



This is a repository copy of *SciBabel: a system for crowd-sourced validation of automatic translations of scientific texts*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/185028/>

Version: Published Version

Article:

Soares, F. orcid.org/0000-0002-2837-1853, Rebechi, R. and Stevenson, M. orcid.org/0000-0002-9483-6006 (2020) SciBabel: a system for crowd-sourced validation of automatic translations of scientific texts. *Genomics and Informatics*, 18 (2). e21. ISSN 2234-0742

<https://doi.org/10.5808/gi.2020.18.2.e21>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

SciBabel: a system for crowd-sourced validation of automatic translations of scientific texts

Felipe Soares^{1*}, Rozane Rebechi², Mark Stevenson¹

¹Computer Science Department, University of Sheffield, Sheffield S38RA, UK

²Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre 91540-000, Brazil

Scientific research is mostly published in English, regardless of the researcher's nationality. However, this growing practice impairs or hinders the comprehension of professionals who depend on the results of these studies to provide adequate care for their patients. We suggest that machine translation (MT) can be used as a way of providing useful translation for biomedical articles, even though the translation itself may not be fluent. To tackle possible mistranslation that can harm a patient, we resort to crowd-sourced validation of translations. We developed a prototype of MT validation and edition, where users can vote for that translation as valid, or suggest modifications (i.e., post-editing the MT). A glossary match system is also included, aiming at terminology consistency.

Keywords: crowdsourcing, linguistics, machine translation, medical informatics applications, PubMed

Availability: Available online under the MIT license at <https://github.com/soares-f/scibabel>.

Introduction

Research in the biomedical domain, particularly about treatments and procedures for humans, can help improve the patient care offered by physicians. Evidence-based medicine is based on the premise that physicians give the best care possible when they base their treatments on reliable scientific evidence. But, although in practice this access is possible, there is a limitation that makes evidence-based medicine out of the reach for many physicians: almost all of its contents are written in English.

During the first half of the 20th century, scientific research was published in a variety of languages. But, as Gordin [1] described in detail, a complex set of factors led to English becoming the language of most scientific publications following the Second World War. Researchers tend to publish in English regardless of their native language. But, while academic researchers are often proficient in English, this may not be true for physicians in non-English speaking countries.

Translation of documents into the languages with which physicians are familiar seems like an obvious way to make the world's scientific production accessible to them. But new research is produced so quickly and its results are published so rapidly that translating the information manually would be impractical. For example, in 2019 alone, more than 10,000 new articles were published in PubMed (PubMed Query: (((“2019”[Date - Publication] : “3000”[Date - Publication])) AND (treatment[Title/Abstract])) AND (procedure[Title/Abstract])) containing the keywords “treatment” and “procedure”—exactly

the kind of articles that would be of interest to physicians. However, there is a technology that could potentially do this translation automatically: machine translation (MT).

MT is a technology to render texts written in one language to another language. Modern MT research began just after the Second World War with the automatic translation of Russian scientific texts to English [2] as part of the scientific response to the Cold War (e.g., see Hutchins[3]). Machine translation research fell into decline soon thereafter due to considerable skepticism about whether practical MT systems were possible within the research community [4], but MT resurged in the 1990s with the advent of more powerful computers and alternative approaches. The field of MT experienced explosive growth after the September 2001 terrorist attacks and is an active area of scientific research [5-8]. This effort has led to a substantial improvement in the quality of translations produced by MT systems [9].

The earliest work on MT for scientific content concentrated on the physical sciences, however the focus of current research is shifting towards biomedical texts, especially due to shared tasks. This difference is important because, while users of translations in other scientific fields can tolerate some amount of error, as they do not have such a strict vocabulary and are not dealing directly with human beings, even a small mistranslation in this domain (e.g., a drug name being incorrectly translated, or a negation being ignored) could lead to disastrous consequences to patients. For example, consider [Supplementary Table 1](#) which shows examples of a simple medical instruction (i.e., “Take two pills orally every day unless you feel dizzy or lightheaded”) usually found in drugs prescriptions translated into Finnish, Korean, Portuguese, Italian, Spanish, Japanese, French, German, Russian, Chinese (simplified) and Ukrainian by Google Translate. The third column contains their translations back into English by an educated native speaker (a common method of evaluating MT, similar to an approach

known as back-translation) [10]. Contraindications that have been incorrectly translated are highlighted in bold font and it can be seen that these occur in six of the 11 translations. This demonstrates the need for automatic translations to be manually checked for critical mistranslations. However, this process is time-consuming and unlikely to scale well. Therefore, we propose a crowdsourced approach to validate automatic translations of biomedical articles and develop a prototype to facilitate such task.

