Il bilancio complessivo di questa prima parte della sperimentazione è estremamente positivo: escludendo, infatti, i comprensibili limiti legati alla traduzione, notiamo che il riconoscimento delle costruzioni perifrastiche e la successiva analisi sintattico-descrittiva vengono compiute senza il minimo margine di errore da *GPT-3.5*. Sia per il *corpus* di Guglielmo IX, sia per quello

di Raimbaut d'Aurenga, la percentuale di perifrasi regolari sul singolo verso correttamente esaminate raggiunge il 100%¹⁰.

4.4 Analisi delle costruzioni perifrastiche irregolari sul singolo verso

Osserviamo ora come lavora *GPT* in un contesto più complesso, ossia quando deve rintracciare quelle costruzioni perifrastiche che abbiamo voluto definire "irregolari", categoria in cui rientrano molteplici tipologie di strutture. Ne riconosciamo (e ne analizzeremo) tre principali:

- a) Le strutture perifrastiche discontinue, che presentano l'inserimento di elementi terzi fra il verbo ausiliare e il verbo principale (generalmente si tratta di "parole-contenuto", quali aggettivi e nomi, ma possono essere anche avverbi).
- b) Le strutture perifrastiche doppie, in cui vi è un unico ausiliare che regge due o più verbi all'infinito (ad esempio «dei auzir e vezzer»).
- c) Le strutture perifrastiche soggette a sovversioni dell'ordine sintattico delle due unità portanti, nelle quali cioè troviamo la costruzione ausiliare + infinito invertita (infinito + ausiliare).

Nei *corpora*-campioni presi in esame, queste perifrasi equivalgono, in termini statistici, rispettivamente al 37,4% per Guglielmo IX (31 costruzioni perifrastiche irregolari sulle 83 complessive) e al 48,7% per Raimbaut d'Aurenga (136 costruzioni perifrastiche irregolari sulle 279 complessive): in ambo i casi, quindi, esse costituiscono meno della metà delle strutture perifrastiche totali.

Gli esempi che ora esporremo, e sui quali poi discuteremo, rappresentano indubbiamente le casistiche che hanno offerto i risultati più interessanti e che spingono, in generale, a maggiori spunti di riflessione.

Il primo di essi è tratto da Raimbaut d'Aurenga: «**vuoill** un bon nou vers **commensar**» (*BdT* 389, 1: v. 4). Si tratta di una perifrasi discontinua (tipologia a), che presenta l'inserimento di elementi "di disturbo" fra il verbo ausiliare e il verbo all'infinito. In questo caso, infatti, fra *vuoill* e *commensar*, viene posto l'oggetto del verbo, formato da ben quattro componenti: articolo + aggettivo + aggettivo + sostantivo. L'interposizione dell'oggetto del verbo fra le due unità principali della costruzione perifrastica è una caratteristica tipica dell'antico-occitano, che si riscontra non raramente tanto in poesia¹¹ quanto in prosa¹² e che può subire talvolta alcune piccole variazioni (ad esempio l'aggiunta del soggetto, anziché dell'oggetto, fra i due verbi).¹³

¹⁰ Tutti i dati statistici raccolti dalla sperimentazione sono riportati in *Appendice*.

¹¹ Si veda solamente, a titolo esemplificativo, Bernart de Ventadorn: «e per so dei la bona fi grazir» (*BdT* 70,1, v. 7); «c'om no pot cor destrenher ses aucire.» (*BdT* 70,12, v. 25).

¹² Ancora una volta, il metro di paragone qui adottato in termini prosastici è costituito dalle *vidas* trobadoriche: cfr. «e fetz lo cor raustir» (Guillem de Cabestaing, *vida* tràdita da *H*); «que li faria tantas messas dire» (Guillem de la Tor, *vida* tràdita da *IK*).

¹³ Cfr. Marcabru: «Assaz pot hom vilaneiar» (*BdT* 293,15, v. 7).

Ad ogni modo, alla richiesta di individuare le costruzioni perifrastiche in quel verso, il modello linguistico risponde così:

- (5) *Nel verso «vuoill un bon nou vers commensar», è presente una costruzione perifrastica: “vuoill commensar”.*

Questa costruzione è formata dal verbo modale “vuoill” (volere), che funge da ausiliare, seguito dal verbo all’infinito “commensar” (cominciare). Insieme, formano la costruzione perifrastica che esprime la volontà di cominciare: “voglio cominciare”.

