statuS Quaestionis
language text culture

Patrizia Giampieri
Università degli Studi di Perugia

Are NMT and LLMs effective in the translation of dialogues?

Abstract

AI has made huge strides forward and is increasingly applied in language learning and audiovisual translation. This paper explores the quality of the machine- and LLMs-driven translation of excerpts of naturally occurring conversations scrutinised by scholars. The dialogues are sourced from literature analyses and examples. After outlining the traits of spoken language, this paper delves into an examination of the renderings of the transcripts of spoken interactions from Italian into American English. To do so, the DeepL NMT platform and the ChatGPT and Gemini chatbots are taken into account. The paper findings bring to the fore three main aspects: 1) NMT is qualitatively less accurate and reliable than LLMs-generated translations; 2) human intervention in LLMS-driven output is necessary because of sporadic inconsistencies and mistranslations, and 3) LLMs prompts must be written carefully and mindfully in order to obtain consistent and accurate results, as unspecific prompts may give rise to a less satisfactory output.

Keywords: spoken interactions; LLMs-driven translations; NMT; LLMs; the grammar of spoken language

1. *Introduction*

1.1 *Aim of the study*

The aim of this paper is to explore and assess the quality of automated translations (from Italian into English) of extracts of spoken conversations. It examines the machine-driven rendering of dialogue excerpts retrieved from naturally occurring interactions. The conversations are sourced from literature studies and investigations on spoken Italian. In other words, this paper scrutinises excerpts of authentic dialogues carried out by Italians which were reported and analysed by scholars in academic works. By doing so, it compares machine-powered translations with the renditions into English proposed by researchers. In this way, it is possible to assess the quality and reliability of LLMs- and NMT-driven outputs.

1.2 *The grammar of spoken English*

Spoken language is unstructured, or loose, and it does not follow the same grammatical rules of the written mode (Thornbury 2005; Carter and McCarthy 2015). Conversations, in fact, are real-time, face-to-face and context-related (Carter and McCarthy 2015).

Spoken interactions are characterised by a context-dependant shared visual element (Conrad 2022), where some components are taken or given for granted by the conversation participants. There are many features which characterise the grammar of speaking. Some of them are discourse markers, turn takers, sharedness markers, face-savers, response tokens, approximators, reformulation devices, deixis, and non-words.

Discourse markers shape the conversation experience and are topic-managing devices (De Cristofaro et al. 2022: 125-126). They can be multi-words of various length (such as “you know” and “and that sort of things”) (McCarthy and McCarten 2019: 5), single words (such as “well” or “so”) (O’Keeffe et al. 2007: 39), and non-words (such as “mm” or “erm”) (O’Keeffe et al. 2007: 142; McCarthy 2010: 7). They serve different purposes. For example, the two-word unit “you know” is very frequent in spoken conversations (McCarthy and McCarten 2019: 5) and, amongst others, it expresses shared knowledge. It can also function as a topic launcher (McCarthy and Carter 2019: 43-44). The phrase “I guess” is a face-saver and a hedging device (Erman 2001: 1341) as it softens the impact of an utterance. Single words such as “well” and “so” can be topic launchers and opening markers, given that they signal the beginning of a shift in the discussion (O’Keeffe et al. 2007: 34).

The single word “so” signals the conclusion of a previous statement or discourse (Carter et al. 2011: 616). Therefore, it serves summarising purposes. Some discourse markers are also response tokens, as they are used as flow-sustaining devices and instances of listenership (O’Keeffe et al. 2007: 141-145; McCarthy 2010). Examples in this regard can be single words such as “yeah”, “right”, phrases such as "is that right?", or even non-words like "mm" (O'Keeffe et al. 2007: 141-145). Spoken interactions also rely on approximation via face-saving strategies. Examples are the expressions "kind of" (and the more colloquial "kinda"), "sort of" (or "sorta"), and "things like that" (Erman 2001; O'Keeffe et al. 2007). In addition, O'Keeffe et al. (2007: 210) refer to the chunk "a lot of" as a vagueness token, given that it does not precisely quantify amounts. Approximators and vague language are used not only to reduce the speaking processing time, but also to take and give things for granted. In this way, they rely on shared knowledge (McCarthy 2010). For instance, the word "like" can be both a vague category marker and a hesitation device (Andersen 2001: 210). Reformulation tools are also used in spoken conversations. They occur when interactants try to refine a previous statement in a more apt way (O'Keeffe et al. 2007: 172), so that their reformulated words or statements function as discourse monitors. Examples are phrases such as "if you like", "so to speak", "I mean", etc. As mentioned, conversations are context-related. For this reason, deictic elements often come to the fore, such as "here", "there", "this", or "that". Finally, non-word vocalisations may have different functions; for example, "er" is a pause filler (Rayson et al. 1997) and is uttered to give thinking time to the speaker; "hmm" and "oh, mm" are pondering devices, and "uh uh" is a response token (Fadden 2008: 124).

1.3 *Neostandard Italian*

Neostandard Italian is a variety of the Italian language defined as *italiano dell'uso medio parlato e scritto* (Sabatini 1985: 171, "average spoken and written Italian"). Scholars posit that it is closer to oral than written language (Ballarè 2020: 476). Be it written or oral, they agree on the fact that it is commonly used by educated people in both formal and informal contexts (Sabatini 1985; see also Berruto 2017: 33). There are many traits characterising neostandard Italian, such as the actualising *ci*, suspended conditionals, the overextension of male object pronouns, left and right dislocations, etc. (Sabatini 1985; Lombardi Vallauri 2016; Giampieri 2025).

The actualising *ci* (Berruto 2017; Miola 2023) is a redundant, clitic pronoun. It is redundant because it does not serve any locative or semantic purpose, as in *c'è* ("there is") or *crederci* ("believe in it") (Sabatini 1985). Also, it should not be interpreted as referring to the plural first-person pronoun (as in *dacci*, "give us"). It often pairs with verbs such as *avere* ("have") and *essere* ("be") to form verbal expressions such as *averci* or *esserci* (Sabatini 1985: 160; Berruto 2017:

46). Another example of the actualising *ci* is in the following question: *ce l'hai le carote?* ("have you got them, the carrots?") (Berruto 2017: 46).