In the proposed system volunteers who are able to read biomedical articles in English and also in another language would check MT output for critical mistranslations and vocabulary adequacy. The purpose of this system is to guarantee that the message in the source text is correctly conveyed in the translation, even though the translated text may lack fluency. Volunteers would accept the proposed translations if they are correct and be able to make editions when appropriate (e.g., incorrect terminology). We expect that our system, named SciBabel, would allow physicians and medical staff not proficient in English to access the most recent advances in medicine, enabling them to provide their patients with better treatment. The source code is available at <https://github.com/soares-f/scibabel>.

Background

An illustration of the recent improvements in MT can be seen from the performance of systems reported in the biomedical track of the Conference on Machine Translation (WMT), which focuses on the translation of PubMed abstracts. Translation quality increased by around 51% (or 16 percentage points) from 2016 to 2019 for English to Spanish. In [Table 1](#) [11-16], we show the MT performance for some language pairs for biomedical texts with dates ranging from 2013 to 2019 for selected language pairs. Note that translation quality is measured automatically using the BLEU

Table 1. Machine translation performance in biomedical article abstract translation and Cochrane reviews

Reference	Language pair	Score (%)
Neveol et al. (2013) [11]	English → French Cochrane Reviews	BLEU: 40
Neves et al. (2016) [12]	English → Portuguese	BLEU: 33.37
	English → Spanish	BLEU: 31.11
Soares and Becker (2018) [13]	English → Portuguese	BLEU: 48.51
	English → Spanish	BLEU: 37.93
Saunders et al. (2019) [14]	English → Spanish	BLEU: 48.93
Soares and Krallinger (2019) [15]	English → Portuguese	BLEU: 49.51
	English → Spanish	BLEU: 47.01
	English → German	BLEU: 35.26
Peng et al. (2019) [16]	English → French	BLEU: 38.29
	English → Chinese	BLEU: 37.09

For years 2018 and 2019, metrics refer to the WMT challenge of the respective years.

score, a common MT metric that relies on the overlapping portions of the generated translations and the manually translated text [17].

In the two most recent WMT conferences (2018 and 2019) interesting results were reported for the English/Portuguese and English/Spanish language pairs. For instance, for the English to Spanish, the number of MT-generated sentences judged by humans as better than human translations was larger than the number of human sentences judged better than MT ones.

When combining the number of times that the best MT was equally good or better than human translation for WMT19, we get an average of 73% of correct translations according to human judgment, with surprising 90% for EN/ES and 82.09% for ZH/EN. This strengthens our point that MT can indeed be used to aid dissemination of biomedical scientific content.

However, as shown in [Supplementary Table 1](#), MT systems can make critical mistakes when considering the usage of a medicine, for instance. It has been shown in literature that even human translation is prone to errors [18]. That is why the translation and localization industry usually has a two-step (or even more) process for translation. That is, at least one additional human is involved in checking the translation already carried out (also called proof-reading) [19].

Crowdsourcing of intensive tasks is not new in science. One example can be the Folding@Home initiative [20], which was popular in the first decade of the years 2000. This initiative consisted of crowdsourcing computational power from regular end-users (that signed to the initiative) to simulate protein folding, drug design, and molecular dynamics. Similarly, Seti@Home [21] tried to follow the same path to search for extraterrestrial life.

The crowdsourcing of manual annotation (or evaluation) was already explored by different authors [22,23]. For instance, the information retrieval (IR) shared tasks can be seen as the pioneers of human distributed annotation. Participants of IR shared tasks would blindly evaluate the participants' automatic predictions. Another example of distributed annotation is the Amazon Mechanical Turk, which pays users to manually annotate tasks. Some authors developed games [24-26] or mobile apps [27] to gather human annotation.

Regarding crowdsourcing of translations, Zaidan and Callison-Burch [28] state that collecting translations by crowdsourcing using non-professionals may lead to low-quality results. They propose the use of distance among translations and LM perplexity to score collected translations to discriminate between "good" and "bad" translations.

