Proceedings of the TERM21 Workshop , pages 33–40
Language Resources and Evaluation Conference (LREC 2022), Marseille, 20-25 June 2022

D-Terminer: Online Demo for Monolingual and Bilingual Automatic Term Extraction

Ayla Rigouts Terryn, Veronique Hoste and Els Lefever
LT3, Language and Translation Technology Team
Department of Translation, Interpreting and Communication – Ghent University
Groot-Brittanniëlaan 45, 9000 Ghent, Belgium
{firstname.lastname}@ugent.be

Abstract

This contribution presents D-Terminer: an open access, online demo for monolingual and multilingual automatic term extraction from parallel corpora. The monolingual term extraction is based on a recurrent neural network, with a supervised methodology that relies on pretrained embeddings. Candidate terms can be tagged in their original context and there is no need for a large corpus, as the methodology will work even for single sentences. With the bilingual term extraction from parallel corpora, potentially equivalent candidate term pairs are extracted from translation memories and manual annotation of the results shows that good equivalents are found for most candidate terms. Accompanying the release of the demo is an updated version of the ACTER Annotated Corpora for Term Extraction Research (version 1.5).

Keywords: automatic term extraction, multilingual term extraction, terminology

1. Introduction

Based on the D-TERMINE (Data-driven Term Extraction Methodologies Investigated) PhD research (Rigouts Terryn, 2021), an online demo, D-Terminer[1], has been developed for automatic term extraction, i.e., the automatic identification of specialised, domain-specific vocabulary in text. The D-Terminer demo supports monolingual term extraction in English, French, Dutch, and German, as well as bilingual automatic term extraction from parallel corpora with pairs of those same languages. The code is open source[2] and the service is freely available, though restrictions apply to the maximum allowed volume of submitted texts. This is an ongoing project with research plans for improvements in many directions, ranging from more advanced term extraction to more customisation and export options. The monolingual methodology has been elaborately described in previous work (Rigouts Terryn et al., 2022), so the current contribution will focus on the methodology and evaluation of the multilingual term extraction.

Accompanying the launch of this demo is the release of an updated version (1.5) of the Annotated Corpora for Term Extraction Research (ACTER) dataset (Rigouts Terryn et al., 2020b), also freely available online under a Creative Commons license (Ayla Rigouts Terryn, Veronique Hoste and Els Lefever, 2022)[3]. Apart from some minor improvements in the annotations themselves (removal of overly long Named Entities and normalisation of accented uppercase "I" character to avoid issues with lowercasing), the main difference is that the annotated terms have now been made available as sequential annotations in the original context, to complement the original format of lists of unique annotations. After a brief overview of the related research, the update of the dataset is discussed. The next section is dedicated to the monolingual term extraction methodology and its implementation into the demo. Next, the methodology and evaluation of the bilingual term extraction are discussed, before concluding with an overview and future research plans.

2. Related Research

Over the past decades, research into monolingual automatic term extraction first evolved from linguistic (e.g., (Justeson and Katz, 1995) and statistical (Sparck Jones, 1972) methodologies to hybrid methodologies. These rule-based hybrid methodologies combine linguistic information like part-of-speech patterns, with statistical metrics used to calculate termhood and unithood (Kageura and Umino, 1996), which measure how related the candidate term (CT) is to the domain, and, in case of candidate multi-word terms, whether the individual components form a cohesive unit. Rule-based hybrid methodologies reached state-of-the-art results for many years, with early work by, a.o., Daille (1994) and Drouin (1997). Variations are still being developed and used more recently as well, e.g. (Kosa et al., 2020; Steingrímsson et al., 2020; Truica and Apostol, 2021). However, (supervised) machine learning methods have become more popular for automatic term extraction, just like for most other areas in natural language processing. Early attempts used algorithms such as AdaBoost (Vivaldi et al., 2001; Patry and Langlais,

[1] D-Terminer demo:
https://lt3.ugent.be/dterminer/

[2] D-Terminer GitHub repository:
https://github.ugent.be/lt3/D-Terminer/

[3] ACTER GitHub repository:
https://github.com/AylaRT/ACTER

2005), RIPPER rule induction (Foo and Merkel, 2010), logistic regression (Nokel, Michael et al., 2012; Fedorenko et al., 2013), and many others. This allowed researchers to combine more information and different kinds of information to detect terms, complementing the traditional linguistic and statistical features, e.g., topic modelling (Bolshakova et al., 2013), consultation of external resources and internet searches (Ramisch et al., 2010), and word embeddings (Wang et al., 2016; Amjadian et al., 2018). The rise of deep learning has seen more neural approaches in recent years, e.g., (Kucza et al., 2018; Shah et al., 2019; Hätty, 2020). The latest trend is the use of language models and sequential methods for automatic term extraction (Gao and Yuan, 2019; Lang et al., 2021), where CTs are detected in their original contexts, usually by classifying each token in text as (part of a) term or not. Most commercial term extraction tools (or tools that include term extraction), e.g., MultiTerm Extract[4] and SketchEngine[5], or online demos by researchers, e.g., TermoStat[6] (Drouin, 2003) and TerMine (Frantzi et al., 2000) rely on rule-based hybrid methodologies.

