Emojilingo: Harnessing AI to Translate Words into Emojis

Francesca Chiusaroli[1,*,†], Federico Sangati[2,†], Johanna Monti[3,†], Maria Laura Pierucci[3] and Tiberio Uricchio[5]

[1]*University of Macerata, Italy*
[2]*Okinawa Institute of Science and Technology, Okinawa, Japan*
[3]*Università Orientale di Napoli*
[4]*University of Macerata, Italy*
[5]*University of Pisa, Italy*

Abstract

This paper presents an AI experiment of translation into emoji conducted on a glossary from Dante Alighieri's *Divine Comedy*. The experiment is part of a project aiming to build up an automated emoji-based pivot language providing an interlingua as a tool for linguistic simplification, accessibility, and international communication: Emojilingo (emojilingo.org). The present test involves human (Emojitaliano) and machine (Chat-GPT) translations in a comparative analysis in order to devise an automated integrated model highlighting emojis' expressive ability in transferring senses, clarifying semantic obscurities and ambiguities, and simplifying language. A first evaluation highlights Chat-GPT's ability to deal with a classic archaic literary vocabulary, also raising issues on managing criteria for better grasping the meanings and forms and about the multicultural extent of content transfer.

Keywords

Emoji, Intersemiotic Translation, Emojitaliano, Emojilingo, Large Language Models

1. Introduction

Consisting today in 3,782 icons, regularly updated by Unicode Consortium,[1] the emoji international catalogue contains signs for 😀 facial expressions, 👷 human gestures, 🤸 people activities, 👨‍🌾 jobs, 🌿 plants, 🐊 animals, 🍝 food, 💡 objects, 🚆 symbols of travel, 🗼 places, 🇵🇱 flags, 🔢 numbers, and 🟥 geometrical forms.

While the visual content seems to be able to provide an encyclopedic list with universal significance, ideally capable of conveying language-independent meanings, the interpretation of emojis is, on the contrary, highly arbitrary. They are strongly subject to ambiguities and variations due to linguistic, cultural, and personal specificities.

The use of emoji has considerably increased over time, and besides complementing written texts in online messages and posts as a nice means to express feelings and emotional statuses, emojis are also used to completely replace verbal language statements [1, 2].

Experiments have been carried out to explore the feasibility of using emojis as language to convey meanings through emoji-only translations. Notable examples include the popular *Emoji Dick* project, the translation into emoji of *Moby Dick* [3], or *Wonderland* [4], an emoji poster created in 2014 to reproduce the full story of Lewis Carroll's *Alice in Wonderland*. These earliest experiments, however, lack codification and, as such, cannot be considered as a language, that is, a shared system in the Saussurean sense [5]. The first translation, in fact, was crowdsourced in a free and creative way, while the second one was an individual and literal translation experiment from English.

A concrete attempt to create a truly codified emoji language can be represented by *Emojitaliano* [6].[2] Emojitaliano is an emoji code originated from a crowdsourcing experiment initiated by a social community, specifically created to share a common emoji language able to counteract the natural polysemy of emojis.[3]

Born with the translation of Collodi's *Pinocchio, The Story of a Puppet* [7] (figure A.1), the structure and glossary of Emojitaliano have been afterwards usefully reapplied for the translation of texts of different genres such

CLiC-it 2024: Tenth Italian Conference on Computational Linguistics, Dec 04 — 06, 2024, Pisa, Italy

*Corresponding author.
†These authors contributed equally.
f.chiusaroli@unimc.it (F. Chiusaroli); federico.sangati@gmail.com (F. Sangati); jmonti@unior.it (J. Monti); m.pierucci6@unimc.it (M. L. Pierucci); tiberio.uricchio@unipi.it (T. Uricchio)
https://docenti.unimc.it/f.chiusaroli (F. Chiusaroli); https://fede.sangati.me (F. Sangati); https://unifind.unior.it/resource/person/2569 (J. Monti); https://docenti.unimc.it/m.pierucci6 (M. L. Pierucci); https://www.dii.unipi.it/user/1906 (T. Uricchio)
0000-0003-1923-3974 (F. Chiusaroli); 0000-0001-6088-415X (F. Sangati); 0000-0002-4563-5988 (J. Monti); 0000-0003-3637-2757 (M. L. Pierucci); 0000-0003-1025-4541 (T. Uricchio)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[1]https://unicode.org/emoji/charts/full-emoji-list.html

[2]https://www.treccani.it/vocabolario/emojitaliano_res-2f30d44e-89c2-11e8-a7cb-00271042e8d9_%28Neologismi%29

[3]https://www.scritturebrevi.it

CEUR-WS.org/Vol-3878/24_main_long.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

as the technical declaratory prose of the Italian Constitution (figure A.2), the Manifesto of non-hostile communication (figure A.6), the narrative prose of classic moral tales (i.e., The Wolf and the Lamb, in figure A.5), Giacomo Leopardi's lyrical poem *L'infinito*[4] (The infinite, in figure A.4).

