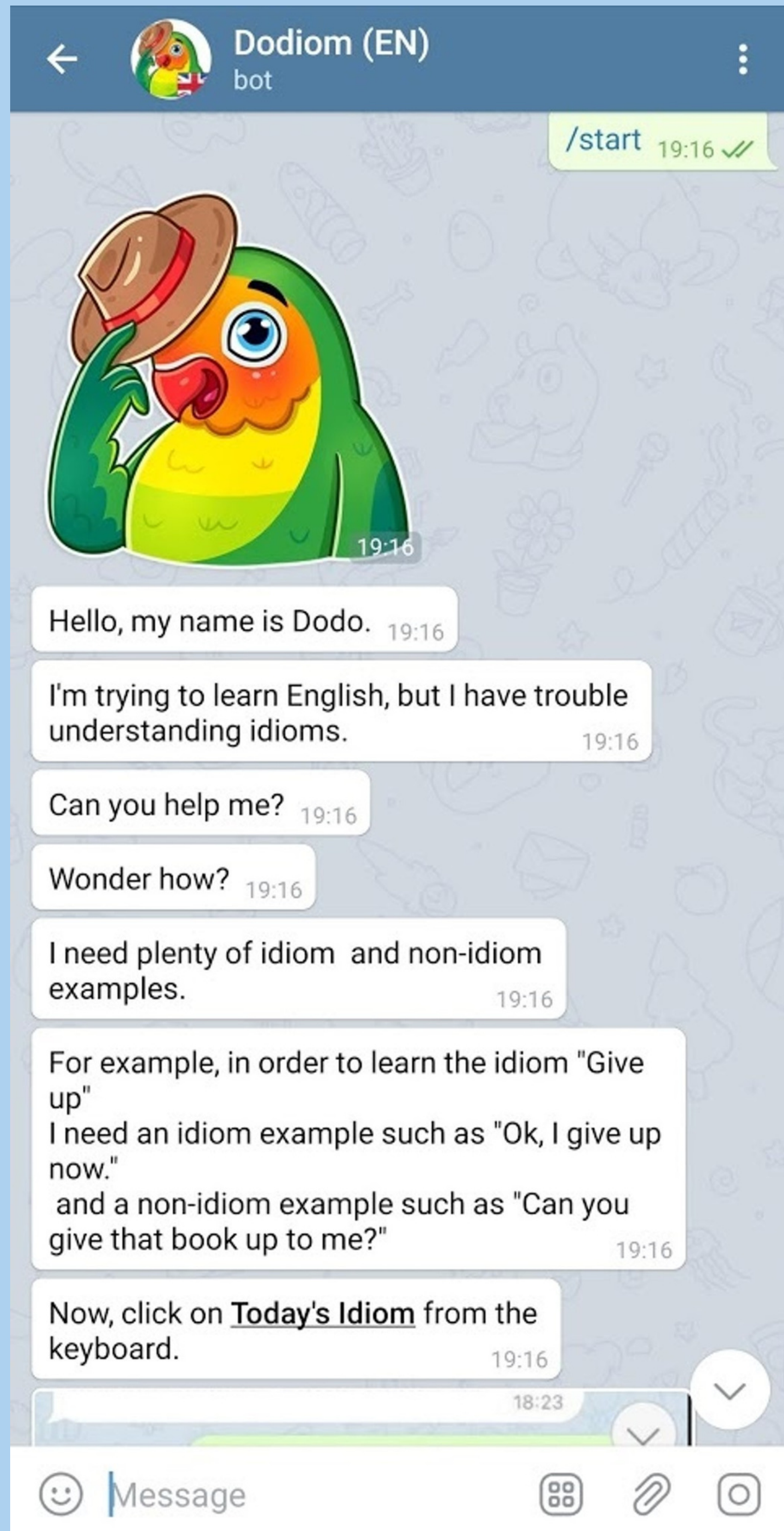


Dodiom: A Gamified Bot Supporting Diversity and Multilinguality for Idiom Corpora Construction

