Results of the Translation Inference Across Dictionaries 2021 Shared Task

Jorge Gracia[1], Besim Kabashi[2,3], and Ilan Kernerman[4]

[1] Aragon Institute of Engineering Research (I3A), University of Zaragoza, Spain
jogracia@unizar.es
[2] Friedrich-Alexander University of Erlangen-Nuremberg, Germany
[3] Ludwig Maximilian University of Munich, Germany
besim.kabashi@fau.de
[4] K Dictionaries, Tel Aviv, Israel
ilan@kdictionaries.com

Abstract. The objective of the Translation Inference Across Dictionaries (TIAD) shared task is to explore and compare methods and techniques that infer translations indirectly between language pairs, based on other bilingual/multilingual lexicographic resources. In this forth edition the participating systems were asked to generate new translations automatically among three languages - English, French, Portuguese - based on known indirect translations contained in the Apertium RDF graph. Such evaluation pairs have been the same during the three last TIAD editions. The main novelty this time has been the use of a larger graph as a basis to produce the translations, which is the Apertium RDF v2, and the introduction of improved evaluation metrics. The evaluation of the results was carried out by the organisers against manually compiled language pairs of K Dictionaries. For the first time in the TIAD series, some systems beat the proposed baselines. This paper gives an overall description of the shard task, the evaluation data and methodology, and the systems' results.

Keywords: TIAD · Apertium RDF · translation inference · lexicographic data

1 Introduction

A number of methods and techniques have been explored in the past aimed at automatically generating new bilingual and multilingual dictionaries based on existing ones. For instance, given a bilingual dictionary containing translations from one language L1 to another language L2, and another dictionary with translations from L2 to L3, a new set of translations from L1 to L3 is produced. The intermediate language (L2 in this example) is called pivot language, and it is possible to use multiple pivots for this purpose. When using intermediate languages, it is necessary to discriminate wrong inferred translations caused by translation ambiguities. The method proposed by Tanaka and Umemura [17] in 1994, called One Time Inverse Consultation (OTIC), identified incorrect translations when constructing bilingual dictionaries intermediated by a third language. This was a pioneering work and it still constitutes a baseline that is hard to beat, as the previous TIAD editions demonstrated. The OTIC method has been further adapted and evolved in the literature, for instance by Lim et al. [9], who grounded on it for their method for multilingual lexicon creation. From a different perspective, other works were proposed that relied on cycles and graph exploration to validate indirectly inferred translations, such as the SenseUniformPaths algorithm by Mousam et al. [10], the CQC algorithm by Flati et al. [2] or the exploration based on cycle density by Villegas et al. [18].

However, previous work on the topic of automatic bilingual/multilingual dictionary generation was usually conducted on different types of datasets and evaluated in different ways, applying various algorithms that are often not comparable. In this context, the objective of the Translation Inference Across Dictionaries (TIAD) shared task is to support a coherent experiment framework that enables reliable validation of results and solid comparison of the processes used. In addition, this initiative aims to enhance further research on the topic of inferring translations across languages.

The TIAD first edition[5] took place in Galway (Ireland) in 2017, co-located with the LDK'17 conference. The second edition[6] in 2019 was co-located with LDK'19 in Leipzig (Germany), and the third one[7] was planned at LREC'20 in Marseille (France) as part of the Globalex Workshop on Linked Lexicography[8]. Although the workshop of the third edition did not take place because of the COVID-19 crisis, the evaluation was run and the results published. Participants in this 3rd edition had the opportunity to present their systems jointly with the contributors to the 4th TIAD edition[9], during the whorkshop that took place in Zaragoza (Spain) at LDK'21. In this paper, we give an overall description of the shard task, the evaluation data and methodology, and the system results of TIAD 2021.

The remainder of this paper is organised as follows. In Section 2, an overall description of the shared task is given. Section 3 describes the evaluation data and Section 4 explains the evaluation process. In Section 5 the system results are reported, and conclusions are summarised in Section 6.