Emojitaliano is based on the assessment of conventional meanings and syntax, capable of guaranteeing the sharing of sense by means of intersemiotic translation, beyond subjective interpretations.[5] Emojitaliano provides a grammatical structure and a shared vocabulary which can be expanded and re-shared with each new translation[8].[6]

Recent experiments have opened new research horizons in evaluating the capability of large language models (LLMs) to translate words or text into emojis. This is predicated on the assumption that, given LLMs are trained on extensive corpora sourced from the internet, they have been exposed to emojis and are able to grasp the semantics of emojis [9]. Recently, Text2Emoji [10] was proposed as an automatic translator, based on a large text-emoji parallel corpus, created by prompting the LLM, Chat-GPT (OpenAI, 2023) and EmojiLM, a sequence-to-sequence model specialised in the text-emoji bidirectional translation. Another translation experiment involving emojis, conducted by [11] is Emojinize. This experiment leverages the power of LLMs to translate text by considering both prior and subsequent contexts, which differs from next-token prediction. Emojinize disambiguates synonyms based on context, unlike a static lookup table, and harnesses the expressive power of combining multiple emojis.

Among the experiments, a first attempt using Chat-GPT to learn the Emojitaliano grammar was also carried out in 2023 by the Emojitaliano research group. Assuming the fundamental role of a conventional syntax as a basis for each shared code [12, 13], the aim was to verify the ability of LLMs to learn and reapply the Emojitaliano grammar rules to produce translations of Pinocchio on its own [14].

In this paper we present a follow-up experiment of automatic translation into emoji, focused on special vocabulary. Chat-GPT's translations of an authorial lexicon have been tested and then compared to the corresponding human solutions.The purpose is to test LLMs capabilities in autonomously rendering complex vocabulary, in the horizon of building a translation tool into emoji as a means of language simplification: the general project and the conlang itself are named Emojilingo and available online on emojilingo.org.

The paper is organised as follows: section 2 introduces the Emojilingo project *Parole di Dante*, the subject being translations in emoji of 365 words (*Parole di Dante*) from Dante's poem *Divine Comedy*. Section 3 presents the AI translation experiment carried out with two versions of Chat-GPT (3.5 and 4) [15] on the 365 Dante's words, with a focus on the method and descriptions of some examples. Section 4 provides an evaluation of the results, also obtained through AI models and through a similarity matrix, and the closing section includes conclusions and ideas on future work.

2. Emojilingo: *Parole di Dante*

The Emojilingo project is presented here as a follow up of Emojitaliano. The general idea is that, through the Emojitaliano community as control group, LLMs technologies can develop and speed up the processes of translation, enable wider and easier dissemination of the code, overcome the barriers of natural languages. The Emojilingo.org website republishes some Emojitaliano translations with English versions (see also Appendix A).

Our translation method pursues a program of conceptual linguistic simplification which can clarify linguistic meanings for the needs of international communication as well as for plain language policies [16]. The outcome never aims to replace the original sources, but rather intends to provide a vehicular code useful to approach and directly understand words in any language [17, 18].

The current work focuses on a new translation, based on Dante's poem *Divine Comedy*, titled *Parole di Dante* (*Dante's words*). It is well known that Dante's vocabulary may be difficult to understand for foreign speakers, and that a similar difficulty may occur for Italian speakers today, too, because of the many archaic or disused poetical words in the poem. Consequently, we believe that translating this vocabulary into emojis can help mitigate these comprehension difficulties and facilitate understanding.

Parole di Dante consists of 365 emojis which are the Emojitaliano translations of 365 Dante's words. The source, together with the original context and explanatory comments, was published during 2021 as a daily social media dissemination event by the Italian Accademia della Crusca.[7] On that occasion, through the participation by the Emojitaliano Twitter/X community,[8] the #emojitaliano and #scritturebrevi social community produced emoji matches for all the 365 words, one per day. *Parole di Dante* (*Dante's words*) is therefore a glossary of Dante Alighieri's *Divine Comedy* translated into emojis, with the corresponding Italian words.