Ambati et al. [29] explored the challenges involved in crowdsourcing translation based on their experiments with Amazon Me-

chanical Turk. Their main findings regarding challenges are related to the large label space, that is, even though there is a finite number of possible translations for a single translations, there is a much larger space of acceptable sentences in the target space, but that may not be adequate or not style compliant. The second one is the small number of bilingual speakers for low-resourced languages. The third one is low quality, as most of the crowd-sourced translators are not professional linguists. Given this scenario, they proposed a framework based on phases to enhance the final quality of crowd-sourced translations. The first step of the translation is done by weak bilingual translators, translations which are revised by bilingual translators and the final step is done by monolinguals of the target language or bilinguals whose mother tongue is the target one. Considering the potential of crowd-sourced annotation, we aimed at developing a prototype of a system to enable the manual evaluation of automatic translations tailored to biomedical texts and post-edition. Our goal was to produce a simple yet usable interface to annotate translations as valid in the target language, while enabling users to make adjustments in the translation to correct possible mistakes.

Design

When idealizing such a tool, we envisioned not to provide perfect and fluent translations, since that would require a considerable effort from users. We are rather interested in finding gross and dangerous MT mistakes, the ones that could completely hinder the interpretation of the article. That is, we are interested in assuring that the translated text conveys the same original message, even though it may not sound completely fluent for a native speaker.

We can see as an example the sentence "*Nehmen Sie jeden Tag zwei Tabletten ein, es sei denn, Ihnen ist schwindelig oder benommen*" in German. The direct translation, as seen in [Supplementary Table 1](#), is "Take two pills every day by mouth unless you feel dizzy or lightheaded." This may not sound natural, but it conveys the message that the dosage is two pills with a daily frequency and the contra-indication is if the person feels dizzy or lightheaded.

Functionalities

The following functionalities were implemented:

- Parallel visualization of the original text and the machine translated version.
- A "voting" system that allows users to flag a particular translation as correct (similar to a "like" in social media).
- An option to edit a suggested translation, allowing users to correct possible mistranslations.
- Only the last translation is available, since this is deemed to be

the one with best quality.

- When editing a translation, a terminology lookup is available. That is, for each matched string in the source text, the suggested translation is shown.

Technical details

In our prototype we aim at providing a simple and easily upgradable interface for document validation and modification. The prototype is coded in Python 3 using the Flask microframework. Our choice of Flask is due to its simplicity regarding back-end and front-end, while being able to scale if required.

For the interface, we opted for the Bootstrap library (<https://getbootstrap.com/>), since it provides responsive mobile-ready frontend components. The functionalities were expanded using JQuery and Javascript.

As for the backend, we took advantage of the SQLAlchemy toolkit (<https://www.sqlalchemy.org/>), which is an ORM (Object Relational Mapper) that abstracts database operations. By using SQLAlchemy, we were able to make the app database agnostic. That is, the user can easily switch among the RDBMS supported by the package without needing to change several parts in the code.

Regarding the translation system behind the prototype, we used an in-house model developed with OpenNMT (<https://opennmt.net/>) which is decoupled from the interface. We do not think that at this point it is extremely relevant to have an online translation system, since new articles can be batch translated overnight, for instance.

For the dictionary, we encourage the usage of UMLS, since it is a very comprehensive asset, already standardized and is available in many languages. Users can also make use of SNOMED CT available in more than one language, when compatible with licensing.

Results

We implemented our prototype following the design specified in Section 3. For such, we first created a simple interface to visualize the translated content in the source language (e.g., English in this case) and target language (e.g., French). In this first screen, bilingual users can check the translation, which is shown in column format. We also introduced a feature that allows users to hover over the source or target sentence and check which sentence it refers to on the other column of the parallel text. After checking the translation, bilingual users can flag (i.e., Like) the translation as good, or perform modifications (editing).

In Fig. 1, we show a screenshot of the article validation step. We

have already included placeholders in the top bar to allow inclusion of alternative MT models as well as access to an Administrator backend which is under development.

In Fig. 2, we included a screenshot of the edition mode for the translated contents. In this view, the text is shown by sentences, with translations displayed as text boxes, such that users can perform post-edition on the suggested text. In addition, we included a glossary functionality, which can help users to guarantee terminology consistency. For this, a dictionary has to be supplied beforehand, and then a simple string matching is used to show the suggested translation. For instance, for the term “estrogen receptors”, the suggested translation in French is “Récepteur des œstrogènes”, while the automatic translation is “récepteurs aux œstrogènes”. Although the automatic translation is not wrong, the suggested term “Récepteur des œstrogènes” is flagged in UMLS (<https://www.nlm.nih.gov/research/umls/index.html>, Unified Medical Language System) as preferred.