Multilingual automatic term extraction aims to not only detect CTs, but cross-lingual candidate term pairs. Multilingual term extraction can be performed on parallel corpora, or comparable corpora. The current contribution focuses on the former, i.e., corpora of translations that can be aligned. As discussed by Foo (2012), methodologies can broadly be divided into two groups: "align-extract" and "extract-align", depending on whether monolingual CTs are extracted first, or whether alignment is performed first (so multilingual clues can be considered for the monolingual extraction). As stated by Repar et al. (2019), the former is the more common. Nevertheless, there are indications that multilingual information can help during the monolingual extraction phase. The TExSIS tool for bilingual automatic term extraction from parallel corpora (Macken et al., 2013) starts by extracting word alignments with GIZA++ (Och and Ney, 2003). Next, rule-based chunking is applied (Macken and Daelemans, 2010), after which "a bootstrapping approach is used to extract language-pair specific translation rules" (p. 11). CTs can then be generated based on the aligned phrases, which are further filtered based on statistical (termhood) measures. Whether alignment is performed before or after extraction, Moses phrases tables (Koehn et al., 2007) and GIZA++ (Och and Ney, 2003) remain some of the most popular methodologies for the alignment (Ivanović et al., 2022). The use of language models is generally more common for extraction from comparable corpora.

[4]https://www.trados.com/products/multiterm-desktop/

[5]https://www.sketchengine.eu/

[6]http://termostat.ling.umontreal.ca/

3. ACTER 1.5

For transparency and to encourage similar research, the launch of the demo is accompanied by an updated version of the ACTER dataset. Since the methodology for monolingual term extraction is trained on ACTER, we start with a brief description of the dataset and update. ACTER was first launched in 2020 (Rigouts Terryn et al., 2020a) and is a dataset with comparable corpora[7] in three languages (English, French, Dutch), and four domains (corruption, dressage, heart failure, wind energy). Terms and Named Entities have been manually annotated with four different labels (Specific Terms, Common Terms, Out-of-Domain Terms, and Named Entities). In total, ACTER contains 18,928 unique annotations in corpora of 719,265 tokens. Originally, annotations were only made available as lists of unique (lowercased) annotations (without context). Version 1.5 now includes sequential annotations with IOB labels (Inside, Outside, Beginning) as well. The way these annotations were obtained is well-documented in both the related paper (Rigouts Terryn et al., 2022) and the readme.md file associated with the dataset. This was necessary since the dataset was already starting to be used in sequential methods (Lang et al., 2021), where the lists of annotations were mapped back to the original text. Since the original annotations were made in context, and creating a sequential dataset from these annotations is not always straightforward (due to nested annotations etc.), the annotations have now been made available in this well-documented sequential IO(B) format so researchers can all start from the same dataset and compare results. Additionally, tokenised versions of the annotations as lists are now included as well, since the original annotations do not always coincide with token boundaries. The monolingual models used for D-Terminer are based on this version of ACTER.

4. Monolingual Term extraction

The monolingual term extraction in the D-Terminer demo is a supervised system, trained on ACTER. The method is described in more detail in a previous publication (Rigouts Terryn et al., 2022), which includes a thorough evaluation. Since the exact same methodology is used for the demo, with even more available training data (no held-out test corpus), results will be similar (perhaps even slightly better) than those reported. With the Flair framework (Akbik et al., 2019), a recurrent neural network was trained to tag each sequential token in a domain-specific text as (part of) a term or not, using the biLSTM-CRF architecture and pretrained multilingual BERT embeddings (Devlin et al., 2019). This methodology was shown to perform well, though results remain highly dependent on the domain, language, and relevance of the training data.
For monolingual term extraction with the D-Terminer demo, users are first prompted to upload a domain-

[7]except for one parallel corpus in the domain of corruption

specific corpus of one or more plain text (.txt) files. In contrast to most currently available term extraction tools, which rely on statistical termhood and unithood metrics, the D-Terminer methodology will perform equally well on a small corpus (or even a single sentence), as on a larger corpus. Of course, a larger corpus of domain-specific texts will result in a more comprehensive and representative overview of terms in the domain. This first version of D-Terminer only tokenises the corpus and does not perform additional linguistic preprocessing.

