









# GPT-based Language Models meets Emojitaliano: A Preliminary Assessment Test between Automation and Creativity

Francesca Chiusaroli<sup>1,†</sup>, Tiberio Uricchio<sup>1,\*,†</sup>, Johanna Monti<sup>2</sup>, Maria Laura Pierucci<sup>1</sup> and Federico Sangati<sup>3</sup>

<sup>1</sup>Università degli Studi di Macerata - <sup>2</sup>Università di Napoli "L'Orientale" - <sup>3</sup>OIST Graduate University, Japan

### Introduction & Emojitaliano

- ➤ We started from the intersemiotic version of 'Pinocchio in Emojitaliano', a crowdsourcing project on Twitter (#emojitaliano) in 2016. Pinocchio in Emojitaliano was published as a book by "apice libri pub." in 2017.
- Emojitaliano is an artificial language created by a human community at the intersection between technology and creativity. Starting from the experience of Pinocchio, Emojitaliano continuously implements assigning grammatical rules on the model of an interlanguage and fixing word-emoji matches based on the released Unicode emoji set. Emojitaliano is available on Telegram in a devoted bot, @emojitalianobot (developed by Federico Sangati).
- We aimed at assessing the skills of GPT4 to handle Emojitaliano grammar and glossary in order to build an automatic translator.
- ➤ A test of re-translation of Pinocchio by GPT4 is presented here.
- Preliminary quantitative results show that GPT4 is able to understand Emojitaliano and provide basic translations.

#### Original text

Capitolo I Come andò che Maestro Ciliegia, falegname trovò un pezzo di legno che piangeva e rideva come un bambino.

- C'era una volta.... — Un re! — diranno subito i miei piccoli lettori.
- No, ragazzi, avete sbagliato. C'era una volta un pezzo di legno. Non era un legno di lusso, ma un semplice pezzo da catasta, di quelli che d'inverno si mettono nelle stufe e nei caminetti per accendere il fuoco e per riscaldare le stanze.
- Non so come andasse, ma il fatto gli è che un bel giorno questo pezzo di legno capitò nella bottega di un vecchio falegname, il quale aveva nome mastr'Antonio, se non che tutti lo chiamavano maestro Ciliegia, per via della punta del suo naso, che era sempre lustra e paonazza, come una ciliegia matura.
- Appena maestro Ciliegia ebbe visto quel pezzo di legno, si rallegrò tutto; e dandosi una fregatina di mani per la contentezza, borbottò a mezza voce:
- Questo legno è capitato a tempo; voglio servirmene per fare una gamba di tavolino.

Detto fatto, prese subito l'ascia arrotata per cominciare a levargli la scorza e a digrossarlo; ma quando fu lì per lasciare andare la prima asciata, rimase col braccio sospeso in aria, perchè sentì una vocina sottile sottile, che disse raccomandandosi:

— Non mi picchiar tanto forte! —

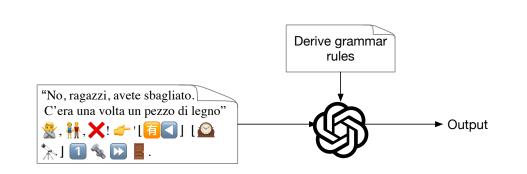
Figuratevi come rimase quel buon vecchio di maestro Ciliegia! Girò gli occhi smarriti intorno alla stanza per vedere di dove mai poteva essere uscita quella vocina, e non vide nessuno! Guardò sotto il banco, e nessuno: guardò dentro un armadio che stava sempre chiuso, e nessuno; guardò nel corbello dei trucioli e della segatura, e nessuno; aprì l'uscio di bottega per dare un'occhiata anche sulla strada, e nessuno. O dunque?...

## 

'| 🔒 🗖 | 📱 🔀 | ৯ 🏋 | 🧷 '00 🚾 🚗 🛨 | 💵 🗶 |. | | | | | | | | | | | | |

**Emojitaliano** 

## **GPT-4** meets Emojitaliano



- We first ensured that GPT-4 did not know about Emojitaliano.
- With an iterative procedure, we gave pairs of Italian text and the corresponding Emojitaliano to GPT-4, asking it to derive grammar rules and the lexicon.

Puoi aiutarmi a derivare delle regole di grammatica?