Suspended conditionals are subordinate clauses which do not appear with a main clause (Lombardi Vallauri 2016: 151). This is possible because in spontaneous conversations, interactants infer meanings from contexts. For instance, the following statement features a suspended conditional clause: *d'altra parte va be', se ti sei scordato* ("anyway, all right, if you forgot") (ibid., 147-148). As can be seen, the subordinate clause *se ti sei scordato* is not accompanied by a main clause.

In the overextension of *gli* (Berruto 2017: 45; Miola 2023: 98), the *gli* male object pronoun (literally "to him") can refer to males, females, and multiple persons. Therefore, expressions such as *gli parlo* ("I talk to him") may actually regard a male or female figure, as well as a group of people.

Left and right dislocations divide a sentence into two parts: left dislocations feature a theme/rheme preference, whereas right dislocations show a rheme/theme order (D'Achille 2019). Left and right dislocations contain a clitic pronoun which functions as an anaphoric or cataphoric element. The clitic pronoun re-states (or retains) a following or preceding object. An example of a right dislocated element appears in the statement *li leggo i giornali* (Berruto 2017: 40) ("I read them, newspapers"). In this case, *li* is a clitic pronoun anticipating the following object (*i giornali*). In left dislocations, the above utterance would be changed into *i giornali li leggo* ("the newspapers, I read them"), where the clitic *li* follows the object.

Finally, it is worthwhile mentioning that spoken Italian, be it standard or neostandard, is hallmarked by truncations and abbreviations (Berretta 1994: 243). They mainly arise from the spontaneity of the speech event and are due to the fact that conversation participants can infer (and fill in) pieces of information from the context.

Discussing the traits of spoken English and spoken (neo)standard Italian is relevant in the present study given that, as stated, it compares NMT- with LLMs-driven translations of dialogue extracts while highlighting whether and to what extent the above elements emerge.

1.4 *LLMs and MT*

In light of the huge strides forward made by LLMs in the last few years, there are numerous applications of LLMs-driven software in language learning and translation. Scholars have investigated how LLMs-based agents and chatbots can improve language learning (Belda-Medina and Cavol-Ferrer 2022; Ericsson and Johansson 2023). Belda-Medina and Cavol-Ferrer (2022) explored how conversational LLMs can be integrated by future educators in language learning classes. The authors posit that the participants' perception on LLMs performances is overall positive, especially with regard to their usefulness and user-friendliness. Nonetheless, the respondents seem somehow reluctant as concerns immediate usages of LLMs as language learning tools. As a matter of fact, they are generally keen to develop interactions and conversations with humans, rather than with LLMs agents (ibid.). Human interactants, in fact, can give supplementary information or advice when needed. Ericsson and Johansson (2023) study the ways LLMs-driven conversational agents can help lower secondary school students improve second language (L2) skills. The students conversed with LLMs interactants to perform day-to-day tasks. At the end of the LLMs-driven experience, they felt emotionally and socially involved and showed overall improvements in their language knowledge.

LLMs-powered solutions have also been applied to improve L2 speaking skills. Vančová (2023) explores the effectiveness of mobile and web applications, chatbots, and intelligent virtual assistants in L2 pronunciation practice. The findings point to a relevant increase in the quality of L2 proficiency. In addition, LLMs-driven solutions can both raise L2 learners' motivation and reduce their oral production anxiety. Other scholars have focused on the use of LLMs in translation. Baños (2023) and Moorkens and Guerberof Arenas (2024), for example, have addressed the audiovisual translation (AVT) field.

Several scholars have scrutinised the translation performances of NMT- and LLMs-powered tools. Jiao et al. (2023) compare target texts produced by ChatGPT and Google Translate. They posit that the chatbot provides satisfactory outputs with recurrent language pairs, whereas it lacks consistency with less frequent language combinations. They also argue that ChatGPT is more effective in the translation of spoken language. Grimm et al. (2024) explore how GPT-4 LLMs help foster medical communication. To this aim, the authors examine the translation performance of GPT-4 systems into English, Spanish and Mandarin. They find that such tools provide accurate medical information in different languages, thereby improving access to medical services. In the same way, Keles et al. (2024) focus on the medical sector and develop an LLM model which delivers satisfactory target texts. As far as the legal field is concerned, Moneus and Sahari (2024) investigate the reliability of translations generated by chatbots such as ChatGPT, ChatSonic and Microsoft Copilot. They postulate that the LLMs do not grasp legal subtleties. For this reason, their outputs are inaccurate. Similarly, Giampieri (2024) discusses the advantages and drawbacks of relying on intelligent machines in the legal domain. More precisely, Giampieri (2024) scrutinises the automated rendition (from English into Italian) of an arbitration clause performed by ChatGPT. The findings point to the fact that LLMs cannot be entirely relied upon as they produce shortcomings at both word and phrase level. This is mostly due to the fact that legal discourse is formulaic and it contains polysemous terms whose disambiguation in context may be challenging. Additionally, it is grounded in legal-system specificities.

Therefore, despite some apparent advantages of LLMs- and NMT-based solutions, there are many instances where they prove to be ineffective and/or unreliable. In this respect, Federici et al. (2023: 151) warn against the pitfalls and allure of NMT- and LLMs-driven systems and claim that “volunteerism and technology lull users into a false security of the translation quality achieved through automated processes”. Also, as mentioned, there are instances of inaccuracies and lexical issues (Giampieri 2023 and 2024), as well as social inequalities and potential discrimination favoured by (biasedly trained) automated systems. For example, AI solutions may not address less widespread languages (Nunes Vieira et al. 2021).