Once the corpus has been uploaded, users are redirected to a new page where they can start the monolingual term extraction. There are three customisable settings pertaining to the training data. The first is to choose between an IOB (Inside-Outside-Beginning) or a binary (IO) tagging scheme. Performance was shown to be similar for both, but can have an impact on the results (e.g., more long terms for IO tagging). The second option concerns the domains on which the system will be trained. Training data will always include all ACTER languages (English, French, Dutch), since the models using multilingual BERT were shown to generalise well across languages. Domain, however, was shown to have a bigger impact on results. To extract terms in a domain that does not resemble any of the domains in ACTER (corruption, dressage, heart failure, and wind energy), it is recommended to use a model trained on the entire dataset. If however, the domain is more closely related, it can be beneficial to use a model trained only on the most similar domain. For instance, the corpus on heart failure is trained on medical abstracts and short papers. These texts contain many terms, and many very specific terms. Therefore, to extract terms in a medical text (even one not related specifically to heart failure), results may be better with the model trained only on the heart failure corpus. More detailed descriptions of the corpora can be found on the demo website. The third and final customisable setting for the monolingual term extraction concerns the types of terms that will be extracted. ACTER contains annotations with four labels: Specific Terms, Common Terms, Out-of-Domain Terms, and Named Entities. Users can select a model that focuses on all, or only on a subset of these labels. Since these three customisable settings are mostly relevant for more advanced users, a standard configuration (IOB labels, all domains, all labels) is offered and recommended.

Results of the monolingual term extraction can be viewed in two ways: either a list of all unique CTs (and their frequencies) in a table, or highlighted CTs in the original texts. These results can also be exported.

5. Bilingual Automatic Term Extraction

5.1. Methodology

For the bilingual automatic term extraction, a bilingual domain-specific corpus can be submitted as a translation memory (one or more .tmx files). First, monolingual term extraction is performed on each language separately, as described above. Users then choose the results of one run of the monolingual extraction in the source language (SL), and one run in the target language (TL), to serve as a starting point for the multilingual extraction. For this multilingual methodology, only CTs that have been extracted in the monolingual phase are considered, so no new instances are added.

Once the appropriate monolingual results for SL and TL have been selected, word alignments are calculated using ASTrED aligned syntactic tree edit distance (Vanroy et al., 2021), which is based on *Awesome Align* (Dou and Neubig, 2021) neural word alignment, that relies on multilingual language models.

Alignment scores per SL and TL CT pair are calculated as 2A + 2B + C, where A = (number of complete matches between SL and TL CT)/(frequency of SL CT), B = average match percentage between SL CT and TL CT, and C = (times SL CT and TL CT occur in same aligned sentence)/(frequency of SL CT). This metric was set experimentally and all alignments with a score of at least 0.5 are currently displayed. The threshold was set low on purpose, to favour recall and provide multiple options which may not always be literal translation, but can still be relevant. As with the monolingual extraction, results can either be viewed in a table as seen in Figure 1, with one or multiple potentially equivalent TL CTs per SL CT, or with the candidate terms in context per document, as in Figure 2, with a parallel scroll for SL and TL texts.

5.2. Evaluation: Annotation

The performance of the multilingual term extraction from parallel corpora was manually evaluated on a bilingual (EN-NL) corpus in the domain of corruption. This corpus is part of the training data for the monolingual term extraction, which means that the results of the monolingual term extraction will be exceptionally good, so the evaluation can focus on the performance of the bilingual alignment. Nevertheless, users should be aware that the multilingual extraction is dependent on the results of the monolingual extractions. The corpus consists mainly of texts from EU institutions, including treaties, reports, and other official communication on the subject of corruption. The English and Dutch parts of the corpus count 52,847 and 54,233 tokens respectively. Monolingual term extraction was performed with the standard settings of the D-Terminer demo (IOB labelling, system trained on all domains and all labels). This resulted in a total of 1129 English CTs and 1367 Dutch CTs. Bilingual extraction was performed as described above, once using English as SL and Dutch as TL, once vice versa. Evaluating the results in both directions was important as this has a considerable impact on results, as will be discussed. Three linguists each annotated 100 EN-NL and 100 NL-EN CT pairs, evaluating both the type of instance and the quality of the alignment. The instances were selected