[4] https://www.scritturebrevi.it/?submit=Search&s=emojitaliano

[5] The Emojitaliano grammar and glossary are registered on @emojitalianobot on Telegram; 3998 members are signed up to @emojitalianobot as of October 16th, 2024.

[6] https://www.scritturebrevi.it/emojitalianobot

[7] https://accademiadellacrusca.it/it/dante

[8] The main accounts involved are @fchiusaroli and @stellissa/Maria Stella Bottai.

Table 1
Translation experiment with Chat-GPT 3.5 (IT) and Chat-GPT 4 (IT) of Dante's "Divina Commedia". For details see section 3.1.

Term	Verse	AC Explanation	Emojitaliano	GPT-3.5 (IT)	GPT-4 (IT)	GPT-4 (IT) Explanation
lurco	Come talvolta stanno a riva i burchi, / che parte sono in acqua e parte in terra, / e come là tra li Tedeschi lurchi	'Mangione, ingordo, beone', latinismo derivato da lurcare 'mangiare avidamente'. Nel verso Dante attribuisce "lurchi" a "li tedeschi".	👄🦛	🦉🍃	🐖🍽️	The pig emoji represents gluttony and greed, while the plate emoji recalls the act of overeating. These two emoji together symbolise a "lurk"-a spoiled and gluttonous individual.
baiulo	Di quel che fé col baiulo seguente, / Bruto con Cassio ne l'inferno latra, / e Modena e Perugia fu dolente. (Paradise Vi, 73)	A Latinism used by Dante to refer to the 'bearer' of the sign of the empire, i.e., the emperor: Dante was convinced that the universal institution of the ancient empire of Rome continued even in his time, with Arrigo VII.	📌🎖️👑👨‍🦲	👜	🚜💼	The 'baiulo' is a professional figure who boasts of transporting something, so I chose the emoji of a worker (rural but representing the ethics of work) and a suitcase to represent the transportation.
accaffi	Poi l'addentar con più di cento raffi, / disser: "Coverto convien che qui balli, / sì che, se puoi, nascosamente accaffi"	Battuta sarcastica dei diavoli a un barattiere; accaffare è voce plebea dal significato certo di 'arraffare, acciuffare', ma d'incerta provenienza.	🦅	🎩🤡	🧗⬇️	An emoji depicting a person climbing or scaling, followed by a downward arrow to represent the action of falling or plummeting.

During a university course exercise,[9] the glossary was later provided with the English matches.[10] A comparison between emoji solutions and English correspondences was discussed in the classroom, bringing out the feasibility of a chain translation method "text-emoji-text". As a subsequent step, the research group tried a LLMs experiment of translation from Italian and English into emojis. The resulting combination of human and AI translations, appropriately selected as will be shown below, is *Parole di Dante* in Emojilingo.[11]

3. The AI translation experiment

In this section we present the translation experiment of the 365 Italian terms from Dante's *Comedy* with Chat-GPT. On the Emojilingo website, the translations chosen by Chat-GPT 4 (from the final evaluation explained in section 4.3) are available.[12]

The very large database on which the LLM architecture is based makes us assume that the machine knows the original text, belonging to the world literary canon, and we may also assume that it knows the English version of the work, as well as it will presumably have available multilingual Dante's glossaries and commentaries. Unlike human translators, who translated on the basis of the explanations provided by the Accademia della Crusca, we decided that the only input to be given to the machine would be author's name and title of the work, instead. This approach allowed us to test Chat-GPT's "autonomous" ability to handle this special lexicon directly.

3.1. Translation Examples

In table 1 we present some examples of the translation experiment with Chat-GPT 3.5 and Chat-GPT 4 for the rendering of some terms, either rare or unusual, or now dismissed. The columns in the table are the following:

- the original Dante's *term*;
- the original *verse* containing the word;
- the explanation by the *Accademia della Crusca* (AC);
- the crowdsourced *Emojitaliano* translation;
- the *Chat-GPT 3.5 (IT)* translation;
- the *Chat-GPT 4 (IT)* translation, and
- its *explanation* by Chat-GPT 4 (IT) itself (translated into English for dissemination).