Conclusion and Further Steps

In this article, we pointed out the importance of making biomedical literature accessible to all healthcare professionals, despite the language they speak. As scientific publication, especially in biomedical sciences, has been fastly growing, manual translation of articles is an untractable approach to make such information multilingual. Thus, we argue that MT can be an alternative to alleviate such bottleneck.

However, despite the increasing performance of MT systems, some critical errors may occur when texts are translated, which can ultimately hinder patient safety. Thus, manual validation/evaluation of translations should be performed to mitigate potential risks. To enable validation to scale to several languages, we point out that crowdsourcing the effort may be a solution. Therefore, we developed a prototype of a system that can allow an easy translation validation and possible edition.

The prototype was developed using Python 3 and Flask (<https://flask.palletsprojects.com/en/1.1.x/>), with Bootstrap for the visual interface. A visualization and edition interface was created, and an Administrator interface is currently under development. We included visual features to help users when doing the validation or editing the text.

As future steps, we envision some important upgrades:

- Ability to export translations into TMX and TXML formats, since they are standard in the localization industry;
- Ability to flag different unit of measurements in translation (e.g., pounds to kilograms), since the numbers need to be converted accordingly;

Pubmed Translation Evaluation Evaluation MT Models Admin

Cytoplasmic ER α and NF κ B Promote Cell Survival in Mouse Mammary Cancer Cell Lines

DOI: 10.1177/096032719701600807
 PubMed: https://www.ncbi.nlm.nih.gov/pubmed/9292286

English	Target Language	Actions
<p>There is a desperate need in the field for mouse mammary tumors and cell lines that faithfully mimic estrogen receptor (ER) expression and activity found in human breast cancers. We found that several mouse mammary cancer cell lines express ER but fail to demonstrate classical estrogen-driven proliferation or transcriptional activity. We investigated whether these cell lines may be used to model tamoxifen resistance by using small molecule inhibitors to signaling pathways known to contribute to resistance. We found that the combination of NFκB inhibition and ER antagonists significantly reduced cell proliferation in vitro, as well as growth of syngeneic tumors. Surprisingly, we found that ER was localized to the cytoplasm, regardless of any type of treatment. Based on this, we probed extra-nuclear functions of ER and found that co-inhibition of ER and NFκB led to an increase in oxidative stress and apoptosis. Together, these findings suggest that cytoplasmic ER and NFκB may play redundant roles in protecting mammary cancer cells from oxidative stress and cell death. Although this study has not identified a mouse model with classical ER activity, cytoplasmic ER has been described in a small subset of human breast</p>	<p>Il existe un besoin désespéré sur le terrain de tumeurs mammaires et de lignées cellulaires de souris qui imitent fidèlement l'expression et l'activité des récepteurs aux œstrogènes (ER) que l'on trouve dans les cancers du sein humain. Nous avons constaté que plusieurs lignées cellulaires de cancer mammaire de souris expriment l'ER mais ne parviennent pas à démontrer une prolifération ou une activité transcriptionnelle induite par les œstrogènes classiques. Nous avons étudié si ces lignées cellulaires peuvent être utilisées pour modéliser la résistance au tamoxifène en utilisant des inhibiteurs de petites molécules pour des voies de signalisation connues pour contribuer à la résistance. Nous avons constaté que la combinaison de l'inhibition de NFκB et des antagonistes ER réduisait de manière significative la prolifération cellulaire in vitro, ainsi que la croissance des tumeurs syngéniques. Étonnamment, nous avons constaté que l'ER était localisée dans le cytoplasme, quel que soit le type de traitement. Sur cette base, nous avons sondé les fonctions extra-nucléaires de l'ER et constaté que la co-inhibition de l'ER et du NFκB entraînait une augmentation du stress oxydatif et de l'apoptose.</p>	<p> 3 </p>

Fig. 1. Interface for translation evaluation. Users can flag the translation as adequate (i.e., Like) or edit the proposed translation using the links in the Actions column.

