Emojitalianobot and EmojiWorldBot
New online tools and digital environments for translation into emoji

Johanna Monti
L’Orientale University
Naples, Italy
jomonti@unior.it

Federico Sangati
Independent Researcher
The Netherlands
federico.sangati@gmail.com

Francesca Chiusaroli
University of Macerata
Italy
f.chiusaroli@unimc.it

Martin Benjamin
EPFL
Lausanne, Switzerland
martin@kamusproject.org

Sina Mansour
EPFL
Lausanne, Switzerland
mansour@ee.sharif.edu

Abstract

English. Emojitalianobot and EmojiWorldBot are two new online tools and digital environments for translation into emoji on Telegram, the popular instant messaging platform. Emojitalianobot is the first open and free Emoji-Italian and Emoji-English translation bot based on Unicode descriptions. The bot was designed to support the translation of Pinocchio into emoji carried out by the followers of the "Scritture brevi" blog on Twitter and contains a glossary with all the uses of emojis in the translation of the famous Italian novel. EmojiWorldBot, an off-spring project of Emojitalianobot, is a multilingual dictionary that uses Emoji as a pivot language from dozens of different languages. Currently the emoji-word and word-emoji functions are available for 72 languages imported from the Unicode tables and provide users with an easy search capability to map words in each of these languages to emojis, and vice versa. This paper presents the projects, the background and the main characteristics of these applications.

Italiano. Emojitalianobot e EmojiWorldBot sono due applicazioni online per la traduzione in e da emoji su Telegram, la popolare piattaforma di messaggistica istantanea. Emojitalianobot è il primo bot aperto e gratuito di traduzione che contiene i dizionari Emoji-Italiano ed Emoji-Inglese basati sulle descrizioni Unicode. Il bot è stato ideato per coadiuvare la traduzione di Pinocchio in emoji su Twitter da parte dei follower del blog Scritture brevi e contiene pertanto anche il glossario con tutti gli usi degli emoji nella traduzione del celebre romanzo per ragazzi. EmojiWorldBot, epigono di Emojitalianobot, è un dizionario multilingue che usa gli emoji come lingua pivot tra dozzine di lingue differenti. Attualmente le funzioni emoji-parola e parola-emoji sono disponibili per 72 lingue importate dalle tabelle Unicode e forniscono agli utenti delle semplici funzioni di ricerca per trovare le corrispondenze in emoji delle parole e viceversa per ciascuna di queste lingue. Questo contributo presenta i progetti, il background e le principali caratteristiche di queste applicazioni.

1 Introduction

Emojitalianobot\(^1\) and EmojiWorldBot\(^2\) are two new translation bots\(^3\) into and from emoji. These two bots were designed starting from the hypothesis of setting up an emoji multilingual dictionary and translator through a process of selection and assessment of conventional semantic values. Translation cases may show how images can convey common and universal meanings, beyond specific peculiarities, so as they can stand as models in the perspective of an interlanguage (Chiusaroli, 2015). The two

\(^1\)https://telegram.me/emojitalianobot/

\(^2\)https://telegram.me/emojworldbot

\(^3\)Computer programmes that carry out repetitive tasks and in their more sophisticated form can also simulate human behaviours.
bots ease the use of emojis but also collect, refine and make available valuable linguistic data by means of crowdsourcing and gamification approaches.

This contribution presents the state-of-the-art concerning the use of crowdsourcing and gamification approaches to linguistics in section 2, the Emojitalianobot and the Pinocchio project in section 3, the EmojiWorldBot in section 4 and finally conclusions and future work in section 5.