3.2. Methods

For our experiment we chose two models of Chat-GPT: Chat-GPT 3.5 (turbo-0125), and Chat-GPT 4 (0613) as our reference models, to examine the differences in machine translations between the two models. In the second phase of the project, we compared and evaluated these two versions against the human translations (Emojitaliano) .

To automatically translate the words into emojis with Chat-GPT we adopted a zero-shot prompting approach

[9] https://docenti.unimc.it/f.chiusaroli/courses/2023/28680
[10] Sourced from https://dante.princeton.edu.
[11] https://emojilingo.org/parole_di_dante_about
[12] In addition, the full data is publicly available both on an online spreadsheet (https://docs.google.com/spreadsheets/d/13vkH3a-C0OpVTm9r5daFg_y0MN8lPASwGICaa72zaGg and on GitHub (https://github.com/EmojiLingo/emojilingo.github.io/tree/main/_chatgpt)).

using OpenAI APIs.[13] Despite the archaic and often obscure vocabulary, as already mentioned, no preliminary training was provided and the only context given in the prompt was the reference to the work's title. The prompt was provided both in Italian and English (the latter using English terms), although the final evaluation was done using the Italian version. The English prompt is here provided:

> I will give you a word from Dante's Divine Comedy and ask you to invent a translation in emoji. Respond with a single translation in 2 lines of plain text (without formatting):
>
> - translation into emoji
>
> - brief explanation of the choice.
>
> The word is '{term}'.

In the next section we will present some comments as well as some evaluation remarks.

4. Evaluation

4.1. Initial Comments

The following are some initial comments on the examples reported in Table 1.

lurco All the translations use an animal to represent the negative qualities expressed in the text, likely due to a plausible interference from the English word 'lurk'. Chat-GPT 3.5 focuses on the environmental nocturnal context instead of the vice of gluttony. The choice of a specific animal, as the pig or the wolf, to convey negative semantic values, reflects an Eurocentric view, which raises issues for the multilingual and multicultural reception.

baiulo The human translator reproduces the complex meaning of the archaic word for "Emperor" as "bearer of the sign of Empire", while Chat-GPT translates it more simply as the action of "carrying", which simplifies but clarifies the direct meaning.

accaffi Both the human and Chat-GPT 4 translations convey the semantic value of rapid movement and aggression, while the Chat-GPT 3.5 version emphasizes the sarcastic tone used to depict despicable characters. The issue of the symbolic representation of the animal icons also emerges here.

4.2. Preliminary study

According to a preliminary evaluation, it is immediately apparent that both versions of Chat-GPT can provide interesting translation solutions of Dante's words and are able to motivate their choices in a meaningful way.

One initial observation is that the translation solutions provided by the two Chat-GPT models often include multiple emojis, with Chat-GPT 3.5 doing so 88% of the time and Chat-GPT 4 at 81%. In contrast the Emojitaliano shows a higher tendency to use single emojis, doing so in 49% of the cases.

The two Chat-GPT versions rarely provide the same translations, except for some terms related to animals, such as 'colubro' or 'lonza'. In some instances, particularly those involving realia (e.g., 'eagle', 'angel', 'book', 'galaxy'), the translations provided by Chat-GPT align with those given by human translators.

In most cases, however, the solutions generated by Chat-GPT 4 differ, as do the accompanying explanations.

The differences between the translations provided by the two versions of Chat-GPT are most often largely disparate. For example, the phrase 'dolenti note' is translated by Chat-GPT 3.5 as 😢🎵 and by Chat-GPT 4 as 🔥🎶. Additionally, there are differences in the ordering of emojis, as observed in the translation of 'occhi di bragia', where Chat-GPT 3.5 uses 🔥👀 and Chat-GPT 4 uses 👀🔥. In other instances, while both versions include a common emoji, they are paired with different additional emojis; for instance, 'inanellare' is translated by Chat-GPT 3.5 as 🔄💍 and by Chat-GPT 4 as 💍⛓️; 'colubro' is rendered with the snake 🐍 in all cases, but Emojitaliano adds the skull 🐍☠️ to convey the accurate meaning of the poisonous animal, as derived from the Accademia della Crusca comment. In some cases, Chat-GPT translations correctly grasp the core idea of the word but dismiss the figurative strength of the original: 'intuarsi', meaning 'intimate and deep understanding' and 'interpenetration between minds', is in fact one of Dante's original coinages (see also 'infuturarsi', etc.). Chat-GPT 3.5 versions of 'intuarsi' as 🔍🔮 and 🔍💡 appear not so poignant as the human literal solution seems more expressing 🧍🪄👆. Sometimes the Chat-GPT version succeeds in reproducing the sense more physically than the human one, as for 'trasumanar' 'to rise above the human', 🧬➡️👴 by Chat-GPT 4 compared with the human version 🪶 and Chat-GPT 3.5 translation 🚀🌌.