To reduce AI-generated risks of inequality, Federici et al. (2023: 151-152) suggest embedding access and usages in platforms, where applications would be constrained or limited to a certain degree. In addition, NMT and LLMs systems may give rise to “hallucinations” (Escribe and Mitkov 2023; Guerreiro et al. 2023), i.e., outright nonsensical output.

1.5 *Research questions*

The research questions that this paper wishes to address are the following ones: 1) Can LLMs- and NMT-driven solutions translate colloquial language in such a manner as to reproduce traits of spontaneous conversations?; 2) What are the differences in the two types of output?, and 3) Which are the most common shortcomings?

To answer these questions, dialogue extracts sourced from literature analyses on spoken Italian are focused on, together with the related suggested (or official) translations into English performed by scholars. For the purpose of this investigation, such dialogues are also rendered automatically by using LLMs and NMT platforms. Therefore, the LLMs- and NMT-generated target texts are compared with the translations proposed by researchers. In this way, traits of spoken Italian are scrutinised in both human-made and automatically generated renditions.

2. *Methodology*

As mentioned, four dialogue extracts (and the related human-made translations into English) are sourced from literature investigations on spoken Italian (Lombardi Vallauri 2016; Voghera and Borges 2017; Miola 2023). The dialogues have been chosen for their richness in spoken language features; hence, for their representativeness of spontaneous interactions and spoken neostandard Italian, according to literature-based studies.

The conversations are translated automatically by using an NMT-driven platform (such as DeepL) and two chatbots (i.e., ChatGPT and Gemini). The language pairs in DeepL are Italian and American English. The prompt inputted in the chatbots is as follows:

Prompt: *Translate into colloquial American English the following dialogue: "[Dialogue]".*

To test the soundness and consistency of the LLMs-generated outputs, at the end of the analysis another prompt is inputted in the chatbots. For this purpose, one of the four source texts is copied in the chatbots dialogue box. The following different prompt is formulated:

Prompt 2: *Translate the following text into English: "[Dialogue]".*

In this way, it is possible to verify the level of consistency of the results obtained when the prompt does not refer to colloquial language.

In the sections that follow, the source dialogue, the official translation (carried out by scholars) and the automated translations into English (i.e., the one by DeepL, ChatGPT, and Gemini) are reported into tables. As mentioned, the official translations are those appearing in scholarly works. Researchers may have chosen to render dialogues rather literally to emphasise spoken features.

Traits of spoken language in the source and target texts are discussed. The quality of the translated output is examined on the basis of lexical, syntactical and semantic aspects, as well as in light of the spoken language traits discussed in the literature.

Therefore, this paper delves into a discussion of the results based on lexical, syntactical and semantic aspects. Additionally, the accuracy and reliability of the rendition of spontaneous language are scrutinised according to the spoken (standard and/or neostandard) language traits as referred to by scholars (Sabatini 1985; Berretta 1994; Lombardi Vallauri 2016; Berruto 2017; D'Achille 2019; Ballarè 2020; De Cristofaro et al. 2022; Miola 2023).

3. *Analysis*

This section analyses and discusses extracts of Italian conversations among two or more interactants and the related official and automatically generated translations into (American) English.

Each table reports an Italian dialogue excerpt, the translation proposed by scholars, and the automated translations performed by DeepL, ChatGPT and Gemini.

The first dialogue is sourced from Lombardi Vallauri (2016: 148). Translation inaccuracies or mistranslations are marked in bold.

Dialogue in Italian	**Official Literal Translation**	**Translation by DeepL**	**Translation by ChatGPT**	**Translation by Gemini**
A: Non so chi gli ha dato un biglietto pe anda' a vede' Costanzo Show chissà se li hanno fatti entra' *B: Ah be' perché no, se ci hai se ci hai l'invito.* *A: Ci aveva il biglietto per due persone.*	A: I don't know who gave them a ticket for Costanzo Show who knows if they were let in. B: Well, why not if you have if you have the invitation. A: He had the tickets for two people.	A: I don't know who gave them a ticket to go see Costanzo Show, I wonder if they let them in. B: Ah well why not, **if you have us** if you have the invitation. A: **There** had the ticket for two people.	A: I don't know who gave them a ticket to go see the Costanzo Show. I wonder if they got in. B: Oh, well, why not? If you have an invite... A: **They** had a ticket for two people.	A: I don't know who gave him a ticket to go see the Costanzo Show. I wonder if they let him in. B: Well, why not? If you have an invitation, **you get in**. A: He had a ticket for two people.

Table 1. Dialogue 1 and related translations

The Italian dialogue shows truncated forms, such as *pe anda'* (meaning *per andare*, "to go"), *vede'* (*vedere*, "see"), and *entra'* (*entrare*, "enter"). There are also discourse markers (*ah be'*), actualising *ci* phrases (i.e., *ci hai l'invito, ci aveva il biglietto*) and a suspended conditional (*se ci hai l'invito*). As can be seen from the statements in Table 1, *ci* retains or re-states the objects that follow (i.e., *l'invito* and *il biglietto*, respectively).

In the official translation, the suspended conditional is rendered literally (i.e., "if you have the invitation"), whereas the redundant *ci* is not addressed.

In the target text by DeepL, the *ci* phrases are mistranslated; in the first case (i.e., *ci hai l'invito*), *ci* is rendered with the first personal pronoun "us": "if you have us"; in the second case (namely, *ci aveva il biglietto*), *ci* is considered as a place adverb and translated as "there".

In the target text by ChatGPT, there is no repetition (the phrase *se ci hai se ci hai l'invito* is translated as "if you have an invitation"). However, the suspended conditional is emphasised by adding suspension dots ("if you have an invite..."). In the final statement (*ci aveva il biglietto*), the subject is plural ("they"), whereas it should be singular ("s/he"). Such a misinterpretation is probably due to the fact that, as mentioned, in spoken language, the pronoun *gli* (literally "to him", appearing in the first statement) can refer to both females and multiple persons.

In the output by Gemini, there is also no repetition, and the suspended conditional phrase is turned into a standard conditional clause ("if you have an invitation, you get in").