Pubmed Translation Evaluation Evaluation MT Models Admin

Cytoplasmic ER α and NF κ B Promote Cell Survival in Mouse Mammary Cancer Cell Lines

DOI: 10.1177/096032719701600807
 PubMed: https://www.ncbi.nlm.nih.gov/pubmed/9292286

English	Translation
<p>There is a desperate need in the field for mouse mammary tumors and cell lines that faithfully mimic estrogen receptor (ER) expression and activity found in human breast cancers.</p>	<p>Il existe un besoin désespéré sur le terrain de tumeurs mammaires et de lignées cellulaires de souris qui imitent fidèlement l'expression et l'activité des récepteurs aux œstrogènes (ER) que l'on trouve dans les cancers du sein humain.</p>
<i>mammary tumors</i>	<i>tumeurs mammaires</i>
<i>estrogen receptors</i>	<i>Récepteur des œstrogènes</i>
<i>breast cancer</i>	<i>cancers du sein</i>
<p>We found that several mouse mammary cancer cell lines express ER but fail to demonstrate classical estrogen-driven proliferation or transcriptional activity.</p>	<p>Nous avons constaté que plusieurs lignées cellulaires de cancer mammaire de souris expriment l'ER mais ne parviennent pas à démontrer une prolifération ou une activité transcriptionnelle induite par</p>

Fig. 2. Interface for translation editions Users can edit the proposed translation to make corrections on mistranslations or terminology adequacy. The prototype also shows suggested translations from terms matched in a dictionary, aiming at providing terminology consistency.

- Include a voting scheme for rollback of manual edits and a “annotation” weight according to the mother tongue of the annotator. In addition, a similar approach for quality assurance as proposed by [29] could be used, by establishing a score for annotators as well as for annotations;

- Develop an additional view to allow annotation transfer between source and target languages.

The last upgrade, related to annotation transfer, can be extremely helpful to create multilingual annotated datasets by leveraging existing annotations in one language. For instance, one could use annotations already made in a document in English to transfer those annotations to a translated text, making annotation quicker and less expensive.

ORCID

Felipe Soares: <https://orcid.org/0000-0002-2837-1853>

Rozane Rebechi: <https://orcid.org/0000-0002-1878-7548>

Mark Stevenson: <https://orcid.org/0000-0002-9483-6006>

Authors' Contribution

Conceptualization: FS, RR. Data curation: RR. Formal analysis: FS. Funding acquisition: MS. Methodology: FS. Writing – original draft: FS, RR, MS. Writing – review & editing: MS.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

The authors would like to acknowledge the University of Sheffield (Computer Science department) for the PhD scholarship of Felipe Soares and the ROIS-DS and DBCLS for the travel support provided.

Supplementary Materials

Supplementary data including one table can be found with this article online at <http://www.genominfo.org>.

References

- Gordin MD. *Scientific Babel: How Science Was Done before and after Global English*. Chicago: University of Chicago Press, 2015.
- Hutchins JW. Machine translation: a concise history. *Comput Aided Transl Theor Pract* 2007;13:11.
- Hutchins JW. *Early Years in Machine Translation: Memoirs and Biographies of Pioneers*. Amsterdam: John Benjamins Publishing, 2000.
- Poibeau T. The 1966 ALPAC report and its consequences. In: *Machine Translation (Poibeau T, ed.)*. Cambridge: MIT Press, 2017. pp. 75-89.
- Garg A, Agarwal M. Machine translation: a literature review. Preprint at <http://arxiv.org/abs/1901.01122> (2018).
- Okpor MD. Machine translation approaches: issues and challenges. *Int J Comput Sci Issues* 2014;11:159-165.
- Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. Preprint at <http://arxiv.org/abs/1409.0473> (2014).
- Cheng Y. *Joint Training for Neural Machine Translation*. Singapore: Springer Singapore, 2019. pp. 25-40.
- Apter E. *Translation-9/11: terrorism, immigration, and the world of global language politics*. *Global South* 2007;1:69-80.
- Rapp R. The back-translation score: automatic MT evaluation at the sentence level without reference translations. In: *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers (Su KY, Su J, Wiebe J, Li H, eds.)*, 2009 Aug 4, Singapore. Stroudsburg: Association for Computational Linguistics, 2009. pp. 133-136.
- Neveol A, Zweigenbaum P, Max A, Yvon F, Ivanishcheva Y, Ravaud P. Statistical machine translation of systematic reviews into French. *Training* 2013;15:366K.
- Neves M, Yepes AJ, Neveol A. The scielo corpus: a parallel corpus of scientific publications for biomedicine. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016 May, Portoroz, Slovenia. Paris: European Language Resources Association, 2016. pp. 2942-2948.
- Soares F, Becker K. *UFRGS participation on the WMT Biomedical Translation Shared Task*. Preprint at <http://arxiv.org/abs/1905.01855> (2018).
- Saunders D, Stahlberg F, Byrne B. *UCAM biomedical translation at WMT19: transfer learning multi-domain ensembles*. Preprint at <http://arxiv.org/abs/1906.05786> (2019).
- Soares F, Krallinger M. *BSC participation in the WMT translation of biomedical abstracts*. In: *Proceedings of the Fourth Conference on Machine Translation (Volume 3: Shared Task Papers, Day 2) (Bojar O, Chatterjee R, Federmann C, Fishel M, Graham Y, Haddow B, et al, eds.)*, 2019 Aug 1-2, Florence, Italy. Stroudsburg: Association for Computational Linguistics, 2019. pp. 175-178.
- Peng W, Liu J, Li L, Liu Q. Huawei's NMT systems for the WMT 2019 Biomedical Translation Task. In: *Proceedings of the Fourth*