Contrariamente ad ogni aspettativa, non sembrano esserci stati problemi per GPT tanto nel riconoscimento della perifrasi, quanto nell’analisi sintattico-descrittiva che ne deriva, nonostante la presenza dei quattro elementi interposti fra l’ausiliare e l’infinito (dei quali il chatbot non tiene giustamente conto). Anche la traduzione, diversamente dai casi analizzati in precedenza, è corretta e completa.

Il secondo caso di perifrasi variabile che prendiamo in considerazione rientra nel *corpus* di Guglielmo IX e si inserisce nella seconda tipologia qui individuata (b): si tratta quindi di una costruzione perifrastica doppia, in cui abbiamo due infiniti ma un unico ausiliare.

- (6) *«c’om puesca vezer ni auzir» (BdT 183, 8: v. 6)*

Nel verso «c’om puesca vezer ni auzir», sono presenti due costruzioni perifrastiche verbali. Le costruzioni sono:

1. “puesca vezer”, in cui il verbo “puesca” funge da ausiliare e “vezer” è il verbo principale. Insieme, formano la perifrasi verbale che esprime il concetto di possibilità o capacità: “possa vedere”.

2. “puesca auzir”, in cui il verbo “puesca” funge da ausiliare e “auzir” è il verbo principale. Insieme, formano la perifrasi verbale che esprime il concetto di possibilità o capacità: “possa ascoltare”.

Come notiamo, GPT-3.5 è stato in grado di rilevare correttamente entrambe le perifrasi, malgrado la seconda non fosse immediatamente preceduta dal suo ausiliare, e ha saputo fornire un’analisi sintattico-descrittiva precisa e una traduzione esaustiva.

L’ultimo caso di costruzione perifrastica irregolare che vogliamo sottoporre all’AI rientra nella terza tipologia (c), ed è un esempio di alterazione sintattica relativa alle due componenti principali della struttura perifrastica: ausiliare e infinito. Il verso in questione è tratto da Raimbaut d’Aurenga, poiché per Guglielmo IX non si registrano casistiche di questo tipo.

- (7) *«Amar la dei se non fas fahizo» (BdT 389, 38a: v. 11)*

Nel verso «Amar la dei se non fas fahizo», è presente una costruzione perifrastica verbale: “amar la dei”. In questa costruzione, il verbo “amar” (amare) è seguito dall’ausiliare “dei” (deve), per formare una costruzione perifrastica che esprime l’obbligo o la necessità dell’azione di amare: “deve amarla”.

L'esito di quest'ultima prova si può nel complesso ritenere soddisfacente, considerato che *GPT* ha compreso l'inversione di ausiliare e infinito. La traduzione, come talvolta avviene, lascia qualche dubbio sulla corretta interpretazione della persona verbale.

In generale, operando sul singolo verso, il modello linguistico di *OpenAI* non incontra particolari difficoltà e, sul piano strettamente analitico, arriva senza eccessivi indugi a risultati coerenti e incoraggianti. Vediamo ora cosa avviene quando i versi da analizzare diventano due.

4.5 Analisi delle costruzioni perifrastiche su due o più versi

Quando si è deciso di sottoporre all'attenzione di *GPT-3.5* due o più versi da esaminare, si è ritenuto opportuno fare un passo indietro e tornare alle sperimentazioni condotte nel test preliminare, prendendo quindi in esame le medesime costruzioni perifrastiche e aggiungendo il verso precedente oppure il successivo. Questo perché, avendo già appurato che il chatbot è in grado di riconoscere quelle specifiche strutture perifrastiche, capiremo che, qualora l'*AI* commetta un errore, con tutta probabilità quell'errore sarà dovuto alla presenza di un verso aggiuntivo, e non potrà essere quindi imputabile alla comprensione della perifrasi in sé. Per le medesime ragioni, si è scelto di non cancellare la cronologia: in questo modo, dinanzi al materiale postogli in analisi, il modello linguistico potrà usufruire di una "memoria" pregressa.

Ad ogni modo, la casistica che ora esporremo risulta essere una delle più problematiche e che ha richiesto maggiore tempo per correggere l'approccio del chatbot.

- (8) *«bors metges er, si-m pot guerir,
mor non, si amau»* (BdT 183, 7: vv. 23-24)

Nel testo fornito, non sono presenti costruzioni perifrastiche. Le costruzioni perifrastiche sono forme verbali composte da un verbo ausiliare (come "essere" o "avere") seguito da un participio passato o un gerundio. Nell'esempio fornito, i verbi "guerir" e "amau" sono semplici forme verbali al presente e al passato, rispettivamente, e non sono parte di costruzioni perifrastiche.