---

[5] `https://tiad2017.wordpress.com/`

[6] `https://tiad2019.unizar.es`

[7] `https://tiad2020.unizar.es`

[8] `https://globalex2020.globalex.link/globalex-workshop-lrec2020-about-globalex-lrec2020/`

[9] `https://tiad2021.unizar.es`

2 Shared task description

The objective of TIAD shared task is to explore and compare methods and techniques that infer translations indirectly between language pairs, based on other bilingual resources. Such techniques would help in auto-generating new bilingual and multilingual dictionaries based on existing ones.

In this fourth edition, the participating systems were asked to generate new translations automatically among three languages: English, French, and Portuguese, based on known translations contained in the Apertium RDF v2.0 graph[10]. As these languages (EN, FR, PT) are not directly connected in this graph, no translations can be obtained directly among them there. Based on the available RDF data, the participants had to apply their methodologies to derive translations, mediated by any other language in the graph, between the pairs EN/FR, FR/PT and PT/EN.

Participants could also make use of other freely available sources of background knowledge (e.g. lexical linked open data and parallel corpora) to improve performance, as long as no direct translation among the studied language pairs were available. Beyond performance, participants were encouraged to consider also the following issues in particular:

1. The role of the language family with respect to the newly generated pairs
2. The asymmetry of pairs, and how translation direction affects the results
3. The behavior of different parts of speech among different languages
4. The role that the number of pivots plays in the process

The evaluation of the results was carried out by the organisers against manually compiled pairs of K Dictionaries (KD), extracted from its Global Series[11], which were not accessible to the participants. A validation data set was made available to participants, upon request, in particular a 5% of randomly selected translations for each language pair. The goal of this validation data is to allow participants to analyse the nature of the data, to run some validation tests, and to analyse negative results.

3 Evaluation data

In this section we briefly describe the input data source that has been proposed in the shared task as a source of known translations, i.e., Apertium RDF, as well as the data used as golden standard, from KD.

3.1 Source data

As mentioned above, the shared task relies on known translations contained in Apertium RDF, which were used to infer new ones. Apertium RDF is the

[10] https://tiad2021.unizar.es/images/ApertiumRDFv2.0_graph.png
[11] https://www.lexicala.com/

linked data counterpart of the Apertium dictionary data. Apertium [3] is a free open-source machine translation platform. The system was initially created by Universitat d'Alacant and is released under the terms of the GNU General Public License. In its core, Apertium relies on a set of bilingual dictionaries, developed by a community of contributors, which covers more than 40 languages pairs.

Apertium RDF [7] is the result of publishing the Apertium bilingual dictionaries as linked data on the Web. The result groups the data of the (originally disparate) Apertium bilingual dictionaries in the same graph, interconnected through the common lexical entries of the monolingual lexicons that they share. An initial version of 22 language pairs was developed by Universidad Politécnica de Madrid and Universitat Pompeu Fabra[12]. A later conversion of the Apertium data into RDF, which we call Apertium RDF v2 in the following, was recently made by Goethe University Frankfurt and University of Zaragoza [6]. It contains 44 languages and 53 language pairs, with a total number of 1,540,996 translations between 1,750,917 lexical entries. In the second and third TIAD editions, the first version of Apertium RDF was used, while in this fourth edition we moved to the larger and richer Apertium RDF v2 graph.

In its first version, Apertium RDF was modeled using the *lemon* model [11] jointly with its translation module [15], while Apertium RDF v2 uses the Ontolex lemon core model to represent the data [12], jointly with the lemon vartrans module[13].

Each original Apertium bilingual dictionary was converted into three different objects in RDF: source lexicon, target lexicon, and translation set. As a result, two independent monolingual lexicons per dictionary were published as linked data on the Web, along with a set of translations that connects them. Note that the naming rule used to build the identifiers (URIs) of the lexical entries allows to reuse the same URI per lexical entry across all the dictionaries, thus explicitly connecting them. For instance the same URI is used for the English word *bench* as a noun: `http://linguistic.linkeddata.es/id/apertium/lexiconEN/bench-n-en` throughout the Apertium RDF graph, no matter if it comes from, e.g., the EN-ES dictionary or the CA-EN one. More details about the generation of Apertium RDF based on the Apertium data can be found at [7].