In a few cases, the translation solutions provided by both versions of Chat-GPT misinterpret the intended meaning of Dante's word. For example, Chat-GPT 3.5 translates 'zeba' (meaning 'goat') as 🐃, erroneously conflating it with a similar-looking word. Similarly, Chat-GPT 4 translates the term as 🐃💩, misinterpreting it as 'cattle dung'.

[13] https://platform.openai.com/docs/api-reference

Table 2
Results of the evaluation task by Chat-GPT 4.

	Emojilingo	GPT 3.5	GPT 4
Preferences	116	110	139
Percentage	31,8%	30,1%	38,1%

4.3. Chat-GPT 4 as evaluation agent

To evaluate Chat-GPT 4's ability to suggest the best translation solutions we organised an evaluation task run by the model itself using the human crowdsourced translation, the Chat-GPT 3 and the Chat-GPT 4 ones.

Also, in this case we adopted a zero-shot prompting approach. The original Italian prompt is translated into English as follows:

> I would like to ask you to evaluate 3 translations from archaic Italian words extracted from Dantes's Divina Commedia into emoji.
> I will provide you with:
> - The Italian word
> - Emoji translation A
> - Emoji translation B
> - Emoji translation C
>
> I ask you to tell me which translation into emoji do you prefer and why. Respond with 2 lines of plain text (without formatting) with the following info:
> - Choice: <emoji string>
> - Explanation: <Brief explanation of the choice>
>
> Here you are:
> - Italian word: term
> - Translation A: emoji1
> - Translation B: emoji2
> - Translation C: emoji3

To ensure more reliable results we instructed the model to perform 10 retries and select the most frequent answer. As the model was evaluating several translations ex aequo, we decided to reiterate the process until a difference was reached between the first and the second preferred translation. The results of this evaluation task are shown in table 2.

The data from the Chat-GPT 4 evaluation shows that Chat-GPT 4 has the highest preference score at 38.1%, followed closely by Emojiitaliano at 31.8%, and Chat-GPT 3.5 at 30.1%. This suggests that Chat-GPT 4 was generally rated more favourably compared to Chat-GPT 3 and Emojitaliano.

The distribution of proportions is essentially symmetrical and balanced. The currently preferred translations have been compiled into a corpus validated as Emojilingo.

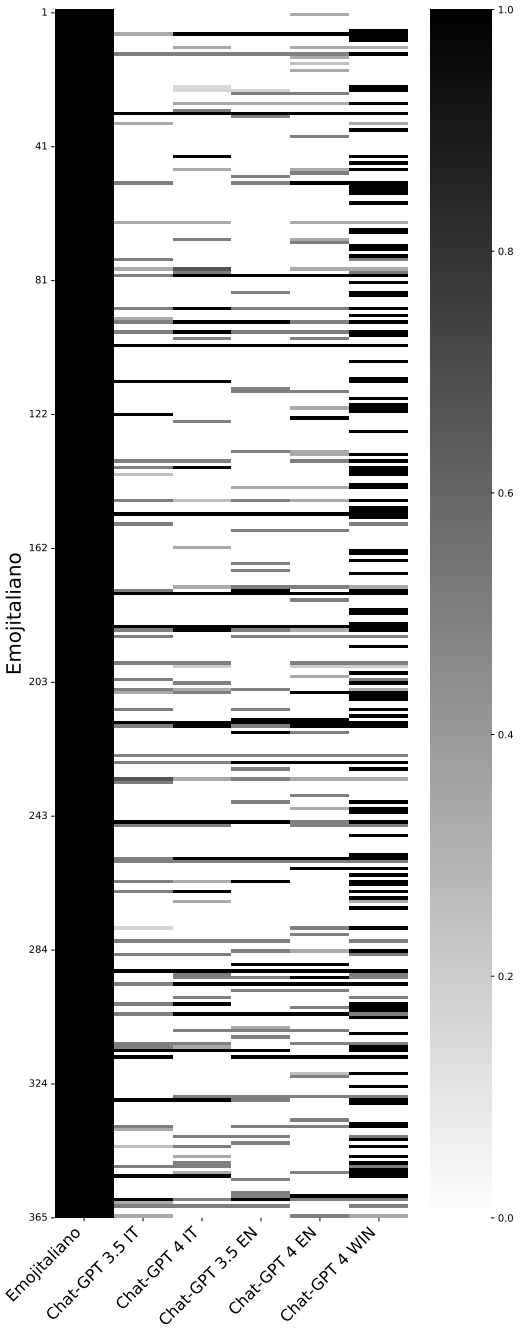

Figure 1: Similarity matrix between Emojitaliano and various versions of Chat-GPT engines.