Gülşen Eryiğit¹Ali Şentaş¹Johanna Monti²

1. ITU NLP - Department of Artificial Intelligence & Data Engineering, Istanbul Technical University

2. UNIOR NLP - Department of Literary, Linguistic and Comparative Studies, University of Naples L'Orientale



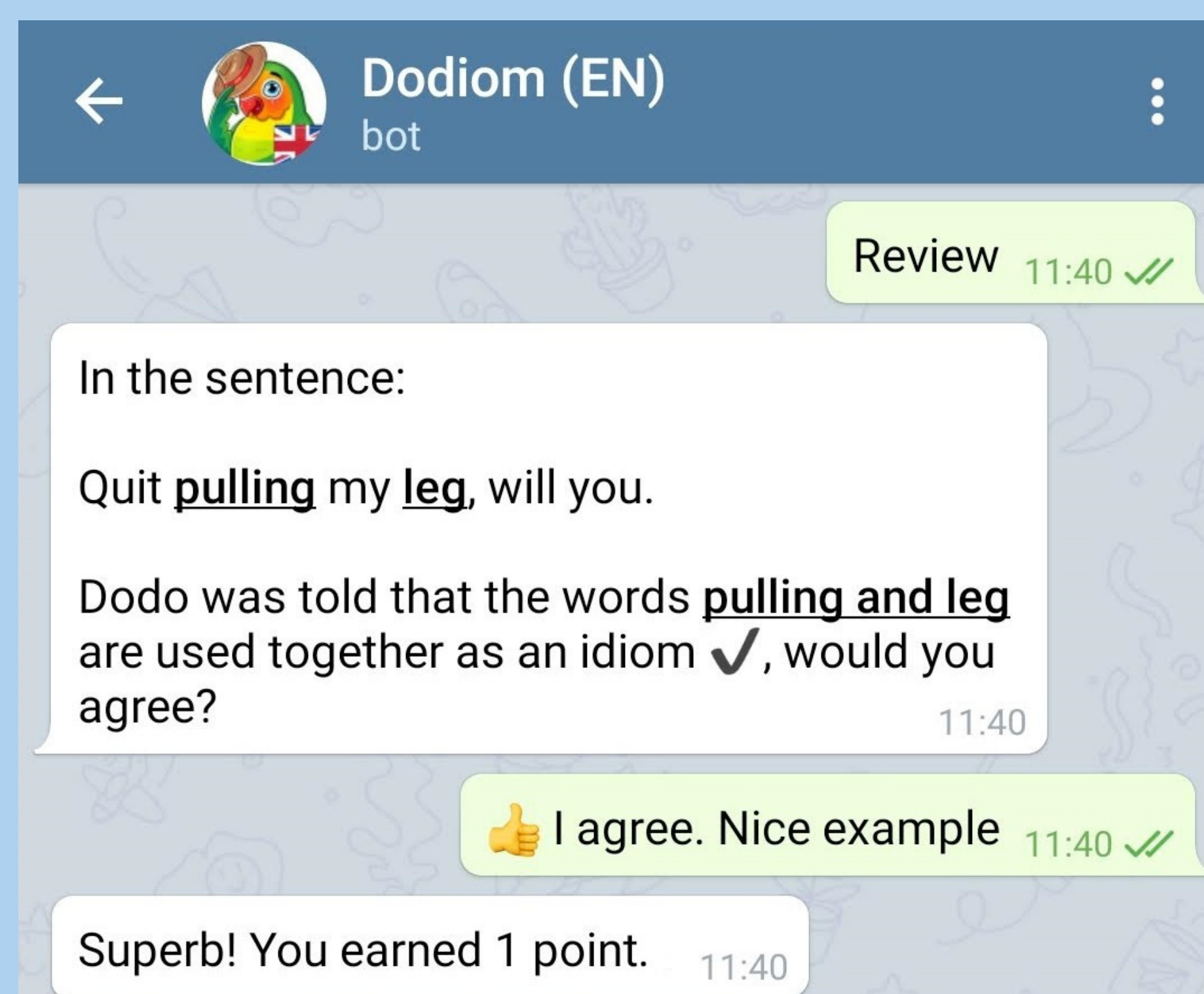
Abstract

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**.



What is it?

- An enjoyable and cooperative environment motivating the volunteers to help the research studies
- An asynchronous multiplayer game for native speakers who compete with each other while providing idiomatic and nonidiomatic usage examples and rating other players' entries
- A Telegram Bot
- An explicit crowdsourcing game
- Collects usage samples for idioms of which the words of the idiom may also commonly be used in their literal meanings within a sentence.
- An open-source project available via Github
- Easily adaptable to new languages via localization files
- Available localizations for English, Turkish, Italian, Russian.
- Designed according to Morschheuser et al. (2018) design principles

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.