Figure 1 illustrates the Apertium RDF v2 unified graph. The nodes in the figure are the languages and the edges are the translation sets between them. All the datasets are available in Zenodo[14]. There is a plan to store the data in a permanent triplestore and expose it through a SPARQL endpoint in the near future, as part to the Prêt-à-LLOD project[15].

There were several ways in which the evaluation data was available to the participants: (i) through the data dumps available in Zenodo, which need to be loaded in a local triplestore, e.g., Apache Fuseki, and queried locally; (ii) through

[12] `http://linguistic.linkeddata.es/apertium/`
[13] `https://www.w3.org/2016/05/ontolex/#variation-translation-vartrans`
[14] `https://tinyurl.com/apertiumrdfv2`
[15] `https://pret-a-llod.eu/`

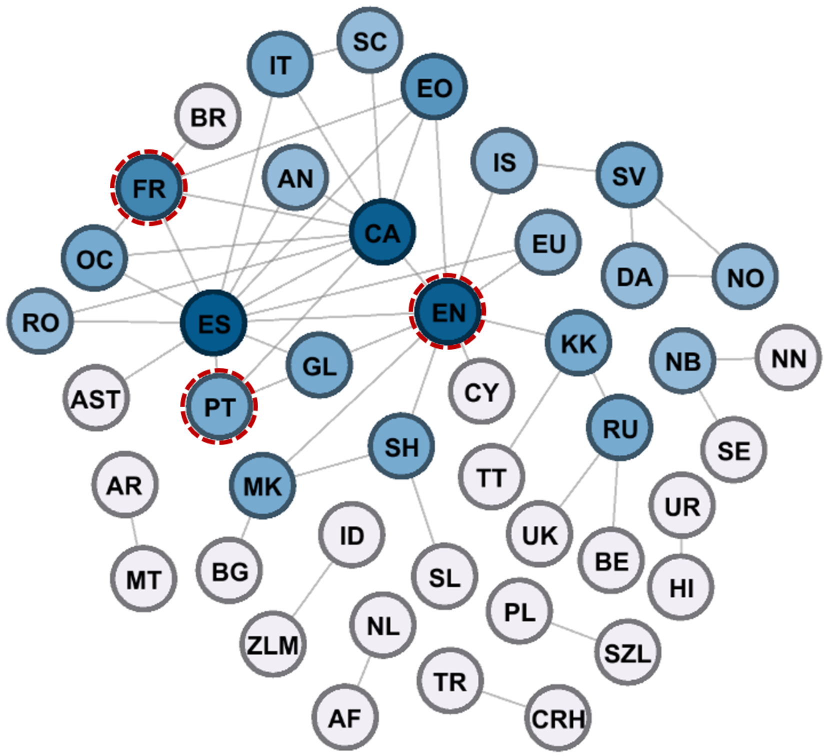

Fig. 1. The Apertium RDF v2 graph. The nodes in the figure represent the monolingual lexicons and the edges are the translation sets between them. The darker the colour, the more connections a node has. We have highlighted the three languages of this evaluation campaign: PT, FR, and EN.

a testing SPARQL endpoint[16], and (iii) in a ZIP file in comma separated values (CSV) format[17], for those not acquainted with semantic web technologies. More details on how to access the data are available in the TIAD 2021 website[18].