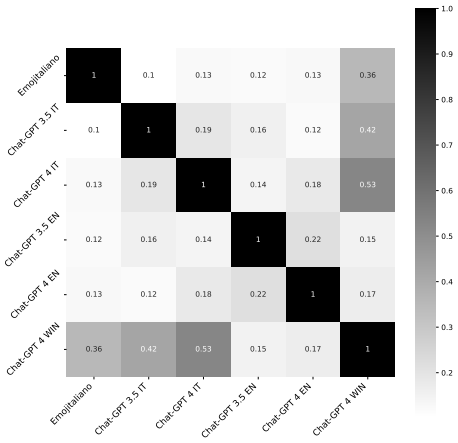

Figure 2: Similarity matrix between all models.

4.4. Similarity Matrix

In figure 1 we report the similarity matrix between all Emojitaliano 365 values and the corresponding values provided by all Chat-GPT engines (both for Italian and English).[14] In figure 2 we report the similarity matrix between all model pairs. In both figures, we include the selection of the Chat-GPT 4 evaluation agent presented in section 4.3, using the label *Chat-GPT 4 WIN*.

The similarity score between two strings, is computed using the *Levenshtein distance*: δ.[15] It is defined as the minimum number of single-character edits (insertions, deletions or substitutions) required to change one word into the other. It is then normalized (i.e., divided by the maximum length of either strings). Finally the similarity is obtained as $1 - \delta$. The similarity is 1 (black) when the two strings are identical, and 0 (white) when they have no emojis in common. In the heatmap of figure 2 the similarity between two models is computed as the mean between all the 365 term-pairs similarities.

5. Conclusion

In this paper we presented a translation experiment into emojis using two versions of Chat-GPT, to compare them with a human version, already available, realized within the framework of the Emojitaliano experience. The present project focuses on an integrated translation program, that combines both human (Emojitaliano) and automated approaches, as a basis for a constructed emoji-based pivot language: Emojilingo. Using a zero-shot prompting approach, both Chat-GPT versions (3.5 and 4) provided emoji translations for 365 words extracted from Dante's *Comedy*, along with explanations for the their own translation solutions. We also had Chat-GPT evaluate the three different translations produced within the Emojitaliano project, alongside those produced by Chat-GPT.

The present experiment substantially succeeds in confirming AI easiness and ability to use emojis to convey linguistic meanings, also managing special and archaic vocabulary. We in fact tested the machine's ability to handle denotative and connotative issues in the different translation choices, i.e. the translation solutions can be multi-faceted, each one catching some of the many semantic features underlying words. In this sense most solutions may be acceptable, such as to demonstrate the versatility of the emoji code to convey the senses.

Within this broad faculty of choice, however, some options seem quite critical, due to the dissimilarity of cultural values expressed by the languages, and by the emojis themselves. That is, a main consequence of using AI for translation, also in emojis, is the reaffirmation of the crucial challenge in international translation: the need for careful attention to specific cultural dimensions during localization [19]. Cultural values underlie texts, words and languages, as, for example, a 'pig' is an 'occidental' symbol for negative concepts as 'dirt' and 'gluttony' (as in 'lurco'), while the animal has a totemic or sacred value elsewhere; likewise, colors, or gestures, take on cultural values according to societies and cannot be accorded univocal international meanings. The choice of an icon as and international multilingual sign cannot override cultural peculiarities. Finally, cultural vocabularies may vary on the basis of literary contexts and textual genres, often conveying suggestions related to signifiers that are now lost. Given the conservative structure of poetical language, emoji translations may therefore need to move beyond the broadness of interlingua to fully convey meanings by reproducing linguistic signs 'verbatim' (es. 'intuarsi' 🧍🪄👆): that is, the literal solution, usually ruled out from the perspective of an international semantic code, becomes substantial to recover the cultural dimension of a literary text [20]. Special care is therefore required in selecting corresponding matches in emoji so that they do not conflict with reception in different countries and societies and so that they do succeed in reaching the core content of the original, which is the main purpose of 'the emojilingua'.