Primo tentativo non riuscito. Un dato è chiaro sin da subito: benché il materiale da analizzare sia aumentato di poco rispetto agli esempi illustrati finora, il modello di intelligenza artificiale inizia a presentare alcuni segni di cedimento. Il primo verso di (8), infatti, era stato sottoposto a *GPT* già in una fase iniziale (2), ed è l'unico contenente la costruzione perifrastica. In seguito al riscontro negativo, comunque, abbiamo provato a rigenerare la risposta, sollecitando il chatbot a ripetere l'analisi. Di seguito i nuovi risultati ottenuti:

- (9) *Nel testo fornito, sono presenti due costruzioni perifrastiche:*

1. *"si-m pot guerir". Questa costruzione è formata dal verbo "pot" (potere) seguito dal verbo "guerir" (guarire) e dalla particella pronominale "m" (mi). Insieme, formano la costruzione perifrastica che esprime il concetto di possibilità: "se mi posso guarire".*

2. *"si amau". Questa costruzione è formata dal verbo "amau" (amare) e dalla particella pronominale "si" (se). Insieme, formano la costruzione perifrastica che esprime il concetto di condizionalità: "se amavo".*

Secondo tentativo parzialmente riuscito. *GPT* ha rilevato due costruzioni perifrastiche: la prima gli era stata già sottoposta in (2), e la analizza allo stesso modo; la seconda, invece, non è chiaramente una struttura perifrastica verbale. A questo punto, anziché generare di nuovo la risposta, abbiamo ritenuto più opportuno far notare l'errore all'intelligenza artificiale, scrivendo chiaramente che "si amau" non è una perifrasi verbale. Il chatbot effettivamente si corregge e fornisce l'esito giusto.

Malgrado i risultati non molto incoraggianti di questa prima analisi, abbiamo proseguito la sperimentazione, sottoponendo al modello linguistico di *OpenAI* un caso totalmente differente, questa volta estratto dal *corpus* di Raimbaut d'Aurenga. Ancora una volta, si tratta di un verso già analizzato, accompagnato da un secondo verso contenente anch'esso una struttura perifrastica: la procedura di riconoscimento che dovrebbe svolgere qui *GPT* è pertanto duplice, perché include entrambe le perifrasi.

- (10) «*Dompna, ieu vos dei grazir*
so q'ieu sai ben far e dir» (BdT 389, 7: vv. 64-65)

Nel testo fornito sono presenti due costruzioni perifrastiche:

1. "*ieu vos dei grazir*". Questa costruzione è formata dal pronome "*ieu*" (io) seguito dal verbo ausiliare "*dei*" (deve) e dal verbo "*grazir*" (ringraziare). Insieme, formano la costruzione perifrastica che esprime l'azione di dovere ringraziare voi: "*devo ringraziarvi*".
2. "*so q'ieu sai ben far e dir*". Questa costruzione è formata dal pronome "*ieu*" (io) seguita dal verbo "*sai*" (so) e dagli infiniti "*far*" (fare) e "*dir*" (dire). Insieme, formano la costruzione perifrastica che indica la conoscenza dell'abilità di fare e dire: "*sapete che io so ben fare e dire*".

Pur dovendo lavorare su due versi, in questo caso il modello di linguaggio fornisce una risposta perfettamente congrua, poiché risulta in grado di identificare le due costruzioni perifrastiche e di analizzarle correttamente. L'unico errore consiste nella traduzione della seconda perifrasi: qui, infatti, leggiamo un "sappiate che", che secondo la comprensione del chatbot equivarrebbe a "so q'". Ciò malgrado, l'esame sintattico svolto è da ritenersi comunque accettabile, soprattutto data la presenza di una costruzione perifrastica irregolare (doppia, tipologia b). Il margine di errore riguardante unicamente l'analisi descrittiva delle strutture perifrastiche e il loro riconoscimento equivale quindi allo 0%.

L'ultima casistica che presentiamo in questa sede è tratta sempre dal *corpus* di Raimbaut d'Aurenga ed è non molto dissimile da (8): prevede, infatti, un primo verso nel quale sono totalmente assenti costruzioni perifrastiche verbali, e un secondo verso, già esaminato dal chatbot (5), che invece ne ospita una. A differenza dell'esempio esposto in precedenza (9), qui il verso che è stato aggiunto a quello ospitante la struttura perifrastica è il precedente anziché il successivo.

- (11) «*et ab nou bel captenemen
vuoill un bon nou vers commensar*» (BdT 389, 1: vv. 3-4)

Nel testo fornito sono presenti due costruzioni perifrastiche:

1. “*ab nou bel captenemen*”. Questa costruzione è formata dalla preposizione “*ab*” (con) seguita dagli aggettivi “*nou*” (nuovo) e “*bel*” (bello) e dal sostantivo “*captenemen*” (comportamento). Insieme, formano la costruzione perifrastica che indica l’associazione di un nuovo e bel comportamento: “*con un nuovo e bel comportamento*”.