- Conference on Machine Translation (Volume 3: Shared Task Papers, Day 2) (Bojar O, Chatterjee R, Federmann C, Fishel M, Graham Y, Haddow B, et al., eds.), 2019 Aug 1-2, Florence, Italy. Stroudsburg: Association for Computational Linguistics, 2019. pp. 164-168.
17. Papineni K, Roukos S, Ward T, Zhu WJ. BLEU: a method for automatic evaluation of machine translation. In: *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 2002 Jul 6-12, Philadelphia, PA*. Stroudsburg: Association for Computational Linguistics, 2002. pp. 311-318.
 18. Daems J, Macken L, Vandepitte S. Quality as the sum of its parts: a two-step approach for the identification of translation problems and translation quality assessment for HT and MT+PE. In: *Proceedings of MT Summit XIV Workshop on Post-Editing Technology and Practice* (O'Brien S, Simard M, Specia L, eds.), 2013 Sep 2, Nice, France. Allschwil: European Association for Machine Translation, 2013. pp. 63-71.
 19. Esselink B. *A Practical Guide to Localization: Language International World Directory*. Amsterdam: John Benjamins Publishing Company, 2000.
 20. Beberg AL, Ensign DL, Jayachandran G, Khaliq S, Pande VS. Folding@home: lessons from eight years of volunteer distributed computing. In: *2009 IEEE International Symposium on Parallel & Distributed Processing, 2009 May 23-29, Rome, Italy*. New York: Institute of Electrical and Electronics Engineers, 2009.
 21. Anderson DP, Cobb J, Korpela E, Lebofsky M, Werthimer D. SETI@home: an experiment in public-resource computing. *Commun ACM* 2002;45:56-61.
 22. Sabou M, Bontcheva K, Derczynski L, Scharl A. Corpus annotation through crowdsourcing: towards best practice guidelines. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), 2014 May, Reykjavik, Iceland*. Paris: European Language Resources Association, 2014. pp. 859-866.
 23. Bontcheva K, Derczynski L, Roberts I. Crowdsourcing named entity recognition and entity linking corpora. In: *Handbook of Linguistic Annotation* (Ide N, Pustejovsky J, eds.). Dordrecht: Springer Netherlands, 2017. pp. 449-464.
 24. Jurgens D, Navigli R. It's all fun and games until someone annotates: video games with a purpose for linguistic annotation. *Trans Assoc Comput Linguist* 2014;2:449-464.
 25. Rokicki M, Chelaru S, Zerr S, Siersdorfer S. Competitive game designs for improving the cost effectiveness of crowdsourcing. In: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM '14), 2014 Nov, Sanghai, China*. New York: Association for Computing Machinery, 2014. pp. 1469-1478.
 26. Munezero M, Kakkonen T, Sedano CI, Sutinen E, Montero CS. EmotionExpert: Facebook game for crowdsourcing annotations for emotion detection. In: *2013 IEEE International Games Innovation Conference (IGIC), 2013 Sep 23-25, Vancouver, BC, Canada*. New York: Institute of Electrical and Electronics Engineers, 2013.
 27. Chen N, Hoi SC, Li S, Xiao X. Mobile app tagging. In: *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM '16), 2016 Feb 22-25, San Francisco, CA, USA*. New York: Association for Computing Machinery, 2016. pp. 63-72.
 28. Zaidan OF, Callison-Burch C. Crowdsourcing translation: professional quality from non-professionals. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, 2011 Jun 19-24, Portland, OR, USA*. Stroudsburg: Association for Computational Linguistics, 2011. pp. 1220-1229.
 29. Ambati V, Vogel S, Carbonell J. Collaborative workflow for crowdsourcing translation. In: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, 2012 Feb 11-15, Seattle, WA, USA*. New York: Association for Computing Machinery, 2012. pp. 1191-1194.