As a whole, it can be stated that NMT gives rise to issues related to the fact that the system does not disambiguate words in context. For this reason, *ci* is either considered as a substitute of a plural first-person pronoun ("us") or a place adverb ("there").

In the chatbot-generated translations, no such shortcomings appear. In particular, *ci* is interpreted correctly, although no redundant (or clitic) pronoun is displayed. Additionally, the suspended conditionals are addressed satisfactorily, although Gemini produces some extra language output.

The next dialogue extract is sourced from Voghera and Borges (2017: 68)

Dialogue in Italian	**Official Literal Translation**	**Translation by DeepL**	**Translation by ChatGPT**	**Translation by Gemini**
S2: Eeh il naso del cane? *S1: Eeh è.* *S2: È rotondo?* *S1: È rotondo, sì.* *S2: C'ha una c'ha una cosetta bianca al centro?* *S1: Sì come se fosse il coso.* *S2: Una mezzaluna diciamo.*	S2: Eh the dog's nose? S1: Is it. S2: Is it round? S1: It is round, yes S2: Does it have a white little thing in the middle? S1: Yes, like it was a thing. S2: A half moon let's say.	S2: Eeh the dog's nose? S1: Eeh is it. S2: Is it round? S1: **Is it round**, yes. S2: Does it have a little white thing in the middle of it? S1: Yes like **it's the thing**. S2: A crescent moon.	S2: Uh, what about the dog's nose? S1: Uh, it's… S2: Is it round? S1: Yeah, it's round. S2: Does it have, like, a little white thing in the middle? S1: Yeah, like a, you know, that thing… S2: Like, sorta like a half-moon, you could say.	S2: Hey, what about the dog's nose? S1: Yeah, what about it? S2: Is it round? S1: Yeah, it is. S2: Does it have like a little white thing in the middle? S1: Yeah, kinda like a… you know. S2: Like a half-moon, I guess.

Table 2. Dialogue 2 and related translations

In the Italian dialogue, *eeh* can be considered as a filled pause; the clitic pronoun *ci* appears in *c'ha*, whereas *una cosetta* ("a little thing"), *il coso* ("the/that thing") and *diciamo* ("say") are instances of approximation.

The official translation does not render the redundant *ci* (as also seen in the previous dialogue). The expression *una cosetta* and *il coso* are translated literally (i.e., "a little thing" and "a thing"), and *diciamo* becomes "let's say".

The DeepL platform turns the statement *è rotondo, sì* into a question (i.e., "is it round, yes") and the phrase *sì come se fosse il coso* (meaning "yes, as if it were that thing") is (mis)translated as "yes like it's the thing".

The two chatbot-driven translations start with "uh" (in ChatGPT) or "hey" (in Gemini). The former is a pause filler, whereas the latter is an attention marker. The dialogues then continue with "what about the dog's nose?", which renders *il naso del cane?* The expression "what about" is added in both cases. Gemini-driven output also contains "yeah, what about it?" in the second turn. Hence, it adds language output as the source expression is only *eeh è* ("uh, it's"). Although an equivalent of the utterance "what about it" does not appear in the source text, the target expression can be considered as adequate[1]. In addi-

---

[1] The expression "what about it" generates 225 hits in the spoken section of the COCA corpus (Davies 2008), where it functions as a new turn.

tion, both chatbot-assisted translations produce a sharedness marker (i.e., "you know"), as well as vagueness markers, such as "that thing", "sorta like", and "you could say" (in ChatGPT), or "kinda like", and "I guess" (in Gemini).

Therefore, it can be stated that the target text generated by DeepL contains some inconsistencies and inaccuracies. Conversely, the translations produced by the two chatbots resemble authentic language as they contain several elements of spoken interactions (although sometimes such traits could be considered as redundant). The next excerpt is sourced from Miola (2023: 7).

Dialogue in Italian	Official Literal Translation	Translation by DeepL	Translation by ChatGPT	Translation by Gemini
TOI003: Ha perso tantissime cose Torino. *TOR001: Mh.* *TOI003: Se uno pensa.* *TOI002: No però si è arricchita parecchio con le olimpiadi a pa guarda prima non c'era.* *TOI003: A partire da un macello di cose che io mi ricordo quando andavo a scuola potevi fare un casì era diventata la città delle delle mh.* *TOI002: Mado' ma prima tu vede mado' ma tu prima vedevi turismo a Torino* *TOI003: Del libro e poi l'ha spostata a Milano.*	TOI003: Torino lost a lot of things. TOR001: Uhm. TOI003: If one thinks. TOI002: No but it developed a lot with the Olympics – look, before there wasn't. TOI003: Starting from a lot of things that I remember when I still went to school you could do a lot it became the city of of uhm. TOI002: God but before you saw God but before you saw tourism in Torino. TOI003: Of books and then they moved it to Milano.	TOI003: **Lost so many things Turin.** TOR001: Mh. TOI003: If one thinks. TOI002: No though**' she** got a lot rich with the Olympics **a pa** look before there was no. TOI003: Starting with a mess of things that I remember when I was in school you could make a **casì** had become the city of mh. TOI002: **Mado'** but first you see **mado'** but you used to see tourism in Turin. TOI003: Of the book **and then moved it** to Milan.	TOI003: Turin lost a ton of stuff. TOR001: Mm-hmm. TOI003: If you think about it... TOI002: No, but it gained a lot with the Olympics though, didn't it? I mean, before that, it didn't have much. TOI003: Yeah, starting from a bunch of stuff. I remember when I was in school, you could do all kinds of things. It had become the city of, like... uh... TOI002: Oh my god, but before, did you see... oh my god, did you see **any** tourism in Turin before**?** TOI003: The book fair, and then they moved it to Milan.	TOI003: Turin has lost so much. TOR001: Hmm. TOI003: If you think about it... TOI002: No, but it got a lot richer after the Olympics. Look at it now, it wasn't like that before. TOI003: Starting from **a total mess**! I remember when I was in school, it was **a disaster**. It had become the city of... of... TOI002: Oh man, but before that, you know, there was tourism in Turin. TOI003: It was all about books. Then they moved it to Milan.