2. “*vuoill un bon nou vers commensar*”. Questa costruzione è formata dal verbo modale “*vuoill*” (volere) seguito dall’articolo indeterminativo “*un*” (un), dagli aggettivi “*bon*” (buono) e “*nou*” (nuovo), e dal sostantivo “*vers*” (verso). Insieme, formano la costruzione perifrastica che indica l’azione di volere iniziare un buon nuovo verso: “*voglio iniziare un buon nuovo verso*”.

Il modello linguistico segnala due costruzioni perifrastiche, di cui solamente la seconda è corretta, ed è la medesima di (5). Mettendo per un attimo da parte l’errore commesso dal chatbot circa il punto 1, osserviamo con più attenzione l’analisi della seconda perifrasi, quella esatta. Notiamo infatti che, benché vi sia stato un complessivo riconoscimento della struttura, lo studio sintattico-descrittivo non corrisponde in pieno a quello effettuato in (5), e presenta alcune problematiche. Anzitutto, il chatbot si preoccupa di esaminare anche quegli elementi terzi interposti fra verbo ausiliare e verbo principale, i quali, come abbiamo visto, rappresentano l’oggetto della perifrasi. Seppure l’analisi di *GPT* sia effettivamente accurata e conforme, è necessario ricordare che le tre “parole-contenuto”, insieme all’articolo, non fanno parte dei componenti della costruzione perifrastica verbale; pertanto, la spiegazione al riguardo è da ritenersi superflua ai fini di questa sperimentazione. In secondo luogo, la descrizione sintattica delle unità perifrastiche presenta una lacuna non indifferente: è assente il verbo all’infinito, il quale, malgrado non venga esaminato, viene comunque tenuto in conto in fase di traduzione. Nel complesso, questo risulta essere il tentativo meno riuscito fra quelli analizzati finora.

Torniamo, comunque, al verso in cui è assente la perifrasi, ma nel quale invece l’*AI* ne registra una. Al pari di (9), abbiamo ritenuto necessario far notare al chatbot l’errore commesso, ma sperimentando una modalità differente per “educare” il modello di linguaggio. Anziché scrivere chiaramente che il costrutto “*ab nou bel captenemen*” non è una perifrasi verbale, abbiamo segnalato a *GPT* che nel testo vi è un’unica struttura perifrastica, senza fare cenno a quale fosse la corretta. Così, è stato lasciato libero arbitrio al chatbot di scegliere quale fosse, a suo parere, la costruzione esatta fra le due individuate. La sua risposta è stata la seguente:

- (12) *Mi scuso per l’errore nella mia risposta precedente. Hai ragione, nel testo fornito è presente solo una costruzione perifrastica: – “vuoill un bon nou vers commensar”. Questa costruzione è formata dal verbo modale “vuoill” (volere) seguito dall’articolo indeterminativo “un” (un), dall’aggettivo “bon” (buono), e dall’aggettivo “nou” (nuovo). Insieme, formano la costruzione perifrastica che indica l’azione di volere iniziare un buon nuovo verso: “voglio iniziare un buon nuovo verso”. Mi scuso per la confusione e grazie per avermelo fatto notare.*

Come avviene consuetamente, il modello linguistico, dando ragione a chi scrive, ammette la confusione e, in questo caso, fornisce la risposta esatta. Era prevedibile, però, che fra le due costruzioni perifrastiche scegliesse quella effettivamente corretta? Evidentemente, isolando le due strutture e dovendo scegliere solamente fra quei due insiemi di parole (senza quindi includere le restanti parti dei due versi), l'intelligenza artificiale riesce a svolgere un lavoro più attento e preciso. Ciò nonostante, anche qui l'analisi sintattico-descrittiva dei componenti perifrastici presenta qualche difficoltà, com'era avvenuto in (11). Sembra, infatti, che *GPT* abbia lavorato esclusivamente sul riconoscimento della struttura, senza preoccuparsi di svolgere una nuova analisi, e riportando quindi quella elaborata in precedenza.

4.6 Analisi delle costruzioni perifrastiche su un testo

Il lavoro effettuato da *GPT-3.5* su due versi ha fornito risultati assai meno efficaci rispetto all'analisi sintattica delle costruzioni perifrastiche svolta sul singolo verso; ciò malgrado, il riscontro del chatbot sembra comunque abbastanza incoraggiante da permetterci di proseguire la sperimentazione.