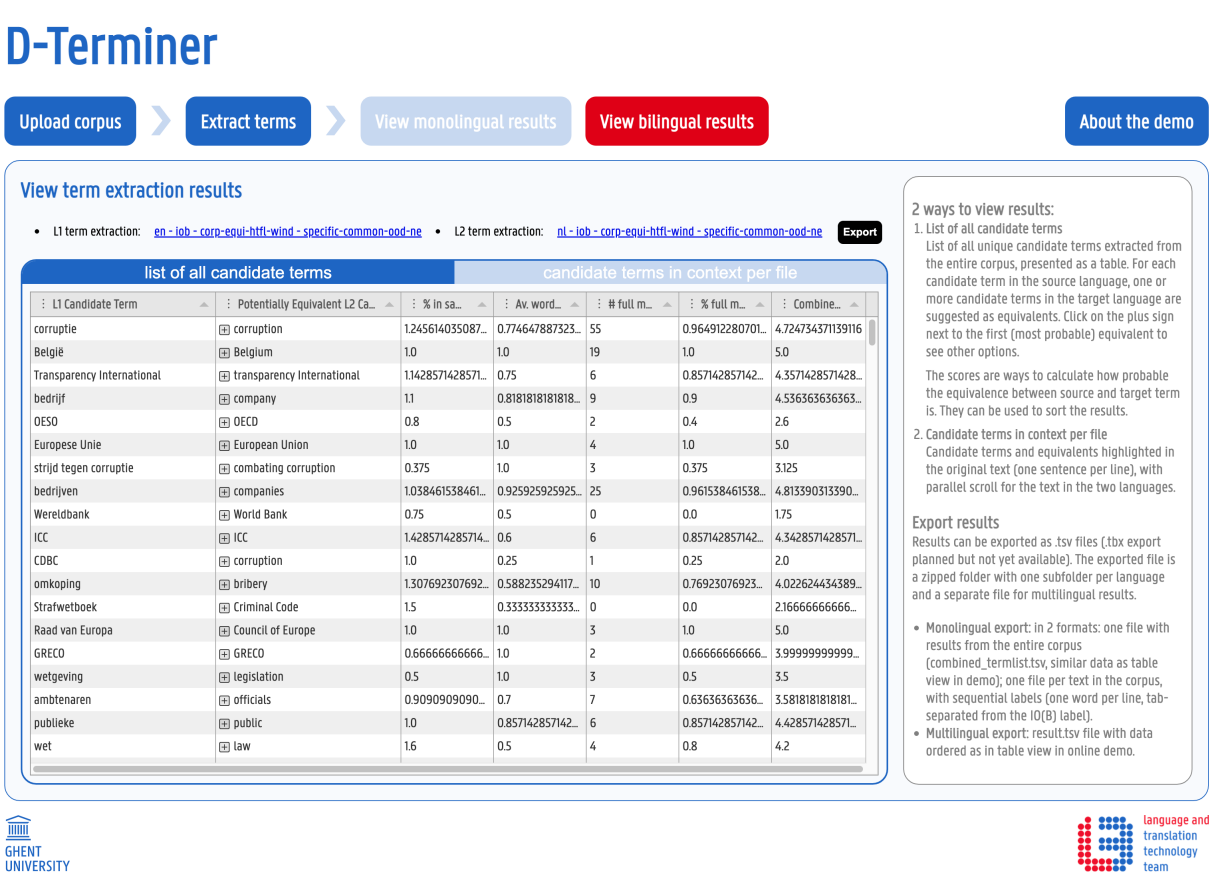

Figure 1: Screenshot of D-Terminer demo, showing multilingual results as list.