- Certo, sarei lieto di aiutarti. Il tuo testo è una traduzione di un capitolo del Pinocchio in quello che hai chiamato "emojitaliano". E' una forma di comunicazione che utilizza gli emoji per rappresentare parole o concetti, è un esercizio interessante per il linguaggio simbolico. Ecco alcune regole di grammatica di base che possiamo derivare da questo testo:
  - 1. Parole di frequente utilizzo: Le parole di uso comune o concetti comuni possono essere rappresentate da un singolo emoji. Per esempio, "legno" viene rappresentato da , "falegname" da , "ridere" da , "piangere" da , "fuoco" da .
  - 2. **Nomi propri**: Per i nomi propri come "Maestro Ciliegia", si utilizzano emoji che rappresentano le parole chiave nel nome (\*\*).
  - 3. **Aggettivi e avverbi**: Gli aggettivi e avverbi possono essere rappresentati utilizzando un emoji che rappresenta l'idea di base combinata con altri emoji che modificano il significato. Per esempio, "vecchio" è rappresentato da , "semplice" da (non di lusso), "bel" da , "paonazza" da , (rosso brillante).
  - 4. **Verbi**: I verbi possono essere rappresentati da emoji che rappresentano l'azione di base. Per esempio, "trovò" viene rappresentato da (cercare e trovare), "piangeva" e "rideva" da (e) e sispettivamente, "accendere" da (i), "riscaldare" da (ii).
  - 5. **Fraseologia**: Alcune frasi comuni o idiomatiche sono rappresentate da una combinazione di emoji. Per esempio, "C'era una volta" è rappresentato da "

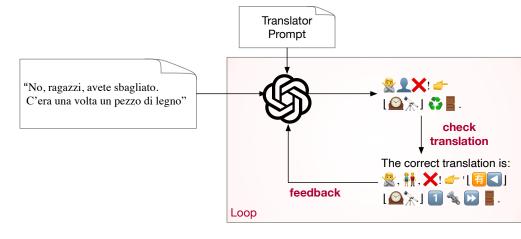
    ' [ ☐ ◀ ] [ △ ↑ ]".
  - 6. **Connettivi e preposizioni**: Alcuni connettivi e preposizioni sembrano essere rappresentati da emoji specifici. Per esempio, "e" da +, "di" da ≥, "in" da √, "per" da ↗, "con" da ⊘.
  - da emoji specifici. Per esempio, "e" da +, "di" da ☑, "in" da ⊋, "per" da ↗, "con" da ∑.

    7. Punteggiatura: La punteggiatura viene utilizzata per indicare la struttura delle frasi, come nella lingua italiana. Ad esempio, le virgolette (") sono utilizzate per indicare il discorso

diretto, i due punti (:) sono utilizzati per introdurre un'elenco o una spiegazione.

- 8. **Sostantivi**: I sostantivi sono rappresentati da emoji che rappresentano l'oggetto o il concetto. Per esempio, "pezzo di legno" è rappresentato da 🎾 📱, "stufe" da 🐠, "camini" da 🕌, "naso" da 👃.
- 9. Numeri e quantità: I numeri e le quantità vengono rappresentate utilizzando le cifre numeriche o gli emoji per le quantità. Per esempio, "una" viene rappresentato da 1, "tutti" da 12. Continue this conversation

## Re-translating Pinocchio



- First test: we verified the translated sentence structures
- Prova a tradurre la seguente frase "Appena riebbe l'uso della parola, cominciò a dire tremando e balbettando dallo spavento"

Spiegazione:

- "Appena" è tradotto come "son".
   "Riebbe l'uso della parola" è tra
- \* "Riebbe l'uso della parola" è tradotto come "i '[, "]".
   \* Qui, "i rappresenta "lui", "[, "] rappresenta l'idea di "riottenere l'uso della parola".
- "[ 😨 🛨 🔈 ] " rappresenta l'idea di "tremare e balbettare", e "🏨" rappresenta "spavento".

Si prega di notare che questa traduzione potrebbe non essere perfetta poiché la traduzione emojitaliano è altamente simbolica e potrebbe variare in base al contesto e alla interpretazione personale.

- Questa è la traduzione corretta "\$\frac{1}{2} \cdot \frac{1}{2} \c
  - Guardando la traduzione che hai fornito, posso notare alcune differenze e potenzialmente identificare alcune regole supplementari nell'emojitaliano:
    - Nomi propri e personaggi: I nomi propri o i nomi dei personaggi, come "Maestro Ciliegia" ([\*\*\*\*\*]), sono racchiusi tra le parentesi [ e ]. Questa regola potrebbe essere utilizzata per distinguere chiaramente i nomi propri o i personaggi da altri concetti o azioni.
       Concetti complessi: Alcuni concetti complessi, come "riottenere l'uso della parola" ([\*\*\*\*\*])
    - '♠:) e "iniziare a parlare" (i '[ a] '♠:), sono espressi tramite una combinazione di emoji e le parentesi [ e ].

      3. **Aggettivi e descrizioni**: Le descrizioni o gli aggettivi, come "tremando" ([ a] ) e

"balbettando" ([ [ ] ac] ), sono anch'essi racchiusi tra le parentesi [ e ]. Questa regola

potrebbe essere utilizzata per distinguere chiaramente le descrizioni o gli aggettivi da altri concetti o azioni.

4. **Emozioni e sentimenti**: Le emozioni o i sentimenti, come "spavento" ( ), possono essere

espressi tramite emoji singoli senza l'uso delle parentesi [ e ].