L'ultimo step di questa ricerca riguarda la registrazione delle perifrasi su una frazione testuale più ampia, come una strofa. Premesso che, come abbiamo visto, il modello linguistico sembra operare meglio su sezioni di materiale molto ridotte, ci chiediamo: è comunque possibile per *GPT* evincere una struttura perifrastica quando essa si trova in un testo non contenente ulteriori perifrasi? La prima casistica che illustreremo è proprio di questo tipo: si tratta di una costruzione, come avvenuto sinora, già presa in esame dall'intelligenza artificiale sul singolo verso (1), inserita questa volta nel contesto della propria stanza poetica.

- (13) «*De lai don plus m'es bon e bel
non vei mesager ni sagel,
per que mos cors non dorm ni ri
ni no m'aus traire adenan,
tro qu'eu sacha ben de la fi,
s'eles aissi com eu deman.*» (BdT 183, 1: vv. 7-12)

Nel testo fornito sono presenti alcune costruzioni perifrastiche verbali. Le costruzioni perifrastiche verbali nel testo sono:

1. "no m'es bon e bel", in cui "es" funge da ausiliare e "bon e bel" è la costruzione verbale principale. La costruzione indica l'attribuzione di un carattere positivo: "è buono e bel".
2. "non vei", in cui "vei" funge da ausiliare e "non" è il verbo principale. La costruzione indica la negazione di un'azione: "non vedo".
3. "mos cors non dorm ni ri", in cui "non dorm" e "ni ri" sono costruzioni verbali perifrastiche che indicano l'azione negativa di "dormire" e "ridere" rispettivamente: "non dormo né rido".
4. "no m'aus traire adenan", in cui "m'aus" funge da ausiliare e "traire" è il verbo principale. La costruzione indica l'azione negativa di "osare" o "avere il coraggio" di fare qualcosa: "non ho il coraggio di tirare".

5. “*qu’eu sacha ben de la fi*”, in cui “*sacha*” funge da ausiliare e “*ben de la fi*” è la costruzione verbale principale. La costruzione indica la conoscenza approfondita di qualcosa: “*so bene della festa*”.

6. “*s’eles aissi com eu deman*”, in cui “*eles*” funge da ausiliare e “*aissi com eu deman*” è la costruzione verbale principale. La costruzione indica una condizione ipotetica: “*se è così come chiedo io*”.

Operare su un testo più lungo come questo (formato da 6 versi e 8 proposizioni totali) è una sfida probabilmente troppo complessa per il chatbot, che va incontro, come notiamo, a un generale disorientamento, fornendo una risposta quasi completamente errata. In altre parole, GPT dà luogo a un elenco di occorrenze (una per ogni verso, in questo caso), che riconosce come costruzioni perifrastiche. Tuttavia, in questa lista, solamente il punto 4 è effettivamente una perifrasi verbale: qui, sia l’analisi sintattico-descrittiva, sia la traduzione equivalgono perfettamente all’esempio (1), laddove la spiegazione dei verbi ausiliare e principale è corretta, mentre la resa in italiano moderno lo è solo parzialmente. Ad ogni modo, benché il modello linguistico abbia saputo riconoscere l’unica struttura perifrastica presente nel testo, si tratta comunque di un solo caso di risposta corretta su sei complessivi, il che equivale ad un tasso di errore dell’83,4%.

In realtà, la casistica appena esposta ci aiuta anche a capire come lavora l’intelligenza artificiale nell’analisi delle costruzioni perifrastiche su una strofa. Nel complesso, infatti, notiamo che la procedura di selezione da parte del chatbot può prendere due differenti strade:

- a) estrarre singoli elementi da una frase (o da un verso), identificandoli come costruzione perifrastica verbale. È l’esempio del punto 2, rispetto a cui GPT afferma che “*vei*” è il verbo ausiliare e “*non*” il verbo principale.
- b) Riconoscere un verso intero come costruzione perifrastica. In questo caso, il verso viene generalmente diviso in due parti, di cui la prima è identificata come verbo ausiliare, mentre la seconda come verbo all’infinito. Prendiamo il punto 6: qui “*eles*” è riconosciuto come ausiliare e “*aissi com eu deman*” come «costruzione verbale principale». L’analisi è chiaramente del tutto errata.

Malgrado i notevoli limiti che riscontra l’intelligenza artificiale quando lavora su una più ampia quantità di materiale, vi è comunque un dato considerevole di questo *modus operandi*. Come si legge nell’esempio (13), ai punti 5 e 6 il chatbot omette dall’analisi le congiunzioni iniziali (rispettivamente, le subordinanti “*si*” e “*tro que*”, entrambe rese con elisione della vocale finale). Si può allora ipotizzare che il modello linguistico di *OpenAI* sia in grado di riconoscere queste congiunzioni e, conseguentemente, di escluderle proprio perché le identifica come tali? Sebbene non sia questa la sede per approfondire la questione, non se ne può negare la rilevanza.

