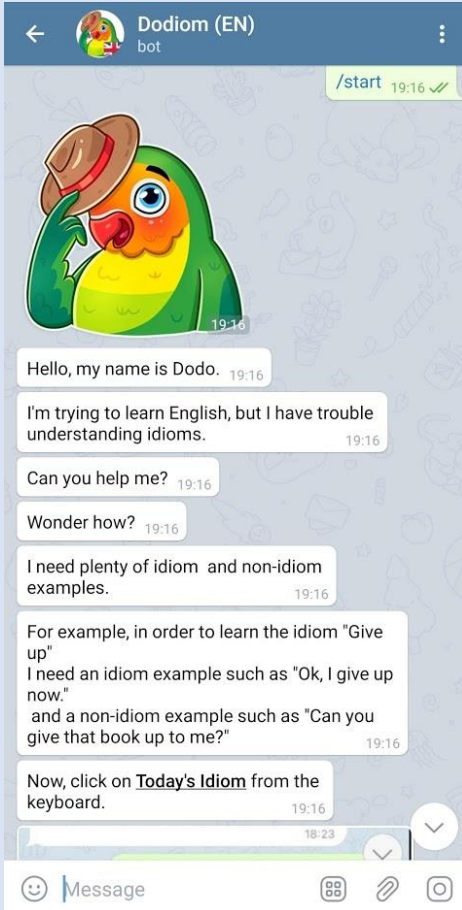


# Dodiom: A Gamified Bot Supporting Diversity and Multilinguality for Idiom Corpora Construction



## Abstract

• Idiomatic expressions are one of the linguistic phenomena with few or rare occurrences within text. That is why traditional methods focusing on their annotations within actual text or their retrieval from existing resources suffer from the **data scarcity** problem.

• "**Gamified Crowdsourcing for Idiom Corpora Construction**" (Eryigit et al., 2022) has been published and reported to provide a solution to this problem by collecting the data via a **crowd-creating & crowd-rating approach**.

• The approach was implemented as a **Telegram bot** and tested on Italian, Turkish, and Russian so far.

• The tool, being easily adaptable to new languages and new phenomena, is seen as a good means of supporting **inter- and intra-language diversity**.

## Key Points

- **First time in the literature:** a gamified approach that uses crowd-creating and crowd-rating techniques to speed up idiom corpora construction for different languages.
- The long-term goal is to **extend and play the game for languages other than the ones in this article**, especially for languages with few lexical resources.