---

[16] Hosted by University of Frankfurt at `http://dbserver.acoli.cs.uni-frankfurt.de:5005/dataset.html`. The queries should be restricted to this graph: `http://linguistic.linkeddata.es/id/apertium-ud`. Since this is for testing purposes, there is no guarantee of a quick and efficient response, and the link may not be persistent long after the evaluation campaign. See an example query at `https://ndownloader.figshare.com/files/26321950`

[17] `https://tiad2021.unizar.es/data/TransSets_ApertiumRDFv2_1_CSV.zip`

[18] See the "how to get the data source" section at `https://tiad2021.unizar.es/task.html`

3.2 Gold standard

The evaluation of the results was carried out by the organisers against manually compiled language pairs of K Dictionaries, extracted from its Global series, particularly the following pairs: BR-EN, EN-BR, FR-EN, EN-FR, FR-PT, PT-FR. The translation pairs extracted from these dictionaries served as a golden standard and remained blind to the participants. Notice that the Brazilian Portuguese variant was used for the translations to/from English (whereas the European Portuguese variant was used with French), which might introduce a bias; however its influence should be equivalent to every participant system thus still allowing for a valid comparison.

Given the fact that the coverage of KD is not the same as Apertium, we took the subset of KD that is covered by Apertium to build the gold standard and allow comparisons, i.e., those KD translations for which the source and target terms are present in both Apertium RDF source and target lexicons. This is shown graphically in Figure 2 for the FR-PT pair.

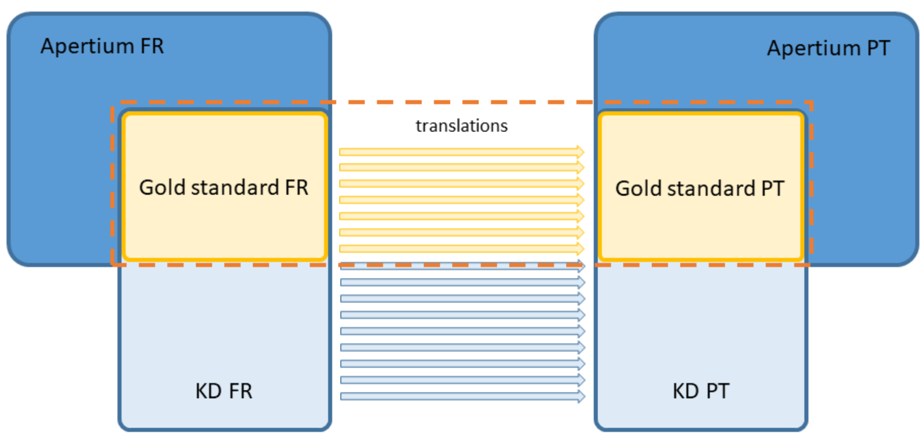

Fig. 2. Gold standard construction for the FR-PT pair. The translations in the dashed area in the middle of the figure constitute the gold standard, selected amongst all the KD translations (for FR-PT) for which both source and target lexical entries are present in their respective Apertium RDF lexicons.

Table 1 shows the size (in number of translations) of the different language pairs in the gold standard. This number might differ with previous TIAD editions because since TIAD'20 the golden standard data have been curated with respect to the initial version in several aspects (see [8]) and, further, the use of a larger Apertium graph in TIAD'21 might have slightly changed the overlap degree between Apertium lexica and KD data.

Table 1. Number of translations per language pair in the gold standard.

Lang. pair	Size
EN-FR	12,453
EN-PT	10,151
FR-EN	16,103
FR-PT	7,982
PT-EN	12,219
PT-FR	6,589

4 Evaluation methodology

The participants run their systems locally, using the Apertium RDF data as known translations, to infer new translations among the three studied languages: FR, EN, PT. Once the output data (inferred translations) were obtained, they loaded the results into a file per language pair in TSV format, containing the following information per row (tab separated):

"source written representation"
"target written representation"
"part of speech"
"confidence score"

The confidence score takes float values between 0 and 1 and is a measure of the confidence that the translation holds between the source and target written representations. If a system does not compute confidence scores, this value had to be put to 1.

4.1 Evaluation process

The organisers compared the obtained results with the gold standard automatically. This process was followed for each system results file and per language pair:

1. Remove duplicated translations (if any).
2. Filter out translations for which the source entry or the target entry are not present in the golden standard (otherwise we cannot assess whether the translation is correct or not). We call *systemGS* the subset of translations that passed this filter, and *GS* the whole set of gold standard translations, in the given language pair.
3. Translations with confidence degree under a given threshold were removed from *systemGS*. In principle, the used threshold is the one reported by participants as the optimal one during the training/preparation phase.
4. Compute the coverage of the system with respect to the gold standard, i.e., how many gold standard entries in the source language were effectively translated by the system (no matter if they were correct or wrong ones).