5 Conclusioni e risultati ottenuti: limiti e potenzialità di GPT-3.5

La sperimentazione che è stata presentata in questo contributo, prendendo in esame categorie sintattiche di varia tipologia, ha inteso evincere limiti e potenzialità di GPT-3.5 rispetto all’analisi sintattica effettuata su un testo in versi. Nel complesso, il chatbot si è prestato a molteplici forme

di esperimenti, dimostrando così la sua natura di laboratorio di ricerca sperimentale. Innegabilmente, i risultati positivi di una larga parte dei test effettuati hanno dimostrato le abilità di questo modello di intelligenza artificiale nell'ambito dell'indagine testuale. Di seguito ne discuteremo le principali.

In primo luogo, *GPT-3.5* sa cos'è una costruzione perifrastica verbale: se questo dato potrà sembrare ovvio – trattandosi di un modello linguistico pre-addestrato, al pari di altri “transformers”, e disponendo di un'enorme gamma di dati – non sarà altrettanto prevedibile che il chatbot sappia riconoscere la perifrasi tanto in situazioni standard, quanto in casi per così dire “non standard”. Si ricordino i seguenti casi, applicati sia sul singolo verso, sia su due versi (le due modalità da cui è stato ricavato un minore tasso di errore):

- a) Casi di perifrasi irregolare e discontinua: (5) e (12).
- b) Casi di doppia perifrasi continua reggente su un unico ausiliare: (6).
- c) Casi di due costruzioni perifrastiche sviluppate su due versi consecutivi: (10).

In queste e in molteplici ulteriori casistiche, inoltre, il modello linguistico ha dimostrato di saper svolgere un'analisi sintattica e sintattico-descrittiva approfondita delle unità perifrastiche principali (verbo ausiliare e verbo principale), senza essere tratto in inganno da elementi terzi interposti nella stessa costruzione perifrastica oppure, più in generale, da altre componenti del verso. Fra gli esempi illustrati, infatti, le uniche difficoltà riscontrate da *GPT* si rilevano in (11).

Un secondo dato raccolto da questa sperimentazione fa riferimento alla parzialmente corretta comprensione da parte del chatbot della lingua provenzale, malgrado i limiti circa la conoscenza delle norme grammaticali dell'antico-occitano a cui la stessa intelligenza artificiale faceva riferimento in (2a). Prendendo in considerazione unicamente l'aspetto interpretativo dei materiali sottoposti, infatti, l'*AI* riesce a fornire una traduzione generalmente esaustiva e spesso attinente al significato reale del verso. Da questo punto di vista, l'evidente analogia grafica di alcune “parole-contenuto” provenzali con vocaboli dell'italiano moderno si è presentata come uno strumento bifronte per il modello linguistico generativo. Se in alcuni casi, infatti, da questa somiglianza il chatbot ha potuto trarre beneficio (come si nota nella maggior parte delle traduzioni relative alle analisi sul singolo verso), in altri casi, invece, la corrispondenza grafica ha generato qualche errore, come in (1). Ad ogni modo, gli errori più comunemente riscontrati nella pratica traduttiva riguardano:

- La persona del verbo ausiliare, alternando prima e terza singolari: (2), (3) e (7).
- Il modo e il tempo verbali, con oscillazioni fra presente indicativo e condizionale: (4).

L'ultima, ma non meno rilevante, capacità dimostrata dal chatbot in questa sede, consiste nel fornire l'esito corretto subito dopo avergli fatto notare l'errore commesso, com'è stato illustrato in (12). Meno efficaci, invece, i tentativi di rigenerazione della risposta, che hanno necessitato di un ulteriore intervento manuale per correggere la strategia di lavoro di *GPT*: è il caso di (9).