Table 3. Dialogue 3 and related translations

In the source transcript, there are several traits of spoken language. For example, in the first statement (*ha perso tantissime cose Torino*), the subject is dislocated to the right periphery; there is a truncated false start (*a pa*, probably referring to *a parte* (*ciò*), "besides"), and truncated expressions such as the word pair *un casì* (from un casino, meaning "a whole lot" in the context) and *mado'* (a truncation of the word *Madonna*, used, in this case, to express surprise or to emphasise one's opinion). The colloquial *un macello di cose* refers to an array of things. There are also repetitions, such as *delle delle* ("of the of the") and *ma prima tu vede* (…) *ma tu prima vedevi* ("but before you sa… but you saw before").

In the official translation, *un casì* is rendered as "a lot"; *un macello di cose* as "a lot of things", and *mado'* as "God". Repetitions also appear (i.e., "it became the city of of" and "but before you saw God but before you saw").

The translation by DeepL contains several syntactical errors, as well as other translation infelicities, such as the verbatim reproduction of truncated expressions in the source language (e.g., *a pa* and *mado'*). Syntactical shortcomings are visible in the phrase "lost so many things Turin", which could have been reformulated as "it lost so many things Turin", thus featuring the same right dislocation as in the source text. Also, there is a wrong subject pronoun (i.e., "she" instead of "it" in the phrase "she got a lot rich" referring to Turin), and an incorrect (or a missing) subject in "and then moved it to Milan", instead of "and then they moved it to Milan" or "and then it was moved to Milan".

The target text by ChatGPT features no truncation, although there are other traits of spoken language, such as tags or confirmation tokens ("didn't it?"); reformulation devices ("I mean"), and responses or engagement tokens ("yeah"). An instance of repetition also appears (*ma prima tu vede mado' ma tu prima vedevi* rendered as "but before, did you see… oh my god, did you see (…) before"). However, the other repeated expression (i.e., *la città delle delle mh*) is changed into approximation: "the city of, like… uh". In addition, the initial suspended conditional *se uno pensa* is added suspension dots ("if you think about it…"). The colloquial *un macello di cose* is rendered as "a bunch of stuff", whereas *fare un casì* becomes "do all kinds of stuff", and *mado*' is translated as "oh my god". The last two sentences are misinterpreted. The question "did you see any tourism in Turin before?", by ChatGpt, should actually be an affirmative statement, namely "you saw (a lot of) tourism in Turin before" (source text: *tu prima vedevi turismo a Torino*). Moreover, the turn that follows must relate to this statement and start with a dependent clause, such as "related to the book fair" (or "of books", as in the official translation) (source text: *del libro*). The translation, conversely, features an independent clause ("the book fair"), which is unrelated to the other interactant's last statement. The phrase also contains the word "fair", although the word *fiera* was never mentioned in the source text.

The text translated by Gemini features the following instances of spoken language: "oh man", rendering *mado'*, sharedness markers ("you know"), and suspension dots in the suspended conditional phrase ("if you think about it…"). There are also repetitions ("it had become the city of… of…"). However, *un macello di cose* and *un casì* are given a negative semantic prosody (i.e., a negative meaning) and are mistranslated as "a total mess" (instead of "a lot of things") and "a disaster" (instead of "a lot"), respectively. The last lines, conversely, are rendered correctly ("there was tourism in Turin", "it was all about books").

The translation by DeepL shows some inaccuracies, such as a wrong word order (or a neglected right dislocation) ("lost so many things Turin"), missing or wrong pronouns ("and then moved it to", instead of "and then they moved it to", and "she", instead of "it"), and neglected truncated colloquialisms (*casì, mado'*).

The target texts by the two chatbots feature interesting colloquial language (e.g., "a bunch of stuff", "oh my god", "oh man"). On the other hand, they contain some inaccuracies (e.g., the incoherent last turn in ChatGPT, and the mistranslation of *un macello di cose* and *un casì* by Gemini).

The next and last dialogue extract is sourced from Miola (2023: 12).

Dialogue in Italian	**Official Literal Translation**	**Translation by DeepL**	**Translation by ChatGPT**	**Translation by Gemini**
TOR004: E com'è che siete finiti là. *TOI054: Non lo so da questa superstrada che dava la cartina oltretutto ero andata con Aldo eh.* *TOR004: Pensa te.* *TOI054: E quindi lui era uno che sapeva girare nel senso guardare la cartina non era un uno che si perdeva ecco.*	TOR004: And how is it that you ended up there? TOI054: I don't know. TOI054: From this highway the map said. TOI054: Besides I went with Aldo uh. TOR004: Just think. TOI054: And so he was one that he knew how to travel I mean look at the map he wasn't a one that got lost okay.	TOR004: And how is it that you **finished** there. TOI054: I don't know from this freeway **that gave the map** moreover I had gone with Aldo eh. TOR004: **Think of you** and so he was someone who knew how to turn in the sense of looking at the map he wasn't one to get lost here.	TOR004: So, how'd you end up there? TOI054: I'm not sure. It was off this highway that was on the map, and I was with Aldo, you know. TOR004: No kidding. TOI054: Yeah, and he was good with directions. Like, he knew how to read the map and didn't get lost or anything.	TOR004: So how did you end up there? TOI054: I don't know, from this highway the map showed. Besides, I went with Aldo, you know. TOR004: Wow, I can't believe it. TOI054: And he was the kind of guy who knew how to get around, you know? He was good at reading maps and never got lost.

Table 4. Dialogue 4 and related translations

In the Italian text, there are deictic elements (*questa superstrada*, "this freeway"); response elicitors (eh, "uh"); engagement or response tokens (*pensa te*, literally "you think", but meaning "really?"); reformulation devices (*nel senso*, "I mean"), and fillers (*ecco*, "now").