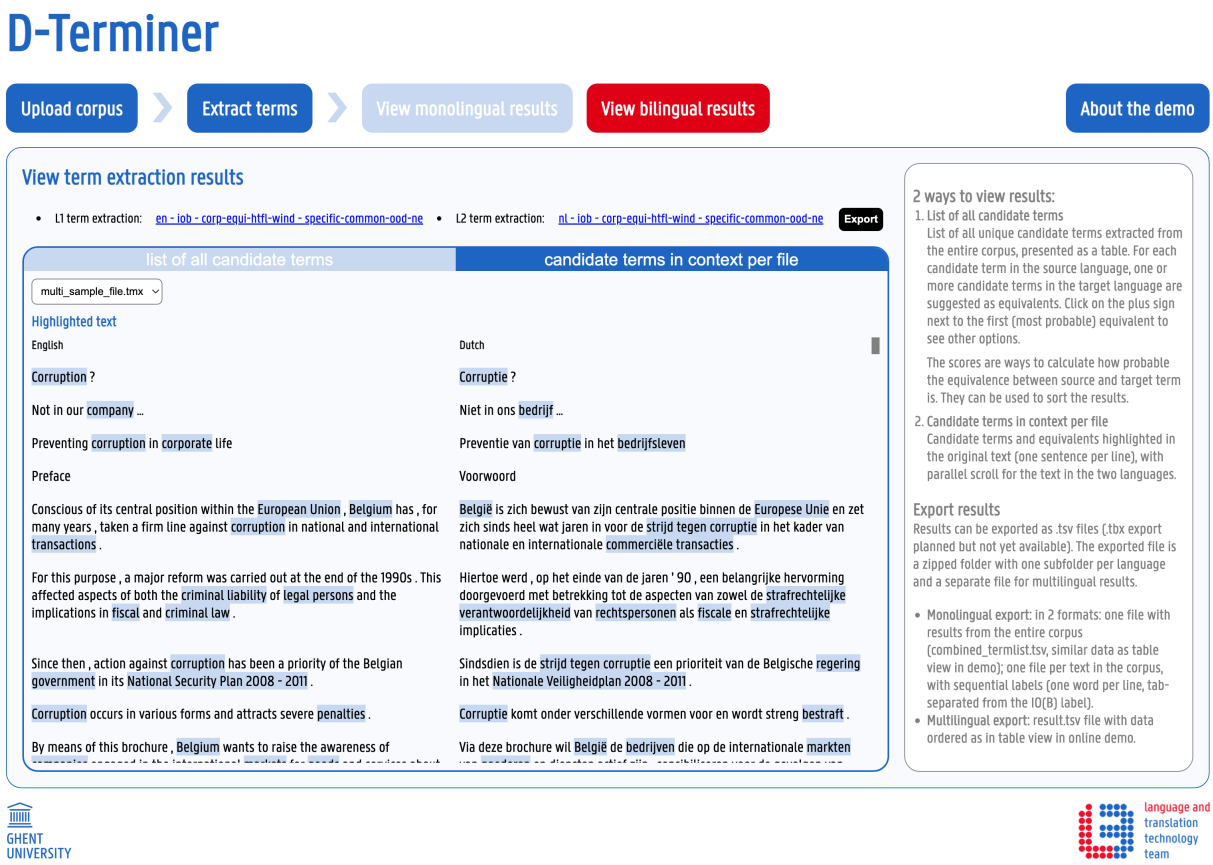

Figure 2: Screenshot of D-Terminer demo, showing multilingual results in context.

by sorting the results by the frequency of the source term, dividing them into 10 sections, and selecting 10 pairs from each section. That way, the evaluation reflects results from different frequency distributions. As there are many CTs that only occur once in the corpus, 41 (EN-NL) and 60 (NL-EN) of the 100 pairs per translation direction were CTs that only occurred once in the entire corpus.

Results were presented to the annotators in a table similar to that used in the online interface (see Figure 1). For each SL CT, annotators had to indicate:

1. Is the **SL CT** a:
   (a) Specific Term (domain- and lexicon-specific),
   (b) Common Term (only domain-specific),
   (c) Named Entity relevant to the domain,
   (d) Named Entity not relevant to the domain, or
   (e) bad candidate (e.g., partial term or Named Entity, clearly neither a term or Named Entity).
2. Is the most highly ranked **TL CT** for the SL CT:
   (a) equivalent,
   (b) equivalent but with a different part-of-speech,
   (c) not equivalent, but useful for a translator, or
   (d) irrelevant

In case the most highly ranked potentially equivalent TL CT was not an exact equivalent (2c or 2d), they also had to indicate whether a correct equivalent was present among the other ranked suggestions and indicate the rank. When the most highly ranked TL CT was found to be a completely irrelevant match (2d) and no exact equivalent was present, they also had to indicate the rank of a potential non-equivalent but relevant TL CT (if present).

Pairwise Cohen's Kappa was used to calculate inter-annotator agreement for annotation tasks 1 (SL CT) and 2 (TL CT). Average agreement (in both translation directions combined) was 0.678 for task 1 and 0.731 for task 2, which are both considered substantial agreement. There were only very small differences between translation directions. Most disagreement on task 1 concerned Specific versus Common Terms (which was expected based on previous experiments (Rigouts Terryn et al., 2020b), especially in this domain). Another recurring issue was differentiating terms from Named Entities, e.g., Named Entities combined with other words/terms (*EU Anti-corruption Reports*) and relevant institutions (*Court of Auditors*). For the annotations of the TL CT, most disagreement was found between the *not equivalent but relevant* and *irrelevant* categories, especially in cases where the suggested equivalent was part of a correct equivalent, e.g., *legal - rechtspersoon [EN: legal person]*, and *anticorruptiestrategie [EN: anti-corruption strategy] - anticorruption*. This is related to the different compounding strategies in Dutch and English (discussed in the next section).

5.3. Evaluation: Results and Discussion

In Table 1, the results of the annotations for both SL and TL CTs can be seen per translation direction and per annotator. The first observation is that the results of the monolingual extraction are very good in both languages. On average, only 5 out of 200 extracted and evaluated CTs were found to be bad candidates. This was expected since the corpus was included in the training data for the monolingual extraction, allowing us to focus on the cross-lingual alignments, i.e., the results of the TL CT evaluation.