Chiaramente, accanto alle non comuni risorse offerte da *GPT-3.5*, vi sono anche alcuni grandi difficoltà legate, probabilmente, alle sue ancora limitate applicazioni nell'ambito dell'analisi del testo. Anzitutto, è necessario notare che vi è sempre un minimo margine di aleatorietà e di variazione nelle risposte fornite dal modello linguistico, conseguentemente a cui potrebbero esserci alcune alterazioni nei risultati delle analisi. Tuttavia, il vincolo più rilevante nell'impiego del chatbot ai fini qui illustrati è un altro, e riguarda l'efficienza e l'accuratezza che *GPT* riesce a garantire. In particolare, sembra che la precisione scientifica relativa al lavoro svolto dall'intelligenza artificiale sia inversamente proporzionale alla quantità di materiale testuale su cui deve operare. In altre parole, il chatbot risulta lavorare in modo più produttivo e trovarsi maggiormente a suo agio con piccole sezioni di testo: non a caso, abbiamo visto come, nel caso di singoli versi, sia che si tratti di costruzioni perifrastiche semplici, sia che si tratti di costruzioni perifrastiche più complesse, l'accuratezza che il modello di linguaggio riesce a raggiungere tocca il 100%, senza il minimo margine di errore. Paradossalmente, questo dato sembrerebbe un estremo controsenso rispetto alla presentazione che gli ingegneri e i ricercatori di *OpenAI* hanno fornito di *GPT-3* relativamente al suo apprendimento e alle sue capacità, poiché sappiamo che il chatbot tende a essere più affidabile quando lavora su maggiori quantità di dati¹⁴. Si potrebbe certamente ipotizzare che *GPT* "segmenti" il materiale fornitogli (in questo caso il verso), analizzando i singoli elementi e traendo delle conclusioni, come abbiamo appurato, praticamente sempre giuste e complete. Questa tesi, però, non spiegherebbe come mai, quando si aumenta il materiale da esaminare, l'accuratezza del lavoro del modello linguistico diminuisca, e la "segmentazione" svolta in un primo momento, poi non funzioni più.

A cosa può essere dovuta, allora, questa ragguardevole efficienza raggiunta sui singoli versi analizzati? Perché il chatbot, nel caso di queste analisi linguistiche, ha lavorato meglio su piccole, anziché su ampie, frazioni di testo? L'abbiamo chiesto allo stesso *GPT*, e abbiamo poi tratto le seguenti conclusioni, seppur sempre congetturali.

- I. La struttura del testo è particolarmente articolata e/o ambigua. Il modello di *AI* è in grado di riconoscere le costruzioni linguistiche, ma la sua capacità di analisi potrebbe essere notevolmente ridotta se il testo proposto risulta particolarmente complesso in molteplici sue parti.
- II. La lirica rappresenta una nuova sfida per l'*AI*. Come sostiene il chatbot stesso: «gli algoritmi di elaborazione del linguaggio naturale, come *GPT*, sono addestrati principalmente su testi in prosa, che rappresentano la struttura linguistica più comune». La poesia, al contrario, presentando costruzioni complesse e creative, mette alla prova su un terreno inedito le capacità dell'algoritmo. Si potrebbe obiettare che non si spiega comunque la precisione raggiunta sul singolo verso: occorre però riconoscere che un singolo verso sarà pur sempre meno ambiguo e articolato rispetto a una strofa oppure a un intero componimento. Questa seconda tesi spiegherebbe, inoltre, come mai il chatbot inizi a riscontrare difficoltà già quando deve analizzare due versi anziché uno.

¹⁴ Cfr. Dale 2021, pp. 113-118.

- III. *GPT-3.5* è stato addestrato principalmente su testi moderni e contemporanei, pertanto, potrebbe avere difficoltà nel comprendere e interpretare correttamente le sfumature linguistiche delle lingue medievali, come l'antico-occitano. Possedendo strutture grammaticali e lessicali che differiscono significativamente dalle lingue moderne, infatti, i testi scritti in queste lingue rendono difficile per l'algoritmo identificare correttamente le costruzioni perifrastiche o comprenderne il significato preciso. È quanto, d'altronde, aveva ribadito la stessa intelligenza artificiale in (2a).

In conclusione, è innegabile che i nuovi e potenti modelli di linguaggio generativo si prestino volentieri a sperimentazioni filologico-linguistiche di vario genere. Tuttavia, è altrettanto necessario riconoscere che questi si trovano ancora in una fase estremamente precaria e rischiosa per poter sostenere analisi testuali e linguistiche ad ampio spettro.

Appendice

Le tabelle riportano i dati statistici relativi alla corretta identificazione e allo studio sintattico-descrittivo delle costruzioni perifrastiche verbali nei due *corpora*-campione. Si noti che le statistiche non tengono conto della traduzione resa dal chatbot.

Tabella 1. Quantità di perifrasi verbali riconosciute e analizzate da *GPT-3.5* sul singolo verso.