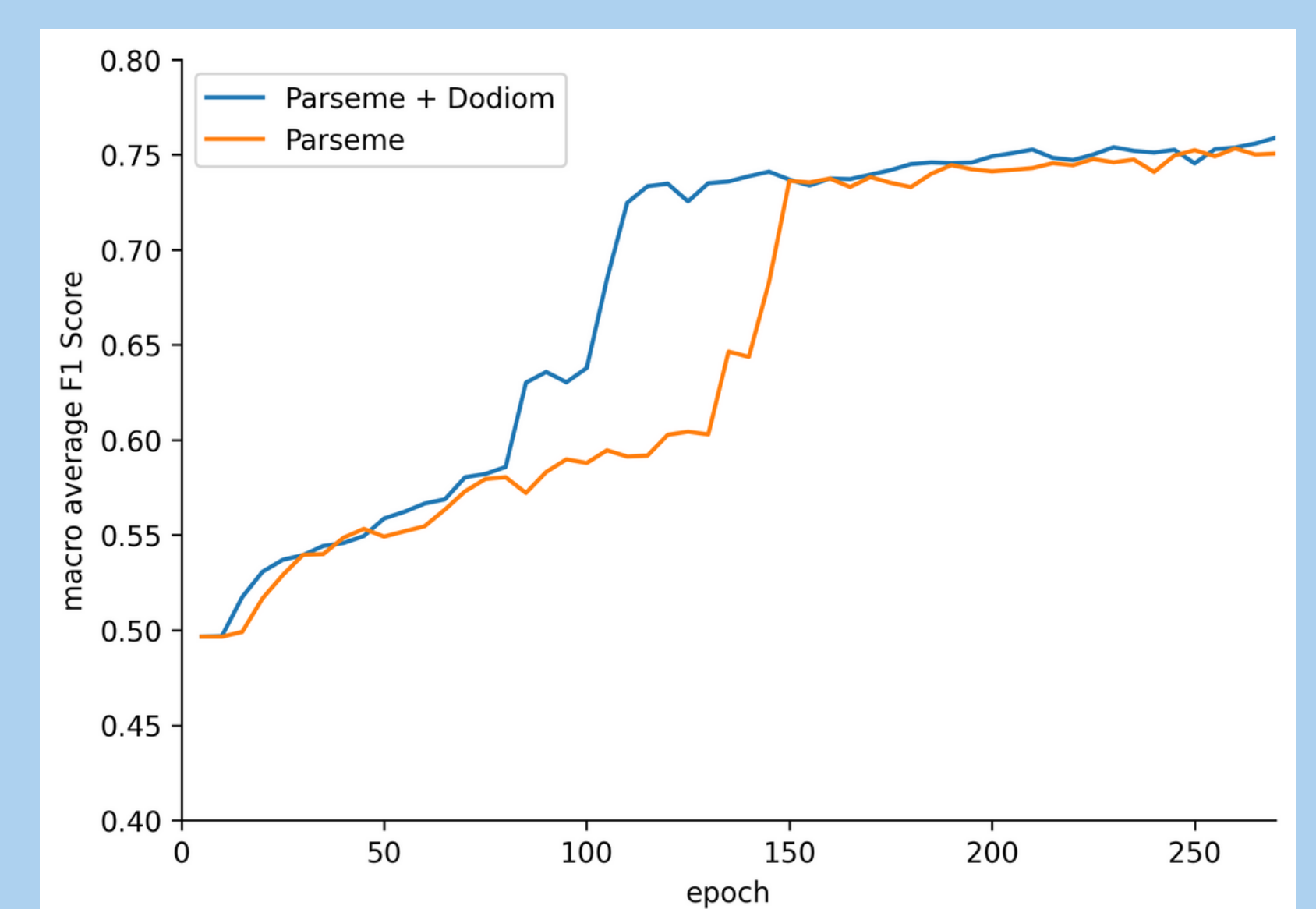
Experiments

Measured motivational and behavioral outcomes:

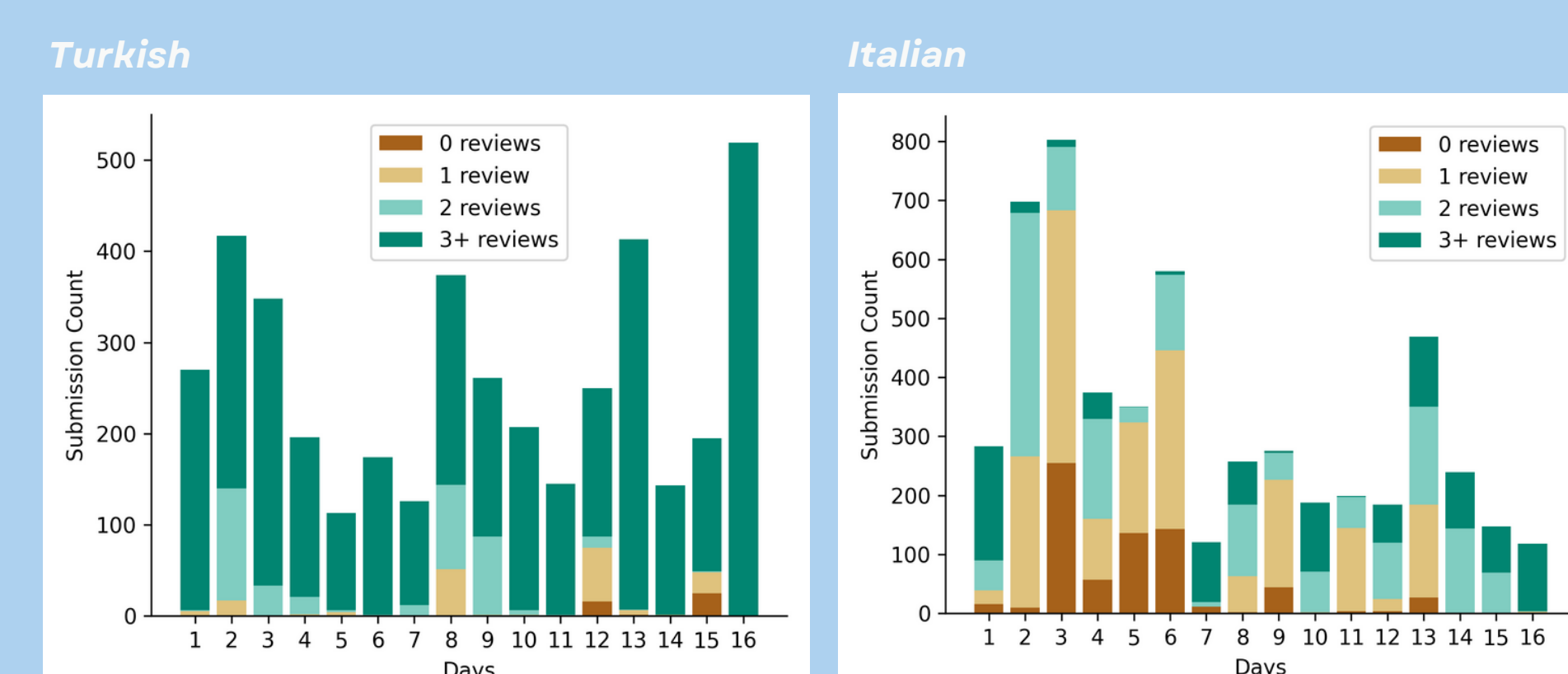
(Morschheuser et al. 2017, 2019)

- System usage,
- Engagement,
- Loyalty,
- Ease of use,
- Enjoyment,
- Attitude,
- Motivation, and willingness to recommend

Impact of Augmenting the Parseme Data Set (Savary et al. 2018) with the Dodiom Data Set on Idiom Identification Performances



Daily Review Frequencies Per Submission



References:

- Savary, A., et al. (2018). PARSEME multilingual corpus of verbal multiword expressions.
- Morschheuser B., Hamari J., Koivisto J. and Maedche A. (2017). Gamified crowdsourcing: Conceptualization, literature review, and future agenda. International Journal of Human-Computer Studies 106, 26–43.
- Morschheuser B., Hassan L., Werder K. and Hamari J. (2018). How to design gamification? a method for engineering gamified software. Information and Software Technology 95, 219–237.
- Morschheuser B., Hamari J. and Maedche A. (2019). Cooperation or competition - when do people contribute more? a field experiment on gamification of crowdsourcing. International Journal of Human-Computer Studies 127, 7–24.

Codes & Application:

- Github: <https://github.com/Dodiom/dodiom>
- Scan for the Telegram bot

Main Article:

- Eryiğit, G., Şentaş, A., & Monti, J. (2022). **Gamified crowdsourcing for idiom corpora construction**. Natural Language Engineering, 1-33.



SCAN ME



SCAN ME



NATURAL LANGUAGE
PROCESSING
RESEARCH GROUP