The official translation addresses the above elements by resorting to the following traits: a demonstrative adjective to render deixis ("this highway"); the non-word "uh" to translate *eh*; the discourse marker "I mean" as a reformulation strategy; the filler "okay" translating *ecco*, and the literal "just think" as a response token (source text: *pensa te*).

In the translation by DeepL, there are lexical, syntactical and semantic issues, such as the incorrect "finished there" (instead of "ended up there"); "from this freeway that gave the map" (instead of "from this freeway that the map showed"), and "think of you" which should be a surprise token (such as "really?" or "no way"). The final filler *ecco* is untranslated.

The outputs by ChatGPT and Gemini appear more engaging. The surprise token *pensa te* is rendered as "no kidding" by ChatGPT and as "wow, I can't believe it" by Gemini. The reformulation token (i.e., *nel senso*) is changed into approximation ("like", by ChatGPT) or sharedness ("you know", by Gemini). Finally, *ecco* becomes "or anything" in ChatGPT, whereas it is unaddressed by Gemini.

As a whole, DeepL performs the worst. The two chatbots, conversely, tackle the conversations slightly better. In particular, engagement tokens, sharedness, and vague category markers are used to render responses and reformulations.

4. Discussion

The sections above showed how NMT output may lack accuracy and can give rise to shortcomings at lexical, syntactical and semantic level. Examples in this regard are the following expressions: "lost so many things Turin"; "make a *casì*"; "from this freeway that gave the map"; "how is it that you finished there", and "think of you". For this reason, dialogues sometimes sounded unauthentic and the language output was overall unsatisfactory.

The conversations produced by the two chatbots, by contrast, featured many traits of spoken language and seemed more natural. Amongst others, the LLMs resorted to reformulation devices (e.g., "like"), approximation (e.g., "all kinds of things"), and sharedness ("you know"). Apparently, ChatGPT performed slightly better, although there were cases where both chatbots generated incoherent texts (e.g., "they had a ticket", by ChatGPT, instead of "he had a ticket", by Gemini; "did you see any tourism in Turin before?", by ChatGPT, instead of "you saw tourism in Turin before", by Gemini, or "a total mess", by Gemini, compared to the more context-adequate "a bunch of stuff", by ChatGPT). Therefore, LLMs seem to provide and resort to colloquial or spontaneous-sounding expressions. Nonetheless, they still lack disambiguation capabilities, thus giving way to translation infelicities. For these reasons, human supervision is necessary to ensure consistency.

It is now interesting to verify whether different prompts may generate the same or different LLMs outputs. For this purpose, the source text of Table 4 is used one more time and the following different prompt is formulated:

Prompt 2: *Translate the following text into English: "[Dialogue]".*

As can be seen, in this case no reference to a dialogue or colloquial American English is made. The results are reported in Table 5 below. The first and third columns show the initial target texts by ChatGPT and Gemini as appearing in Table 4 (hence, they are the translations by ChatGPT and Gemini resulting from the first prompt). The second and fourth column report the new target texts by ChatGPT and Gemini (i.e., they are the translations obtained from the second prompt). Mistranslations are marked in bold. Words that are neglected in the second automated translation process are bold italicised in the first and third columns.

ChatGPT (first prompt – colloquial language)	**ChatGPT (second prompt – standard English)**	**Gemini (first prompt – colloquial language)**	**Gemini (second prompt – standard English)**
TOR004: So, how'd you end up there? TOI054: I'm not sure. It was off this highway that was on the map, and I was with Aldo, you know. TOR004: No kidding. TOI054: ***Yeah***, and he was good with directions. ***Like***, he knew how to read the map and didn't get lost ***or anything***.	TOR004: And how did you end up there? TOI054: I don't know, from this highway that was on the map, and I had gone with Aldo, you know. TOR004: Really? TOI054: And so he was someone who knew how to navigate, **meaning** he could read the map; he wasn't someone who got lost.	TOR004: So how did you end up there? TOI054: I don't know, from this highway the map showed. Besides, I went with Aldo, ***you know***. TOR004: Wow, I can't believe it. TOI054: And he was the kind of guy who knew how to get around, you know? He was good at reading maps and never got lost.	TOR004: So how did you end up there? TOI054: I don't know, from this **superhighway** the map indicated. Besides, I went there with Aldo. TOR004: **Imagine that**. TOI054: And so he was someone who knew how to get around, **meaning** he could read a map and wasn't the type to get lost.

Table 5. LLMs-generated translations resulting from two different prompts

It is noticeable that there are a number of differences in the two sets of target texts. In ChatGPT, the second column shows no confirmation token (i.e., "yeah", bold italicised in the first column) or approximation device (namely, "like" and "or anything", also bold italicised in the first column). Also, the word "meaning" is used to literally translate *nel senso*.

In Gemini, “superhighway” is a mistranslation. The new target text also omits the sharedness marker “you know” after “Aldo” (bold italicised in the third column). In addition, the expression “wow, I can’t believe it” is turned into a plainer “imagine that”, and “meaning” literally renders *nel senso*.

In light of the above, it can be stated that prompts can greatly affect LLMs-generated results. Therefore, formulating accurate prompts is crucial.

5. *Conclusions*

The research questions posed by this paper were the following ones: 1) Can LLMs- and NMT-driven solutions translate colloquial language in such a manner as to reproduce traits of spontaneous conversations?; 2) What are the differences in the two types of output?, and 3) Which are the most common shortcomings?

The answer to the first question is mostly negative. Although the LLMs-driven translations featured elements of seemingly spontaneous language, the target texts contained a number of inconsistencies and infelicities.

In answering the second question, the differences in the types of outputs have been discussed at length in the sections above. Namely, LLMs-driven translations apparently perform better, although some incoherent rendering comes to the fore. Therefore, human supervision and post-editing are mandatory.

The most common shortcomings (third and last question) are concerned with lexical and grammatical issues, as well as with inaccurate traits of spoken language (especially if prompt formulation is too general or unspecific).