	Guglielmo IX		Raimbaut d'Aurenga	
Perifrasi regolari totali	52		143	
	N.	%	N.	%
Perifrasi regolari esaminate da <i>GPT</i>	52	100	143	100
Perifrasi regolari non esaminate da <i>GPT</i>	0	0	0	0
Perifrasi irregolari totali	31		136	
Perifrasi irregolari esaminate da <i>GPT</i>	22	71	114	83,9
Perifrasi irregolari non esaminate da <i>GPT</i>	9	29	22	16,1

Tabella 2. Quantità di perifrasi verbali riconosciute e analizzate da *GPT-3.5* su due versi.

	Guglielmo IX		Raimbaut d'Aurenga	
Perifrasi totali	83		279	
	N.	%	N.	%
Perifrasi esaminate da <i>GPT</i>	51	61,5	179	64,1
Perifrasi non esaminate da <i>GPT</i>	32	38,5	100	35,9

Tabella 3. Quantità di perifrasi verbali riconosciute e analizzate da GPT-3.5 su una strofa.

	Guglielmo IX		Raimbaut d'Aurenga	
Perifrasi totali	83		279	
	N.	%	N.	%
Perifrasi esaminate da GPT	38	45,8	124	44,5
Perifrasi non esaminate da GPT	45	54,2	155	55,5

Riferimenti bibliografici

BdT

Alfred Pillet, Henry Carstens, *Bibliographie der Troubadours*, Halle, Niemeyer, 1933.

Boutière - Schutz - Cluzel 1964

Jean Boutière, Alexander Herman Schutz, Irénée-Marcel Cluzel, *Biographies des Troubadours: Textes provençaux des XIII^e et XIV^e siècles*, Paris, Nizet, 1964.

Brambilla 1964

Franca Brambilla Ageno, *Il verbo nell'italiano antico: ricerche di sintassi*, Napoli/Roma, Ricciardi, 1964.

Burrows 2002

John Burrows, "Delta": *A Measure of Stylistic Difference and a Guide to Likely Authorship*, «Literary and Linguistics Computing», 17/3 (2002), pp. 267-287, <https://doi.org/10.1093/lc/17.3.267>.

Canettieri 2013

Paolo Canettieri, *Le impronte digitali dell'autore. Un metodo di attribuzione automatizzata per i testi delle letterature romanze*, «Le forme e la storia», 6/1 (2013), pp. 229-244, <https://doi.org/10.1400/214374>.

Colella 2020a

Gianluca Colella, *I modi del verbo e la modalità*, in Maurizio Dardano (a cura di), *Sintassi dell'italiano antico. La prosa del Duecento e del Trecento. La frase semplice*, vol. II, Roma, Carocci, 2020, pp. 203-233.

Colella 2020b

Gianluca Colella, *Le perifrasi verbali*, in Maurizio Dardano (a cura di), *Sintassi dell'italiano antico. La prosa del Duecento e del Trecento. La frase semplice*, vol. II, Roma, Carocci, 2020, pp. 234-258.

Corti 2005

Maria Corti, *La lingua poetica avanti lo stilnovo: studi sul lessico e sulla sintassi*, a cura di Giancarlo Breschi e Angelo Stella, Firenze, Edizioni del Galluzzo per la Fondazione Ezio Franceschini, 2005.

Ciotti 2023

Fabio Ciotti (a cura di), *Digital Humanities. Metodi, strumenti, saperi*, Roma, Carocci, 2023.

Dale 2021

Rober Dale, *GPT-3: What's it good for?*, «Natural Language Engineering», 27/1 (2021), pp. 113-118, <https://doi.org/10.1017/S1351324920000601>.

Di Girolamo 2021

Costanzo Di Girolamo, *I trovatori*, Torino, Bollati Boringhieri, 2021.

Frasca 1992

Gabriele Frasca, *La furia della sintassi. La sestina in Italia*, Napoli, Bibliopolis, 1992.

Milone 1998

Luigi Milone, *El trobar 'envers' de Raimbaut d'Aurenga*, Barcelona, Columna, 1998.

Pasero 1973

Nicolò Pasero (a cura di), *Guglielmo IX d'Aquitania. Poesie. Edizione critica*, Modena, Mucchi, 1973.

Pattison 1952

Walter T. Pattison, *The Life and Works of the Troubadour Raimbaut d'Orange*, Minneapolis, University of Minnesota Press, 1952.

Picone 1977

Michelangelo Picone, *Osservazioni sulla poesia di Raimbaut d'Aurenga*, «Vox Romanica», 36 (1977), pp. 28-37, <https://doi.org/10.5169/seals-28576>.

Squartini 2011

Mario Squartini, *Verbal Periphrases in Romance. Aspect, Actionality and Grammaticalisation*, Berlin/New York, Mouton De Gruyter, 2011.

Sitografia

BEdT

<http://bedt.it/>

ChatGPT

<http://chat.openai.com>