Ricorda che queste sono soltanto mie interpretazioni basate sulla traduzione fornita, e che l'emojitaliano, come tutte le lingue simboliche, può avere regole e convenzioni che variano a seconda del contesto e dell'interpretazione personale.

- Second test: we verified how GPT-4 treats figurative language
- Prova a tradurre questa frase "E perchè gli era entrato addosso una gran paura, si provò a canterellare per farsi un po' di coraggio."
- Utilizzando le regole di grammatica dell'emojitaliano che abbiamo dedotto finora, la frase "E perché gli era entrato addosso una gran paura, si provò a canterellare per farsi un po' di coraggio" può essere tradotta come segue:

Rivediamo le componenti della frase:

- La frase originale è "E perché gli era entrato addosso una gran paura, si provò a canterellare per farsi un po' di coraggio."

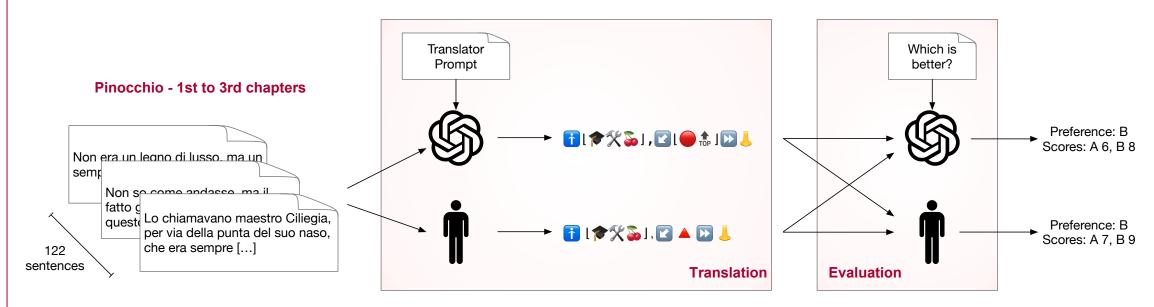
  La tua traduzione in emojitaliano è " + ∠ [ ♦ ※ 5] '[ 6] ◊ ◄] ↓ ▼ 1 ♀ ♣, 1 '[ ﴿ ◄] '[ ﴿ ♣] '[ ﴿ ♣] '[ ﴿ ﴿ ﴿ ♣] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] 'الْ ﴿ ﴾] '[ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿ ﴾] '[ ﴿ ﴿

**ቇ**░ ↗'[██♥┺] [Ŵ�]"

Dalla tua traduzione possiamo dedurre alcune ulteriori regole:

- 2. "→ 💌 📵 🙊 🍰 ": Questa combinazione sembra riferirsi a "addosso una gran paura". L'emoji → potrebbe rappresentare "addosso" o "su di sé", l'emoji 💌 potrebbe rappresentare "un", e l'emoji 😡 rappresenta "paura". L'emoji 🏚 può essere utilizzato per indicare un intensificatore come "grande" o "molto".
- 3. ""[ ♣ ◀]": Questa combinazione sembra rappresentare "si provò". Quindi, l'emoji ♣ potrebbe rappresentare un tentativo o un esperimento, o in questo caso "provare".
- 4. "'[ • • ] ": Questa combinazione sembra rappresentare "canterellare". Quindi, l'emoji potrebbe rappresentare "parlare" o "dire", l'emoji rappresenta "cantare", e l'emoji potrebbe essere utilizzato per indicare un diminutivo o una forma più piccola o più leggera di un'azione, in questo caso "canterellare".
- 5. "'[➡�→]": Questa combinazione sembra rappresentare "farsi". Quindi, l'emoji ➡ e l'emoji → potrebbero essere utilizzati per indicare un'azione rivolta a se stessi, in questo caso "farsi".
- 6. "[ @ ]": Questa combinazione sembra rappresentare "un po". Quindi, l'emoji @ potrebbe essere utilizzato per indicare una piccola quantità di tempo o un breve periodo, o in questo caso "un po".
- ➤ After several iterations, GPT-4 was able to correctly respect grammar and creatively translate text into interesting combinations of emojis.

#### **Evaluation**



Human evaluators put more emphasis on the correct structure of the sentences while GPT-4 generally reported better scores for creativity and direct matching of the emojis.

	<b>GPT-4 Evaluation</b>			<b>Human Evaluation</b>		
Translation by	Human	GPT-4	Equal	Human	GPT-4	Equal
Preferences	27	51	17	46	36	13
Average Score	7.23	7.80	/	7.34	7.21	/

#### **Conclusions**

- ➤ The intensive crowdsourcing experience made Emojitaliano a unique case among the few examples of integral translations in emoji.
- The emoji set is continuously expanded. Translating Pinocchio into Emojitaliano today would certainly involve new, and sometimes more relevant, pairings, synonymic pairs that do not exclude the previous ones.
- ➤ Teaching Emojitaliano to GPT-4 (and the like) does not mean replacing a human translator with a machine, but rather is like having a tool to enhance human work.