The paper findings revolve around three main aspects: 1) NMT-powered translation is qualitatively less accurate and reliable than LLMs-generated translations as far as the translation of dialogues is concerned; 2) human intervention in LLMs-driven output is necessary because of sporadic inconsistencies or mistranslations, and 3) LLMs prompts must be written carefully and mindfully in order to obtain consistent and accurate output, as different prompts may give rise to completely different target texts.

With regard to the first aspect, this paper showed that NMT target texts are not completely accurate, at least as far as spontaneous dialogues are concerned. For this reason, the automatic translation produced by standard NMT platform is unreliable. NMT, in fact, gives rise to syntactical, semantic and lexical issues, revolving around word order, wrong pronouns, wrong interpretations of the source text, non-detection of truncated forms of colloquialism, and so on. As concerns the second finding, LLMs-driven dialogues are by far more accurate and they seem more naturally sounding. Namely, they contain elements of spoken language (i.e., sharedness markers, approximators, response tokens, engagement tokens etc.). Nonetheless, their target texts may sometimes be inaccurate. In some cases, Gemini added lexical content, although this did not heavily affect or modify the final output. In addition, some inconsistencies could be noticed, such as statements turned into questions; wrong interpretations of colloquial expressions (in particular, wrong semantic prosody), and incorrect agreement. Therefore, human intervention is still necessary in order to cater for such shortcomings. Thirdly, this paper highlighted how prompts must be formulated as carefully as possible, because omitting specific information or details may lead to different (or inconsistent) outcomes.

As a whole, the results obtained corroborate scholars' scepticism about simplistic or unscrutinised approaches to machine-driven translations (Federici et al. 2023; Giampieri 2023 and 2024; Moneus and Sahari 2024). To some extent, the findings by Jiao et al. (2023) are confirmed, as LLMs performed better than NMT in the rendition of spoken language. Nonetheless, users should not be enticed by the allure of fast and easily obtained automatic translations, as many pitfalls lie ahead. Although the language patterning produced by LLMs are natural-seeming, human supervision cannot be ruled out.

The automatic translation of spoken language into English and/or Italian is a complex task that ought to take into account several elements. As noticed in the sections above, there are syntactical and lexical aspects which play a significant role. Therefore, the patterns of spoken language can vary considerably in the two languages. Indeed, there are different ways in which meanings and nuances are conveyed, due to diverse and distinct linguistic structures and/or cultural values. Automated systems should be trained to address such nuances accurately to produce acceptable translations and renderings.

For reasons of space, this paper did not deal with the translations of dialogues into British English. Therefore, this is an area that other researchers might consider in the future.

The limits of this paper lie in the fact that only a few dialogue excerpts were taken into account. In addition, the dialogue extracts were not particularly long and they envisaged only a few turns. Therefore, scholars could tackle more extended dialogues by retrieving them from corpora of spoken language. In this way, they could tap into varied situations and conversations.

There are also other limitations. Namely, the present investigation can only be regarded as an exploratory study of a few dialogue excerpts, but not as a systematic study or assessment of machine-driven translation quality.

Bibliography

Andersen, Gisle. 2001. *Pragmatic Markers and Sociolinguistic Variation*. Amsterdam: John Benjamins.

Baños, Rocío. 2023. "Key challenges in using automatic dubbing to translate educational YouTube videos." *Linguistica Antverpiensia*, New Series: Themes in Translation Studies, 22: 61-79.

Belda-Medina, Jose, and Calvo-Ferrer, José R. 2022. "Using Chatbots as AI Conversational Partners in Language Learning." *Applied Sciences* 12, no. 17. https://doi.org/10.3390/app12178427.

Berretta, Monica. 1994. "Il Parlato Italiano Contemporaneo". In *Storia della Lingua Italiana. Vol 2° (Scritto e Parlato)*, edited by Luca Serianni and Pietro Trifone, 239-270. Torino: Einaudi.

Berruto, Gaetano. 2017. "What is Changing in Italian Today?". In *Towards a New Standard: Theoretical and Empirical Studies on the Restandardization of Italian*, edited by Massimo Cerruti, Claudia Crocco and Stefania Marzo, 31-60. Berlin / Boston: De Gruyter Mouton.

Carter, Ronald, and McCarthy, Michael. 2015. "Spoken Grammar: Where Are We and Where Are We Going?". *Applied Linguistics* 1: 1-21.

Carter, Ronald, McCarthy, Michael, Mark, Geraldine, and O'Keeffe, Anne. 2011. *English Grammar Today*. Cambridge: Cambridge University Press.

Conrad, Susan. 2022. "What can a Corpus tell us about Grammar?". In *The Routledge Handbook of Corpus Linguistics.* Second Edition, edited by Anne O'Keeffe and Michael J. McCarthy, 221-234. New York: Routledge.

D'Achille, Paolo. 2019. *L'Italiano contemporaneo*. Bologna: Il Mulino.

Davies, Mark. 2008. *The Corpus of Contemporary American English (COCA): 520 million words*, 1990-present. http:/corpus.byu.edu/coca/ (accessed: 11/11/2025).

De Cristofaro, Elisa, Crocco, Claudia, Badan, Linda, and Plevoets, Koen. 2022. "Discourse Markers and Turn-Planning at The Pragmatics-Prosody interface: The Case of Allora in Spoken Italian". *Journal of Pragmatics* 198: 125-141.

Ericsson, Elin, and Johansson, Stefan. 2023. "English speaking practice with conversational AI: Lower secondary students' educational experiences over time." *Computers and Education: Artificial Intelligence* 5. https://doi.org/10.1016/j.caeai.2023.100164.

Erman, Britt. 2001. "Pragmatic markers revisited with a focus on you know in adult and adolescent talk". *Journal of Pragmatics*, 33: 1337-1359.