5. Compute precision as P =(#correct translations in systemGS) / |systemGS |
6. Compute recall as R =(#correct translations in systemGS) / |GS |
7. Compute F-measure as $F = 2 * P * R/(P + R)$

The precision/recall metrics calculated after applying steps 1 to 3 correspond to what in [4] is defined as *both-word precision* and *both-word recall.* The idea is to reduce the penalization to a system for inferring correct translations that are missing in the golden standard dictionary because human editors might have overlooked them when elaborating the dictionary. Note that in previous TIAD editions we only filtered out translations for which the source entry was not present in the translation (step 2), which led to computing the so-called one-word precision/recall, thus only partially covering such a goal.

4.2 Baselines

We have run the above evaluation process with results obtained with two baselines, to be compared with the participating systems' results:

Baseline 1 - Word2Vec. The method uses Word2Vec [14] to transform the graph into a vector space. A graph edge is interpreted as a sentence and the nodes are word forms with their POS tag. Word2Vec iterates multiple times over the graph and learns multilingual embeddings (without additional data). We used the Gensim[19] Word2Vec implementation. For a given input word, we calculated a distance based on the cosine similarity of a word to every other word with the target-POS tag in the target language. The square of the distance from source to target word is interpreted as the confidence degree. For the first word the minimum distance is 0.6^2, for the others it is 0.8^2. Therefore multiple results are only in the output if the confidence is not extremely weak. In our evaluation, we applied an arbitrary threshold of 0.5 to the confidence degree[20].

Baseline 2 - OTIC. In short, the idea of the One Time Inverse Consultation (OTIC) method [17] is to explore, for a given word, the possible candidate translations that can be obtained through intermediate translations in the pivot language. Then, a score is assigned to each candidate translation based on the degree of overlap between the pivot translations shared by both the source and target words[21]. In our evaluation, we applied the OTIC method using Spanish as pivot language, and using an arbitrary threshold of 0.5.

Note that in the TIAD'21 edition the Word2Vec baseline, although based on the same principles, has been re-implemented and re-trained to be adapted to the new Apertium RDF v2 dataset, thus leading to different (generally better) results than in the previous TIAD editions. The OTIC baseline, although it does not need re-training, was also re-run for TIAD'21 to be adapted to the new

[19] `https://radimrehurek.com/gensim/`
[20] The code can be found at `https://github.com/kabashi/TIAD2020_word2vec`
[21] You can find the code at `https://gitlab.com/sid_unizar/otic`

Apertium RDF v2 dataset. The results are generally worse than in TIAD'20 (with the smaller Apertium RDF v1 graph).

Strictly speaking, these are not baselines as they are conceived in other shared tasks, meaning naive approaches with a straightforward implementation, but state-of-the-art methods to solve the task.

5 Results

In this section we review the participating systems in TIAD 2021 and their evaluation results.

5.1 Participating systems

Four teams participated in the shared task, contributing with fourteen systems or system variants, which is a record in TIAD (there were four submitted systems in TIAD'17, eleven in TIAD'19, and nine in TIAD'20). Table 2 lists the participant teams and systems.

The first team, Ahmadi et al.[1], presented a range of approaches that mainly relied on the MUSE[22] and VecMap[23] unsupervised cross-lingual word embedding mappings to create the new translation pairs (ULD_MUSE and ULD_vecmap). They also built two regression models based on the analysis of graph features: ULD_graphSVR and ULD_onetaSVR which are two support vector regression models, respectively based on the translation graph and the previous ULD_oneta system [13] that participated in TIAD'20. Another experimental yet unfinished approach was also presented, which exploited the Multilingual BERT (ULD_mbert).