2 Crowdsourcing and gamification

Crowdsourcing, i.e., the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call (Howe, 2006) is becoming a widespread practice on the Internet to develop linguistic resources (dictionaries, glossaries, translation memories, etc.) or services (translation, localisation, fansubbing, etc.) (Monti, 2012, 2014). It allows the large scale involvement of users who contribute with their knowledge, their ideas, and their skills, in this way performing an active role in the achievement of a common goal. Crowdsourcing can be used for the creation, maintenance and sharing of lexical/terminological data such as: i. lexical resources for online dictionaries, e.g., Wiki platforms such as Wiktionary and Omegawiki, and recent forays by more traditional dictionary publishing companies like Collins, Oxford, and Macmillan; ii. terminological resources for online terminological databases, like TermWiki, the terminological counterpart of Wiktionary or Taas; iii. lexical and semantic resources for Natural Language Processing (NLP) tasks, such as Word Sense Disambiguation (WSA), Sentiment Analysis, Computer Aided Translation, Machine Translation and so on, using platforms for distributing parts of large development projects to professional or occasional lexicographers such as Mechanical Turk. To the best of our knowledge only very few projects so far have been tailored to mobile devices to gather linguistic data in the field, (i) to collect dialect data as in Dialectbot, (ii) to document endangered languages as in Aikuma and Ma Iwaidja, or (ii) to gather grammaticality judgments (Madnani et al., 2011). The social dimension of these types of activities is sometimes connected and fed by social communities, where users discuss problems, give suggestions, and exchange ideas (Brabham, 2012; McGonigal, 2011). In order to loyalize social communities and improve their engagement, gamification is used very often. The use of games is a very effective tool for active participation since it provides a strong motivational framework which pushes people to act for good. Some effective uses of games are to create new habits or modify wrong actions. Wang et al. (2013) list Games with a purpose (GWAPs) among the different types of crowdsourcing. Some good examples of games with a purpose in the lexicographic field are Phrase Detectives and JeuxDeMots. The main advantage of GWAPs is their high attractiveness, because people love playing games and it is easier to obtain their contribution in this way in comparison to other forms of crowdsourcing. The difficulty in designing such games is to match attractiveness with usefulness, i.e. an attractive game which produces valuable data.

3 Emojitalianobot and the Pinocchio project

Emojitalianobot is the first open and free Emoji-Italian translation bot on Telegram. It was developed to support the translation project of Pinocchio in emoji launched on Twitter in February 2016 by F. Chiusaroli, J. Monti and F. Sangati. The translation of the famous children’s novel was carried out by the followers of the Scritture brevi blog (by F. Chiusaroli and F.M. Zanzotto) and the first fifteen chapters have been translated, which correspond to the original novel published by Collodi in 1881. Every day tweets with sentences taken from the novel were posted on Twitter and the followers suggested their translations in emoji.

10http://www.aikuma.org/aikuma-app.html
12When a player without any special knowledge is put into a gaming environment and has to make decisions to win the game under the pressure of time or any game mechanics’ constraints.
13https://anawiki.essex.ac.uk/phrase detectives/
14http://www.jeuxdemots.org/jdm-accueil.php
15http://www.treccani.it/lingua_italiana/speciali/ludolinguistica/Chiusaroli.html
16https://www.scritturebrevi.it/
in emoji; at the end of each day, the official version of the translations was validated and published. Translators used Emojitalianobot that contains (i) the Emoji-Italian dictionary, (ii) the Emoji-English descriptions based on Unicode and (iii) a glossary with all the uses of emoji in the translation of Pinocchio. The project was associated with the Emojitalia discussion group on Telegram, where users met to discuss problems, solutions, suggest improvements of the bot, in addition to the translation choices for Pinocchio and communicate in emoji. The Pinocchio translation project therefore allowed to crowdsource different linguistic data connected with the use of emojis as actual means of communication and not just simple graphics to express amusement or interest. In this respect the main findings of the project are twofold: the need to recur to compound multi-emoji expressions in order to express concepts which are not represented in the current set, as well as a related simple grammar to express syntactic relations among emojis, past and future tenses, etc. Unlike previous literary translation project in emojis, such as the translations of Moby Dick or Alice in Wonderland, this is the first attempt of a collective shared emoji code (vocabulary and grammar) based on a word for word translation totally in emojis. Emojitalianobot is an ideal test bench to experiment with new approaches like crowdsourcing and gamification in the field of Natural Language Processing (NLP). The Pinocchio project, games and features available in the bot to learn or guess the meaning of emoji are devised indeed both to enjoy the bot while using it and at the same time to give the opportunity to users to develop linguistic descriptions of emoji tailored on their actual perceptions. The most important reward for playing with the bot is the awareness of helping develop a linguistic resource for one’s mother tongue, and the pride in contributing to it.