Future research will always need a human evaluation of automated outcomes, carried on by a team with extensive expertise in cross-cultural perspectives, and with a deep understanding of cultural values of emojis. This will help to limit unrestricted creativity and ensure a wide common comprehension of Emojilingo, and its highest exportability.

[14] For clarity we include also Emojitaliano in the the first column, which is all black as it is identical to the original values used as reference.

[15] https://en.wikipedia.org/wiki/Levenshtein_distance

Acknowledgement

This work has been funded by the European Union - NextGenerationEU under the Italian Ministry of University and Research (MUR) National Innovation Ecosystem grant ECS00000041 - VITALITY - CUP D83C22000710005. The views and opinions expressed in the paper are the authors' alone and remain entirely their responsibility.

References

[1] M. Danesi, The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet, Bloomsbury Publishing, London, New York, 2017.

[2] V. Evans, The emoji code : how smiley faces, love hearts and thumbs up are changing the way we communicate, Michael O'Mara Books Limited, London, 2017.

[3] H. Melville, Emoji Dick, Or, The Whale, Lulu.com, 2010. URL: https://www.emojidick.com.

[4] J. Hale, Wonderland, https://www.joehale.info/visual-poetry/wonderland.html, 2014. Accessed: 2024-10-17.

[5] F. de Saussure, Course in General Linguistics, Duckworth, London, [1916] 1983. (trans. Roy Harris).

[6] J. Monti, F. Chiusaroli, F. Sangati, Emojitaliano: A Social and Crowdsourcing Experiment of the Creation of a Visual International Language, in: M. M. Soares, E. Rosenzweig, A. Marcus (Eds.), Design, User Experience, and Usability: UX Research and Design, Springer International Publishing, Cham, 2021, pp. 426–441.

[7] F. Chiusaroli, J. Monti, F. Sangati, Pinocchio in Emojitaliano, Apice Libri, 2017.

[8] J. Monti, F. Sangati, F. Chiusaroli, M. Benjamin, S. Mansour, Emojitalianobot and emojiworldbot - new online tools and digital environments for translation into emoji, in: Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016), 2016.

[9] A. Srivastava, A. Rastogi, A. Rao, A. A. M. Shoeb, A. Abid, A. Fisch, ..., Z. Wu, Beyond the imitation game: Quantifying and extrapolating the capabilities of language models, 2023. URL: https://arxiv.org/abs/2206.04615. arXiv:2206.04615.

[10] L. Peng, Z. Wang, H. Liu, Z. Wang, J. Shang, Emojilm: Modeling the new emoji language, 2023. URL: https://arxiv.org/abs/2311.01751. arXiv:2311.01751.

[11] L. H. Klein, R. Aydin, R. West, Emojinize: Enriching any text with emoji translations, 2024. URL: https://arxiv.org/abs/2403.03857. arXiv:2403.03857.

[12] F. Chiusaroli, La scrittura in emoji tra dizionario e traduzione, in: Proceedings of the Second Italian Conference on Computational Linguistics, CLiC-it 2015, 2015. URL: https://api.semanticscholar.org/CorpusID:191807868.

[13] F. Chiusaroli, Da emojipedia a pinocchio in emojitaliano: l'"emojilingua" tra scritture e riscritture, in: Homo Scribens 2.0. Scritture ibride della modernità, S. Lubello (Ed.), 2019.

[14] F. Chiusaroli, T. Uricchio, J. Monti, M. L. Pierucci, F. Sangati, GPT-based Language Models meet Emojitaliano: A Preliminary Assessment Test between Automation and Creativity, in: Proceedings of the Ninth Italian Conference on Computational Linguistics, Venice, 2023.

[15] OpenAI, J. Achiam, S. Adler, S. Agarwal, L. Ahmad, ..., B. Zoph, Gpt-4 technical report, 2024. URL: https://arxiv.org/abs/2303.08774. arXiv:2303.08774.

[16] R. Willerton, Plain Language and Ethical Action: A Dialogic Approach to Technical Content in the 21st Century, ISSN, Taylor & Francis, 2015. URL: https://books.google.co.jp/books?id=9lWsCQAAQBAJ.