The second team, Goel and Grover [5], proposed the Augmented Cycle Density (ACD) algorithm, which combines two state of the art methods that require no sense information or parallel corpora: Cycle Density (CD) [18] and One Time Inverse Consultation (OTIC)[17]. For TIAD'21, the authors chose Catalan as a pivot language to compute the OTIC component of their framework.

The third team, Steingrímsson et al. [16], tackled the problem from two directions. First, they collected translation candidates by pivoting through intermediary dictionaries, and second, they applied a score to the candidate list by running an ensemble of word alignment tools on parallel corpora and comparing frequency of alignments to frequency of word co-occurrence in the parallel texts.

Finally, the fourth team used two different approaches. The first one was based on word embeddings: a corpus of pseudo-sentences with the translations and their POS was created and used to train an embeddings space. Then, candidate translations were obtained by measuring cosine similarity. This method is very similar to our Word2Vec baseline. The second method was based on Cross-Lingual Word Embeddings (CLWE), using MUSE as a source of mappings between the source language and a pivot, and between a pivot and the

[22] https://github.com/facebookresearch/MUSE

[23] https://github.com/artetxem/vecmap

Table 2. Participant systems.

Team	System	Comment
S. Ahmadi, A. K. Ojha, S. Banerjee, and J. McCrae (National University of Ireland Galway) [1]	ULD_graphSVR	SVR model based on the translation graph
	ULD_onetaSVR	SVR model on top of ONETA
	ULD_oneta2	unsupervised document embedding, machine translation and graph analysis
	ULD_MUSE	based on the MUSE cross-lingual embedding mappings
	ULD_vecmap	based on the Vecmap cross-lingual embedding mappings
	ULD_mbert	based on Multilingual BERT
S. Goel and K. Shaanjeet (International Institute of Information Technology, Hyderabad, Telangana, India) [5]	ACDcat	Augmented Cycle Density, a cycle based approach combined with OTIC with Catalan as a pivot
S. Steingrímsson, H. Loftsson, and A. Way (Reykjavik University, Iceland and Dublin City University, Ireland) [16]	PivotAlign-F	pivoting and word alignment, promoting F-measure
	PivotAlign-R	pivoting and word alignment, promoting Recall
	PivotAlign-F	pivoting and word alignment, promoting Precision
T. B. Tuan and C. Ramisch (Aix-Marseille University)	TUANWEsg	embeddings (skip-gram)
	TUANWEcb	embeddings (CBOW)
	TUANMUSEca	cross-lingual embeddings with MUSE (CA as a pivot)
	TUANMUSEes	cross-lingual embeddings with MUSE (ES as a pivot)

target language. Notice that we cannot refer to a detailed description of the system because the authors decided not to publish their system description paper, nor to participate in the workshop. We still include their result here for completeness.

5.2 Evaluation results

The complete evaluation results per system and per language pair are accessible in the TIAD 2021 website[24]. In order to give an overview of the results, we include here Table 3, which shows the averaged results, evaluated by using the confidence threshold that every participant reported as optimal according to their internal tests.

[24] See `https://tiad2021.unizar.es/results.html` under the section "Evaluation results".

Table 3. Averaged system results, ordered by F-measure in descending order.

System	Precision	Recall	F-measure	Coverage
PivotAlign-R	0.71	0.58	0.64	0.77
PivotAlign-F	0.81	0.51	0.62	0.68
ACDcat	0.75	0.53	0.61	0.75
TUANWEsg	0.81	0.47	0.59	0.76
TUANWEcb	0.81	0.47	0.59	0.76
ULD_graphSVR	0.70	0.49	0.57	0.69
PivotAlign-P	0.86	0.24	0.37	0.33
baseline-Word2Vec	0.69	0.23	0.33	0.40
ULD_MUSE	0.29	0.41	0.33	0.65
baseline-OTIC	0.78	0.18	0.29	0.28
ULD_onetaSVR	0.76	0.10	0.17	0.14
TUANMUSEca	0.86	0.10	0.16	0.16
TUANMUSEes	0.87	0.08	0.13	0.14
ULD_oneta2	0.64	0.07	0.13	0.11
ULD_vecmap	0.36	0.01	0.01	0.02
ULD_mbert	0.00	0.00	0.00	0.11