Since its release on Telegram, the project was an instant success, becoming a viral web phenomenon thanks to the Scritture brevi community and the Pinocchio translation in emojis, so that the bot has now almost 750 users. The Pinocchio translation project in emojis counts 611 tweets, 980 glossary entries which correspond to 2127 words, of which 185 are multi-emojis, i.e. compound emojis, such as 🍳 🍽️ for the Italian word peggio (worst).

4 EmojiWorldBot

On the basis of (both linguistic and technological) experience with Emojitalianobot, the three Italian researchers together with Martin Benjamin and Sina Mansour of the Kamusi Project International and EPFL (Switzerland) designed a new bot on Telegram in April 2016: EmojiWorldBot, a multilingual dictionary that uses Emoji as a pivot language from dozens of different languages. Currently the emoji-word and word-emoji functions are available for 70 languages imported from the Unicode tables and provide users with an easy search capability to map words in each of these languages to emojis, and vice versa. Looking at the UNICODE descriptions (see Fig. 1) it is apparent that emojis are not annotated in a coherent way across languages, so some languages have more descriptions and some others, especially underrepresented languages, have less or in the most cases some languages are not represented at all.

Figure 1: Annotations in Romance languages

Our first goal with EmojiWorldBot is therefore to reach a uniform and comprehensive list of tags across multiple languages with a precise mapping between any language pair, which may serve to bootstrap a massive multilingual dictionary. The bot currently features:

- emoji-to-word and word-to-emoji translation for more than 70 languages
- Eggs, a tagging game for people to contribute to the expansion of these dictionaries or the creation of new ones for any additional language. Users can suggest additional tags for single emojis in any language (for example adding egg to the tag list for 🍳 in English).

---

17The translation of Pinocchio in emoji can be followed on Twitter using #emojiitaliano.

18https://kamusi.org/
19http://www.unicode.org/cldr/charts/29/annotations/
• inline queries: type EmojiWorldBot and a word, and it will suggest a set of emojis for that word you can send in any Telegram conversation
• the possibility to add new languages. To date 56 new languages were added, such as Latin, Esperanto, Sardinian among others.

The basic idea of the Eggs game is to collect new tags to associate with emojis as shown in Fig. 2.

With fewer than 2000 official emojis, stretching the boundaries of their communicative potential makes them more useful. However, it also makes the dictionary more essential, so that someone who receives 🐔 in a chat in any language might look to see if it signifies something other than an eggplant. In the future, Eggs will experiment with multi-emoji terms (METs), building on the work of the Pinocchio translation project to Emoji, in an effort to build a larger pictorial vocabulary that is comprehensible across languages (Chiusaroli, 2015). A new version of the bot is already under development. It will feature Ducks, a second game where users are asked to map tags from a source language (e.g. English) to a target language (e.g. Swahili). In the example of Figure 1, several Romanian users would be shown the sense-specific definition of grin from Wordnet and all of the emojis that have been attached to that definition, and be asked which of the options among fată încântată fetă and încântare, if any, is a good translation. The game would also be played for face and grinning face.

5 Conclusions
We described the EmojiItalianobot and the EmojiWorldBot projects. Combining crowd-
sourcing, gamification and a smartphone app is a powerful strategy to collect, improve and refine valuable linguistic data easily and in a short time particularly for less-resourced languages (Benjamin and Radetzky, 2014). These may be the first crowdsourcing projects of this type to use bots for linguistic data collection and validation and are unique in their attempts at engaging participants for different languages.

References


Francesca Chiusaroli. 2015. La scrittura in emoji tra dizionario e traduzione. CLiC it, page 88.