[17] F. Chiusaroli, Emoji e semplificazione linguistica, in: Comunicare il patrimonio culturale. Accessibilità comunicativa, tecnologie e sostenibilità, FrancoAngeli, 2021, pp. 164–193.

[18] C. Bliss, Semantography (Blissymbolics): A Simple System of 100 Logical Pictorial Symbols, which Can be Operated and Read Like 1+2, A logical writing for an illogical world, Semantography (Blissymbolics) Publications, 1965.

[19] J. Monti, F. Chiusaroli, Il codice emoji da oriente a occidente : standard unicode e dinamiche di internazionalizzazione, in: RILA : Rassegna Italiana di Linguistica Applicata, Bulzoni, Roma, 2017, pp. 83–101. URL: http://digital.casalini.it/10.1400/256182.

[20] F. Chiusaroli, Alla ricerca della traduzione universale. L'infinito leopardiano tra lingue artificiali e codici non verbali, Macerata, eum, 2022, pp. 73–100.

A. Emojitaliano Works

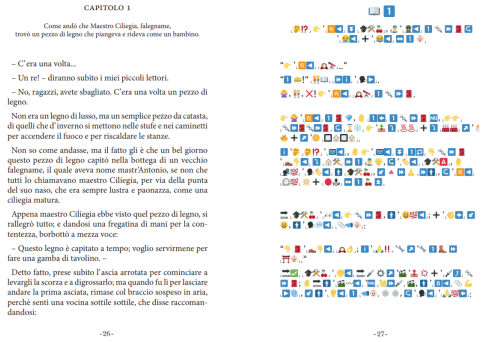

Figure A.1: "Pinocchio" in Emojitaliano

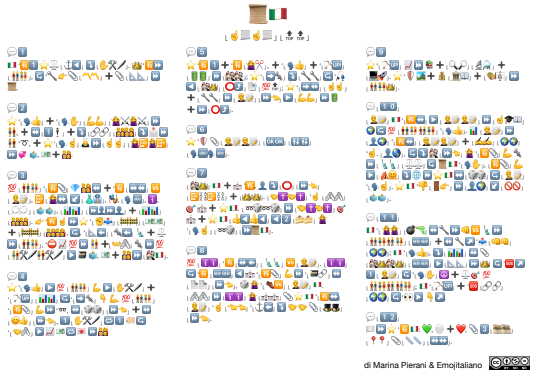

Figure A.2: "Italian Constitution" in Emojitaliano

Figure A.3: "Divina Commedia" in Emojitaliano

Figure A.4: L'"Infinito" in Emojitaliano

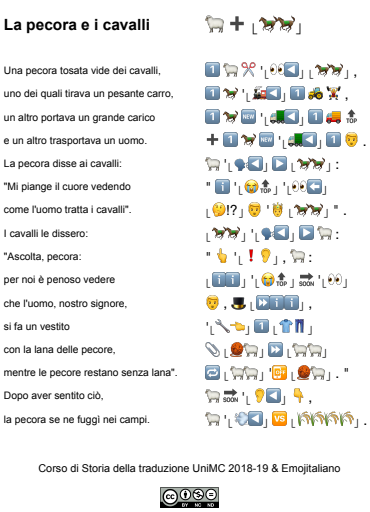

Figure A.5: "La pecora e i cavalli" in Emojitaliano

Figure A.6: “Manifesto della comunicazione non ostile” in Emojitaliano

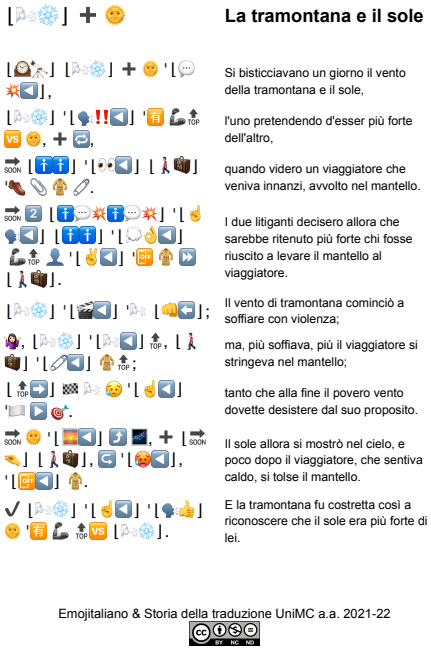

Figure A.7: “La tramontana e il sole” in Emojitaliano