5.3 Discussion

As can be seen in Table 3, there are a number of participating systems that obtained better results than both baselines in terms of F-measure. This is new in this edition, since in previous TIAD editions there was no system beating both baselines. Interestingly, the OTIC method, based on purely graph exploration and dated back to 1994, systematically outperformed more contemporary methods based on word embeddings and distributional semantics, which gives an idea of the difficulty of the task. This year's results confirm our intuition that OTIC was not an upper bound and that there were still much room for improvement for more recent methods. In particular, PivotAlign-R and PivotAlign-F gave the best results in terms of F-measure, closely followed by the ACDcat, TUANWEsg, TUANWEcb and ULD-graphSVR.

Of course a direct comparison between this year's results with those of previous TIAD editions is only relative because of two reasons: (1) the use of a new development dataset in TIAD'21, which is the larger Apertium RDF v2 graph instead of Apertium RDF v1, which lead to a different gold standard (Apertium-KD intersection) and (2) the use of improved metrics in the evaluation, as explained above: both-word precision and both-word recall. It seems also clear that the availability of a larger graph to infer translations had a positive effect in the system's performance.

Note that the precision values shown in Table 3 are conservative since there is a small but undefined number of false negatives (correct translations that are not present in the gold standard) that can be found in the results. For example, from the EN→FR set of translations: "wizard"→"sorcier" (noun), "abandon"→"quitter" (verb) or "dump"→"vider" (verb).

6 Conclusions

In this paper we have given an overview of the 4th Translation Inference Across Dictionaries (TIAD) shared task, and a description of the results obtained by the 14 participating systems and two baselines. In this edition, the participating systems were asked to generate new translations automatically among English, French, Portuguese, based on known indirect translations contained in the Apertium RDF graph. This time, a new larger version of the data graph was used, that is Apertium RDF v2. The evaluation of the results was carried out by the organisers against manually compiled pairs of K Dictionaries.

The results are good and illustrate improvement in the area of translation inference across dictionaries, despite the difficulty of the task. However, we consider that the task is far from being solved, with much room for improvement and many other aspects and languages to be explored.

7 Acknowledgements

This work has been supported by the European Union's Horizon 2020 research and innovation programme through the projects Prêt-à-LLOD (grant agreement No 825182) and Elexis (grant agreement No 731015). This work is also based upon work from COST Action CA18209 – NexusLinguarum "European network for Web-centred linguistic data science", supported by COST (European Cooperation in Science and Technology). It has been also partially supported by the Spanish project PID2020-113903RB-I00 (AEI/FEDER, UE), by DGA/FEDER, and by the *Agencia Estatal de Investigación* of the Spanish Ministry of Economy and Competitiveness and the European Social Fund through the "Ramón y Cajal" program (RYC2019-028112-I).

References

1. Ahmadi, S., Ojha, A.K., Banerjee, S., McCrae, J.P.: Nuig at tiad 2021: Cross-lingual word embeddings for translation inference. In: Proc. of LDK 2021 workshops and tutorials. CEUR-WS (September 2021)
2. Flati, T., Navigli, R.: The CQC Algorithm: Cycling in Graphs to Semantically Enrich and Enhance a Bilingual Dictionary (Extended Abstract). In: Proce. of the 23th International Joint Conference on Artificial Intelligence. pp. 3151–3155. IJCAI '13, AAAI Press (2013)
3. Forcada, M.L., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J.A., Sánchez-Martínez, F., Ramírez-Sánchez, G., Tyers, F.: Apertium: a free/open-source platform for rule-based machine translation. Machine Translation **25**(2), 127–144 (2011)
4. Goel, S., Gracia, J., Forcada, M.L.: Bilingual dictionary generation and enrichment via graph exploration [UNDER REVIEW]. Semantic Web Journal (2021), `http://www.semantic-web-journal.net/content/bilingual-dictionary-generation-and-enrichment-graph-exploration`