The multilingual results are good as well, but with a bigger difference between the languages. For most SL CTs, the most highly ranked potentially equivalent TL CT was evaluated as an actual valid equivalent of the SL CT. For the remainder of this contribution, the evaluation of the suggested equivalent (TL CT) will be based on majority voting, i.e., correct if at least 2 annotators label the TL CT as 2a or 2b. The most highly ranked TL CT was evaluated as a correct equivalent 75.5% of the time. For another 12.5%, an exact equivalent was found among the more lowly ranked suggestions, leaving only 12% of all evaluated CTs without any exact equivalents among the suggested TL CTs. For those 12%, a relevant suggestion was found in most cases and only in 4.5% of the evaluated cases, no relevant suggestion was made at all, including Specific Terms, Common Terms, and Named Entities. Looking at these instances in more detail, a number of explanations can be found. The first and most common cause for a lack of good equivalents in the TL is that the appropriate equivalent was not always extracted during the monolingual extraction phase. For instance, the Dutch CTs *standaardclausules* and *clausules* [EN: *standard clauses* and *clauses*] could not be matched to their English equivalents, because the English forms were not extracted as CTs. This regularly happens because of the different compounding rules in English in Dutch. In Dutch, there are many single-word compounds, of which the equivalent would be written in two words in English. In some cases, this means the Dutch compound is considered a term or Named Entity, while only a part of the English equivalent would be considered as such. For instance, the Dutch *WTO-partners* seems to be a relevant term or Named Entity, but is written as 2 separate words in English (*WTO partners*), where it is logical to extract only *WTO*, so the Dutch CT cannot be matched to the complete equivalent in English, because the latter has not been extracted. Another recurring issue is when the correct equivalent is not present in the source segments, either due to bad alignment, or rephrasing in the translations.

Of the 9 instances for which no relevant or useful equivalents were found at all, only 2 occur more than once. The first is a bad CT: *BUILDING*, which is part of an all-caps title and falsely identified as a CT. It occurs 7 times in total (mostly lowercased in general contexts). The second CT that occurs more than once

		EN-NL				NL-EN			
		Ann1	**Ann2**	**Ann3**	**Av.**	**Ann1**	**Ann2**	**Ann3**	**Av.**
SL CT	a. Specific Term	46	42	59	**49**	50	48	62	**53**
	b. Common Term	22	24	16	**21**	21	25	14	**20**
	c. Relevant Named Entity	16	12	10	**13**	19	15	15	**16**
	d. Irrelevant Named Entity	13	15	14	**14**	9	10	9	**9**
	e. Bad Candidate	3	7	1	**4**	1	2	0	**1**
TL CT	a. Equivalent	79	78	81	**79**	63	61	63	**62**
	b. Equivalent, different POS	3	2	3	**3**	2	1	2	**2**
	c. Not equivalent, relevant	9	9	5	**8**	15	14	3	**11**
	d. Irrelevant, with ranked equiv.	4	4	5	**4**	9	11	16	**12**
	e. Irrelevant, no ranked equiv.	5	7	6	**6**	11	13	16	**13**

Table 1: Annotations of SL CT and most highly ranked option for potentially equivalent TL CT, per language direction and per annotator, including average over all annotators. Since there are 100 instances per experiment, the numbers can be interpreted as percentages.

and for which no good equivalent was found is *instrumentalities*, which occurs twice. The correct equivalent (*hulpmiddelen*) is a more common word in Dutch and was not found by the monolingual extraction. Overall, the system performs slightly better on CTs that occur more than once. For CTs that occur twice or more, the most highly ranked potentially equivalent TL CT is correct 88% of the time, versus 69% of the time for SL CTs that occur only once. The label of the SL CT has a small impact (exact impact depends on annotator), but performance is consistently best for Named Entities (83%) and worst for Specific Terms (71%), with Common Terms in between (80%) (numbers based on SL CT annotations of Annotator 2).

There are a few instances of equivalents with different parts-of-speech, e.g., *investing - investering [EN: investment]*, though this is relatively rare (about 5 out of 200 annotated instances). On average, 8 to 11 percent was annotated as non-equivalent but relevant. Most of these concern pairs where one of the CTs is an equivalent of part of the other CT, e.g., *Court of Auditors - Europese Rekenkamer [EN: European Court of Auditors]*, and *basisdelicten [EN: predicate offences] - offences*. Sometimes, they also concern bad SL CTs, e.g., one which has part of a footnote attached due to bad tokenisation: *criminal justice[54 - strafrecht [EN: criminal justice]*.

6. Conclusion and Future Work

The current contribution describes D-Terminer, an online demo for monolingual and bilingual automatic term extraction. The monolingual extraction is a supervised system trained on annotated data that uses a recurrent neural network to detect terms in context. Users can also upload a parallel corpus in the form of a translation memory to perform bilingual term extraction, and automatically detect potentially equivalent term pairs. Future work on this demo will include more export options (e.g., export as TBX, export of only validated CTs), more advanced monolingual term extraction (combining language model with features), and more linguistic preprocessing (to, e.g., be able to group CTs by lemma). In addition to the online demo, version 1.5 of the ACTER dataset was released, which makes sequential annotations available to users to support research on supervised neural methodologies for term extraction.