Escribe, Maria, and Mitkov, Ruslan. 2023. "Applying Incremental Learning to Post-editing Systems: Towards Online Adaptation for Automatic Post-editing Models." In *Corpora and Translation Education, Advances and Challenges*, edited by Jun Pan and Sara Laviosa, 35-62. Singapore: Springer Nature Singapore.

Fadden, Lorna. 2008. *Prosodic profiles: suspects' speech during police interviews*. PhD dissertation. Burnaby, BC, Simon Fraser University.

Federici, Federico M., Declerq, Christophe, Díaz Cintas, Jorge, and Baños Piñero, Rocío 2023. "Ethics, automated processes, machine translation, and crises." In *Towards responsible machine translation: Ethical and legal considerations in machine translation*, edited by Helena Moniz, and Carla Parra Escartín, 135-156. Berlin: Springer. https://doi.org/10.1007/978-3-031-14689-3_8.

Giampieri, Patrizia. 2023. *Legal Machine Translation Explained: MT in legal contexts*. Newcastle Upon Tyne: Cambridge Scholars Publishing.

Giampieri, Patrizia. 2024. "AI and the BoLC: Streamlining legal translation". *Comparative Legilinguistics* 58: 67-90.

Giampieri, Patrizia. 2025. *Spoken English and Spoken Italian: Grammar and Translation*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Grimm, David R., Lee, Yu-Jin, Hu, Katherine, Liu, Longsha, Garcia, Omar, Balakrishnan, Karthik, and Ayoub, Noel F. 2024. "The utility of ChatGPT as a generative medical translator". *Eur Arch Otorhinolaryngol* 281: 6161-6165. https://doi.org/10.1007/s00405-024-08708-8.

Guerreiro, Nuno M., Voita, Elena, and Martins, André F.T. 2023. "Looking for a needle in a haystack: A comprehensive study of hallucinations in neural machine translation." *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, May 2023. https://doi.org/10.48550/arXiv.2208.05309.

Jiao, Wenxiang, Wang, Wenxuan, Huang, Jen-tse, Wang, Xing, Shi, Shuming, and Tu, Zhaopeng. 2023. “Is ChatGPT a Good Translator? Yes with GPT-4 as the Engine”. *ArXiv*. 2301.08745. https://doi.org/10.48550/arXiv.2301.08745.

Keles, Bunyamin, Gunay, Murat and Caglar, Serdar I. 2024. “LLMs-in-the-loop Part-1: Expert Small AI Models for Bio-Medical Text Translation”. *ArXiv* abs/2407.12126 https://doi.org/10.48550/arXiv.2407.12126.

Lombardi Vallauri, Edoardo. 2016. “Insubordinated Conditionals in Spoken and Non-Spoken Italian”. In *Insubordination*, edited by Nicholas Evans and Honoré Watanabe, 145-169. Amsterdam: John Benjamins.

McCarthy, Michael. 2010. “Spoken fluency revisited.” *English Profile Journal* 1, no. 1. https://doi.org/10.1017/S2041536210000012.

McCarthy, Michael, and McCarten, Jeanne. 2019. “Interaction Management in Academic Speaking.” *Linx* 79: 1-19. https://doi.org/10.4000/Linx.3611.

Miola, Emanuele. 2023. “The New Ne: An Incipient Restandardization Process in Contemporary Italian”. *Sociolinguistica* 37, no. 1: 95-113. https://doi.org/10.1515/soci-2022-0019.

Moneus, Ahmed M., and Sahari, Yousef. 2024. “Artificial intelligence and human translation: A contrastive study based on legal texts.” *Heliyon* 10. https://doi.org/10.1016/j.heliyon.2024.e28106.

Moorkens, Joss, and Guerberof Arenas, Ana. 2024. “Artificial intelligence, automation and the language industry”. In *Handbook of the Language Industry: Contexts, Resources and Profiles*, edited by Gary Massey, Maureen Ehrensberger-Dow, and Erik Angelone. Berlin, 71-98. Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110716047-005.

Nunes Vieira, Lucas, O’Hagan, Minako, and O’Sullivan, Carol. 2021. “Understanding the Societal Impacts of Machine Translation: A Critical Review of the Literature on Medical and Legal Use Cases.” *Information, Communication & Society* 24 no. 11: 1515-1532. https://doi.org/10.1080/1369118X.2020.1776370.

O’Keeffe, Anne, McCarthy, Michael, and Carter, Ronald. 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.

Rayson, Paul, Leech, Geoffrey, and Hodges, Mary. 1997. "Social differentiation in the use of English vocabulary: some analyses of the conversational component of the British National Corpus." *International Journal of Corpus Linguistics* 2, no. 1: 133-152.

Sabatini, Francesco. 1985. "'L'Italiano dell'uso Medio': Una Realtà tra le Varietà Linguistiche Italiane". In *Gesprochenes Italienisch in Geschichte und Gegenwart*, edited by Günter Holtus and Edgar Radtke, 154-184. Tübingen: Gunter Narr Verlag.

Thornbury, Scott. 2005. *How To Teach Speaking*. Harlow: Pearson Education.

Vančová, Hana. 2023. "AI and AI-powered tools for pronunciation training." *Journal of Language and Cultural Education* 11, no 3: 12-24.

Voghera, Miriam, and Borges, Carla. 2017. "Vagueness Expressions in Italian, Spanish and English Task-Oriented Dialogues". *Normas* 7, no. 1: 57-74. http://dx.doi.org/10.7203/normas.7.10424.

Online resources
ChatGPT: https://openai.com/chatgpt
DeepL: https://www.deepl.com/en/translator
Gemini: https://gemini.google.com/

Patrizia Giampieri, MSc, PhD, is an Associate Professor of English at the University of Perugia (Italy). She is a member of the Executive Board of EST (European Society for Translation Studies). She has authored books and academic papers on spoken language, English as a second language, legal English and legal translation, corpus-based translation, machine translation, AI-driven translation, EAP, and English for Special Purposes. Her last scientific monograph is titled "Spoken English and Spoken Italian: Grammar and Translation" (Cambridge Scholars Publishing, 2025).