5. Goel, S., Grover, K.S.S.: From pivots to graphs: Augmented cycle density as a generalization to one time inverse consultation. In: Proc. of LDK 2021 workshops and tutorials. CEUR-WS (September 2021)
6. Gracia, J., Fäth, C., Hartung, M., Ionov, M., Bosque-Gil, J., Veríssimo, S., Chiarcos, C., Orlikowski, M.: Leveraging Linguistic Linked Data for Cross-Lingual Model Transfer in the Pharmaceutical Domain. In: Fu, B., Polleres, A. (eds.) Proc. of 19th International Semantic Web Conference (ISWC 2020). pp. 499–514. Springer (2020), http://link.springer.com/10.1007/978-3-030-62466-8_31
7. Gracia, J., Villegas, M., Gómez-Pérez, A., Bel, N.: The apertium bilingual dictionaries on the web of data. Semantic Web **9**(2), 231–240 (2018)
8. Kernerman, I., Krek, S., McCrae, J.P., Gracia, J., Ahmadi, S., Kabashi, B. (eds.): Proceedings of Globalex 2020 Workshop on Linked Lexicography. ELRA (2020), https://www.aclweb.org/anthology/2020.globalex-1.0/
9. Lim, L.T., Ranaivo-Malançon, B., Tang, E.K.: Low Cost Construction of a Multilingual Lexicon from Bilingual Lists. Polibits **43**, 45–51 (2011)
10. Mausam, Soderland, S., Etzioni, O., Weld, D.S., Skinner, M., Bilmes, J.: Compiling a Massive, Multilingual Dictionary via Probabilistic Inference. In: Proc. of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1. pp. 262–270. ACL '09, Association for Computational Linguistics, Stroudsburg, PA, USA (2009)
11. McCrae, J., Aguado-de Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., Wunner, T.: Interchanging lexical resources on the Semantic Web. Language Resources and Evaluation **46**, 701–719 (2012)
12. McCrae, J.P., Bosque-Gil, J., Gracia, J., Buitelaar, P., Cimiano, P.: The OntoLex-Lemon Model: Development and Applications. In: Electronic lexicography in the 21st century. Proc. of eLex 2017 conference, in Leiden, Netherlands. pp. 587–597. Lexical Computing CZ s.r.o. (sep 2017), https://elex.link/elex2017/wp-content/uploads/2017/09/paper36.pdf
13. McCrae, J.P., Arcan, M.: NUIG at TIAD: Combining unsupervised NLP and graph metrics for translation inference. In: Proceedings of the 2020 Globalex Workshop on Linked Lexicography. pp. 92–97. European Language Resources Association, Marseille, France (May 2020), https://aclanthology.org/2020.globalex-1.15
14. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space. In: Proc. of International Conference on Learning Representations (ICLR) (2013)
15. Montiel-Ponsoda, E., Gracia, J., Aguado-De-Cea, G., Gómez-Pérez, A.: Representing translations on the semantic Web. In: Proc. of the 2nd International Workshop on the Multilingual Semantic Web (MSW) at ISWC '11. vol. 775. CEUR Press (2011)
16. Steingrímsson, S., Loftsson, H., Way, A.: Pivotalign: Leveraging high-precision word alignments for bilingual dictionary inference. In: Proc. of LDK 2021 workshops and tutorials. CEUR-WS (September 2021)
17. Tanaka, K., Umemura, K.: Construction of a Bilingual Dictionary Intermediated by a Third Language. In: COLING. pp. 297–303 (1994)
18. Villegas, M., Melero, M., Bel, N., Gracia, J., Bel, N.: Leveraging RDF Graphs for Crossing Multiple Bilingual Dictionaries. In: Proc. of 10th Language Resources and Evaluation Conference (LREC'16) Portorož (Slovenia). pp. 868–876. European Language Resources Association (ELRA), Paris, France (may 2016)