7. Acknowledgements

We thank Michaël Lumingu for his help in creating the online D-Terminer demo.

8. Bibliographical References

Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., and Vollgraf, R. (2019). FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 54–59, Minneapolis, USA. Association for Computational Linguistics.

Amjadian, E., Inkpen, D. Z., Paribakht, T. S., and Faez, F. (2018). Distributed Specificity for Automatic Terminology Extraction. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 24(1):23–40.

Bolshakova, E., Loukachevitch, N., and Nokel, M. (2013). Topic Models Can Improve Domain Term Extraction. In David Hutchison, et al., editors, *Advances in Information Retrieval*, volume 7814, pages 684–687. Springer Berlin Heidelberg, Berlin, Heidelberg.

Daille, B. (1994). *Approche Mixte Pour l'Extraction de Terminologie : Statistique Lexicale et Filtres Linguistiques*. PhD thesis in applied sciences, Université Paris Diderot - Paris 7, Paris, France.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*.

Dou, Z.-Y. and Neubig, G. (2021). Word alignment by fine-tuning embeddings on parallel corpora. In *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*.

Drouin, P. (1997). Une Méthodologie d'Identification Automatique des Syntagmes Terminologiques : l'Apport de la Description du Non-terme. *Meta: Journal des traducteurs*, 42(1):45–54.

Drouin, P. (2003). Term Extraction Using Non-Technical Corpora as a Point of Leverage. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 9(1):99–115.

Fedorenko, D., Astrakhantsev, N., and Turdakov, D. (2013). Automatic Recognition of Domain-specific Terms: An Experimental Evaluation. In *Proceedings of the Ninth Spring Researcher's Colloquium on Database and Information Systems*, volume 26, pages 15–23, Kazan, Russia.

Foo, J. and Merkel, M. (2010). Using Machine Learning to Perform Automatic Term Recognition. In *Proceedings of the LREC 2010 Workshop on Methods for Automatic Acquisition of Language Resources and Their Evaluation Methods*, pages 49–54, Valetta, Malta. European Language Resources Association.

Foo, J. (2012). *Computational Terminology: Exploring Bilingual and Monolingual Term Extraction*. Thesis, Linköping Institute of Technology at Linköping University, Linköping.

Frantzi, K., Ananiadou, S., and Mima, H. (2000). Automatic Recognition of Multi-Word Terms: The C-value/NC-value Method. *International Journal of Digital Libraries*, 3(2):117–132.

Gao, Y. and Yuan, Y. (2019). Feature-less End-to-end Nested Term extraction. In Jie Tang, et al., editors, *Proceedings of Natural Language Processing and Chinese Computing*, pages 607–616, Cham. Springer International Publishing.

Hätty, A. (2020). *Automatic Term Extraction for Conventional and Extended Term Definitions Across Domains*. Ph.D. thesis, Universitat Stuttgart, Stuttgart.

Ivanović, T., Stanković, R., Todorović, B. Š., and Krstev, C. (2022). Corpus-based Bilingual Terminology Extraction in the Power Engineering Domain. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, April.

Justeson, J. and Katz, S. (1995). Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text. *Natural Language Engineering*, 1(1):9–27.

Kageura, K. and Umino, B. (1996). Methods of Automatic Term Recognition. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 3(2):259–289.

Koehn, P., Zens, R., Dyer, C., Bojar, O., Constantin, A., Herbst, E., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., and Moran, C. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions - ACL '07*, pages 177–180, Prague, Czech Republic. Association for Computational Linguistics.

Kosa, V., Chaves-Fraga, D., Dobrovolskyi, H., and Ermolayev, V. (2020). Optimized Term Extraction Method Based on Computing Merged Partial C-Values. In *Information and Communication Technologies in Education, Research, and Industrial Applications. ICTERI 2019*, volume 1175 of *Communications in Computer and Information Science*, pages 24–49. Springer International Publishing, Cham.

Kucza, M., Niehues, J., Zenkel, T., Waibel, A., and Stüker, S. (2018). Term Extraction via Neural Sequence Labeling a Comparative Evaluation of Strategies Using Recurrent Neural Networks. In *Proceedings of Interspeech 2018, the 19th Annual Conference of the International Speech Communication Association*, pages 2072–2076, Hyderabad, India, September. International Speech Communication Association.

Lang, C., Wachowiak, L., Heinisch, B., and Gromann, D. (2021). Transforming Term Extraction: Transformer-Based Approaches to Multilingual Term Extraction Across Domains. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3607–3620, Online. Association for Computational Linguistics.

Macken, L. and Daelemans, W. (2010). A Chunk-Driven Bootstrapping Approach to Extracting Translation Patterns. In David Hutchison, et al., editors, *Computational Linguistics and Intelligent Text Processing*, volume 6008 of *Lecture Notes in Computer Science*, pages 394–405. Springer Berlin Heidelberg, Berlin, Heidelberg.

Macken, L., Lefever, E., and Hoste, V. (2013). TExSIS: Bilingual Terminology Extraction from Parallel Corpora Using Chunk-based Alignment. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 19(1):1–30.

Nokel, Michael, Bolshakova, E.i., and Loukachevitch, Natalia. (2012). Combining Multiple Features for Single-word Term Extraction. In *Proceedings of Dialog 2012*, pages 490–501.

Och, F. J. and Ney, H. (2003). A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51.

Patry, A. and Langlais, P. (2005). Corpus-Based Terminology Extraction. In *Terminology and Content Development - Proceedings of the 7th International Conference on Terminology and Knowledge Engineering*, pages 313–321, Copenhagen, Denmark.

Ramisch, C., Villavicencio, A., and Boitet, C. (2010). Mwetoolkit: A Framework for Multiword Expres-

sion Identification. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pages 662–669, Valetta, Malta. European Language Resources Association.

Repar, A., Podpečan, V., Vavpetič, A., Lavrač, N., and Pollak, S. (2019). TermEnsembler: An Ensemble Learning Approach to Bilingual Term Extraction and Alignment. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 25(1):93–120.

Rigouts Terryn, A., Hoste, V., Drouin, P., and Lefever, E. (2020a). TermEval 2020: Shared Task on Automatic Term Extraction Using the Annotated Corpora for Term Extraction Research (ACTER) Dataset. In Béatrice Daille, et al., editors, *Proceedings of the LREC 2020 6th International Workshop on Computational Terminology (COMPUTERM 2020)*, pages 85–94, Marseille, France. European Language Resources Association.

Rigouts Terryn, A., Hoste, V., and Lefever, E. (2020b). In No Uncertain Terms: A Dataset for Monolingual and Multilingual Automatic Term Extraction from Comparable Corpora. *Language Resources and Evaluation*, 54(2):385–418.

Rigouts Terryn, A., Hoste, V., and Lefever, E. (2022). Tagging Terms in Text: A Supervised Sequential Labelling Approach to Automatic Term Extraction. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 28(1).

Rigouts Terryn, A. (2021). *D-TERMINE: Data-driven Term Extraction Methodologies Investigated*. Doctoral thesis, Ghent University, Ghent, Belgium.

Shah, S., Sarath, S., and Shreedhar, R. (2019). Similarity Driven Unsupervised Learning for Materials Science Terminology Extraction. *Computación y Sistemas*, 23(3):1005–1013.

Sparck Jones, K. (1972). A Statistical Interpretation of Term Specificity and its Application in Retrieval. *Journal of documentation*, 28(1):11–21.

Steingrímsson, S., orbergsdóttir, Á., Danielsson, H., and Ornolfsson, G. T. (2020). TermPortal: A Workbench for Automatic Term Extraction from Icelandic Texts. In *Proceedings of the 6th International Workshop on Computational Terminology (COMPUTERM 2020)*, pages 8–16, Marseille, France. European Association for Machine Translation.

Truica, C.-O. and Apostol, E.-S. (2021). TLATR: Automatic Topic Labeling Using Automatic (Domain-Specific) Term Recognition. *IEEE Access*, 9:76624–76641.

Vanroy, B., De Clercq, O., Tezcan, A., Daems, J., and Macken, L. (2021). Metrics of Syntactic Equivalence to Assess Translation Difficulty. In Michael Carl, editor, *Explorations in Empirical Translation Process Research*, volume 3, pages 259–294. Springer International Publishing, Cham.

Vivaldi, J., Màrquez, L., and Rodríguez, H. (2001). Improving Term Extraction by System Combination Using Boosting. In Luc Raedt et al., editors, *Proceedings of the 12th European Conference on Machine Learning (ECML 2001)*, volume 2167, pages 515–526, Berlin, Heidelberg. Springer Berlin Heidelberg.

Wang, R., Liu, W., and McDonald, C. (2016). Featureless Domain-Specific Term Extraction with Minimal Labelled Data. In *Proceedings of Australasian Language Technology Association Workshop*, pages 103–112, Melbourne, Australia.

9. Language Resource References

Ayla Rigouts Terryn, Veronique Hoste and Els Lefever. (2022). *ACTER Annotated Corpora for Term Extraction Research, version 1.5*. distributed via CLARIN: http://hdl.handle.net/20.500.12124/38.